FACTORS CONTRIBUTING TO INCOMPLETE REPORTING OF PATIENT RISK FACTOR INFORMATION ON HIV/AIDS CASE REPORT FORMS IN THE STATE OF GEORGIA

By

Greg Bautista
Degree to be awarded: M.P.H.
Career MPH

Patrick Sullivan, PhD, DVM
Associate Professor, Department of Epidemiology,
Rollins School of Public Health at Emory University

3/29/13
Date

Yohance Whiteside, PhD
Epidemiologist, Centers for Disease Control & Prevention (CDC)

3/29/13
Date

Melissa Alperin, MPH, MCHES
Chair, Career MPH Program
Rollins School of Public Health at Emory University

4/2/13
Date

Distribution Agreement

In presenting this Thesis as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my Thesis in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this Thesis. I retain all ownership rights to the copyright of the Thesis. I also retain the right to use in future works (such as articles or books) all or part of this Thesis.

Gus Bantsa
Signature of Student

April 2, 2013
Date

FACTORS CONTRIBUTING TO INCOMPLETE REPORTING OF PATIENT RISK FACTOR INFORMATION ON HIV/AIDS CASE REPORT FORMS IN THE STATE OF GEORGIA

By

Greg Bautista
Degree to be awarded: M.P.H.
Career MPH

Thesis Committee Chair: Patrick Sullivan, PhD, DVM

An abstract of
A Thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements of the degree of
Master of Public Health in the Career MPH program
2013

Abstract

FACTORS CONTRIBUTING TO INCOMPLETE REPORTING OF PATIENT RISK FACTOR INFORMATION ON HIV/AIDS CASE REPORT FORMS IN THE STATE OF GEORGIA

By
Greg Bautista

HIV/AIDS case reporting is required in all fifty states and includes reporting of patient demographics, laboratory data, treatment and a brief history of patient risk behavior. An algorithm developed in the 1980's by the Centers for Disease Control and Prevention (CDC) assigns each case a transmission category based on patient history variables. Among cases with sufficient risk history information, the three transmission categories which collectively account for the vast majority of HIV/AIDS cases are male sexual contact with another male (MSM); heterosexual contact with a person known to have HIV infection or at least with a person at increased risk of HIV infection (based on a history of MSM, IDU or receipt of blood products) and receipt of non-prescribed drugs by injection, intravenously, intramuscularly, or subcutaneously ("injection drug use" or IDU). Unfortunately, many cases are reported with little or no risk history information and thus do not meet the criteria for any of the CDC-defined transmission categories. According to the Georgia Department of Public Health (DPH), each year a significant percentage of HIV/AIDS cases diagnosed in Georgia lack complete patient risk history information. For example, in 2011 70% of HIV cases diagnosed among males (n=2,002) and 92% of HIV cases diagnosed among females (n=787) were in the category of "no identified risk" (NIR). This is a growing problem of public health significance as health departments depend on the completeness of surveillance data to monitor changes in HIV/AIDS incidence, track the burden of disease, plan programs and services, allocate limited resources for HIV care, and develop strategies for targeting prevention interventions to populations most at risk of infection. A literature review was conducted to identify factors that have been shown to be associated with incomplete reporting of patient information as well as strategies with demonstrated effectiveness for addressing this challenge. Also, a logistic regression analysis was conducted of 25,022 adult AIDS cases in the CDC AIDS Public Information Dataset (APIDS). This analysis identified the following variables as significantly associated with the binary outcome of having or not having sufficient risk information for transmission category classification: patient age at diagnosis, race/ethnicity, residence in a large metropolitan area, birth in the United States and sex at birth.

FACTORS CONTRIBUTING TO INCOMPLETE REPORTING OF PATIENT RISK FACTOR INFORMATION ON HIV/AIDS CASE REPORT FORMS IN THE STATE OF GEORGIA

By

Greg Bautista
Degree to be awarded: M.P.H.
Career MPH

Thesis Committee Chair: Patrick Sullivan, PhD, DVM

A Thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements of the degree of
Master of Public Health in the Career MPH program
2013

Table of Contents

List of Figures

List of Tables

List of Attachments

Attachment #1:	Key informant interview guide
Attachment #2:	Georgia confidential HIV/AIDS case report form (2 pages)
Attachment #3:	CDC 2011 Adult HIV Confidential Case Report Form (4 pages)
Attachment #4:	Codebook for the APIDS public access database
Attachment #5:	Exempt study approval letter (Emory IRB)
Attachment #6:	Code for logistic regression model in SAS 9.2
Attachment #7:	SAS output

1. Background

1.1. Introduction

Throughout the United States health departments require the reporting of all cases of Human Immunodeficiency Virus (HIV) and Acquired Immune Deficiency Syndrome (AIDS) for surveillance of HIV/AIDS incidence and prevalence but a significant and growing proportion of cases are reported with missing patient risk history information. This is a growing problem of public health significance as health departments depend on the completeness of surveillance data to monitor changes in HIV/AIDS incidence, track the burden of disease, plan programs and services, allocate limited resources for HIV care, and develop strategies for targeting prevention interventions to populations most at risk of infection.

The importance of HIV/AIDS case reporting

The collection and reporting of accurate, timely and complete HIV/AIDS case information is of critical importance to state and local health departments for a number of reasons. The completeness and accuracy of HIV cases reported in a particular jurisdiction directly impacts the amount of Federal funds received by that jurisdiction for HIV/AIDS surveillance (Page, 2010) and medical care for uninsured residents living with HIV under the Ryan White CARE Act. (Glynn, 2007; Nash, 2007) Equally important, errors in the reporting of HIV/AIDS cases (such as the reporting of incomplete information on a case form or not reporting a case at all) are of concern because such errors may contribute to the appearance of false trends. (Nwanyanwu, 1993)

1.2. Problem statement and context: A growing nationwide problem

In the CDC Enhanced HIV/AIDS Reporting System (eHARS), cases are classified as "No Risk Reported" (NRR) if the initial report form lacked sufficient patient risk factor information to meet the criteria for inclusion in an HIV transmission category as defined by the Centers for Disease Control and Prevention (CDC). (CDC, 2005) If a health department is unable to identify risk factors for NRR cases through follow-up investigation within 12 months of the initial diagnosis date, such cases are then classified as "No Identified Risk" (NIR). The three HIV transmission categories which collectively account for the vast majority of HIV/AIDS cases are 1) male sexual contact with another male (MSM); 2) heterosexual contact with a person known to have HIV infection or at least with a person at increased risk of HIV infection (based on a history of MSM, IDU or receipt of blood products) and 3) receipt of non-prescribed drugs by injection, intravenously, intramuscularly, or subcutaneously ("injection drug use" or IDU). An informed response to the epidemic factors into consideration the proportion of cases comprised by these groups and differences that may exist among them in terms of behaviors and socio-demographic circumstances. However, the large proportion of incomplete patient transmission category data hampers the ability to develop and implement an informed response to the epidemic.

With time, the problem of incomplete reporting of HIV/AIDS patient risk information has worsened. In 1985, fewer than 5% of AIDS cases reported to the CDC were initially lacking risk factor information. With more laboratories reporting AIDS cases and increasing numbers of persons tested for HIV, health departments began receiving a larger proportion of cases initially reported without a risk factor mainly because laboratories are generally not setup to interview patients on their history of risk behavior. (Glynn, 2007) Since then, the proportion of cases

categorized as NIR/NRR has climbed steadily and reached 35% in 2004. (McDavid, 2006) Data from this period is not available to analyze trends in the proportion of HIV (non-AIDS) cases categorized as NIR/NRR because the majority of states did not initiate name-based HIV reporting until the mid-1990s. (Kaiser Family Foundation, 2011)

Figure 1 shows the number of adult and adolescent AIDS cases reported from 1981 to 2002 and the percentage of cases classified as NIR during this period. From 1981 to 1991 the percentage of reported AIDS cases classified as NIR was relatively low and even included eight years with a percentage below 5%. Beginning in 1992, the percentage gradually increased by approximately two to three percentage points each year to a peak of 29% in 2002 while the total number of reported AIDS cases generally decreased.

A similar trend was noted for the Atlanta Metropolitan Statistical Area (MSA). As shown in Figure 2, the total number of AIDS cases diagnosed and reported in the Atlanta MSA from 1981 to 2002 followed a trend similar to that seen at the national level. The annual increase in the NIR percentage may be attributed to a number of factors including the significant overall increase in the volume of newly-diagnosed and living HIV/AIDS cases.

According to the Georgia Department of Public Health (DPH), each year a significant percentage of HIV/AIDS cases diagnosed in Georgia lack complete patient risk history information. For example, in 2011 70% of HIV cases diagnosed among males (n=2,002) and 92% of HIV cases diagnosed among females (n=787) were in the category of "no identified risk" (NIR). (DPH, 2013; Tables 1-2)

Figure 1. Number of AIDS cases diagnosed from 1981 to 2002 among adults and adolescents and percentage classified as "No Identified Risk" (NIR), adjusted for reporting delays -- APIDS dataset

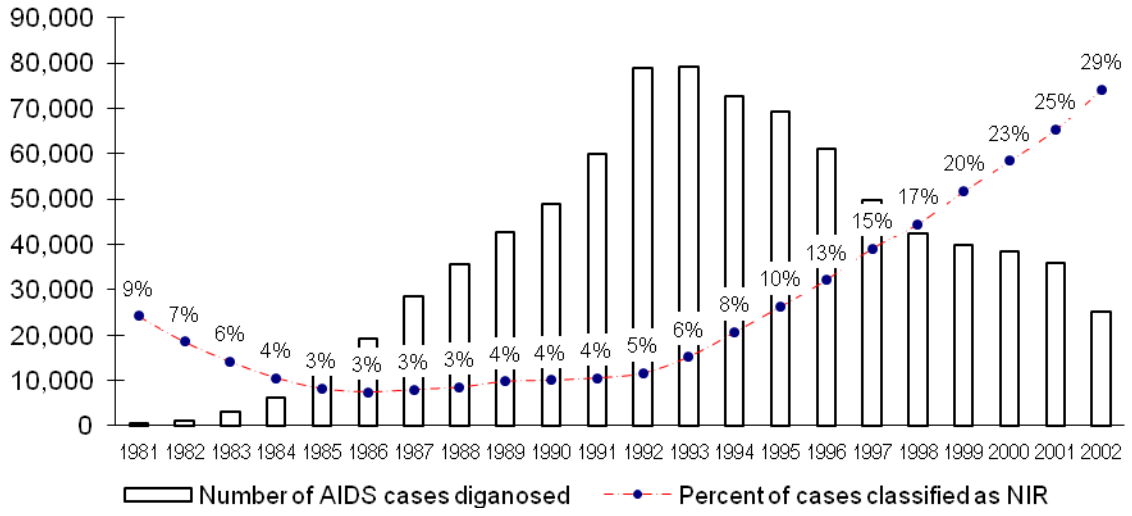


Chart created using data from CDC APIDS dataset. Data http://www.cdc.gov/hiv/topics/surveillance/resources/software/apids/

Figure 2. Number of AIDS cases diagnosed from 1981 to 2002 among adults and adolescents in the Atlanta MSA and percentage classified as "No Identified Risk" (NIR), adjusted for reporting delays -- APIDS dataset

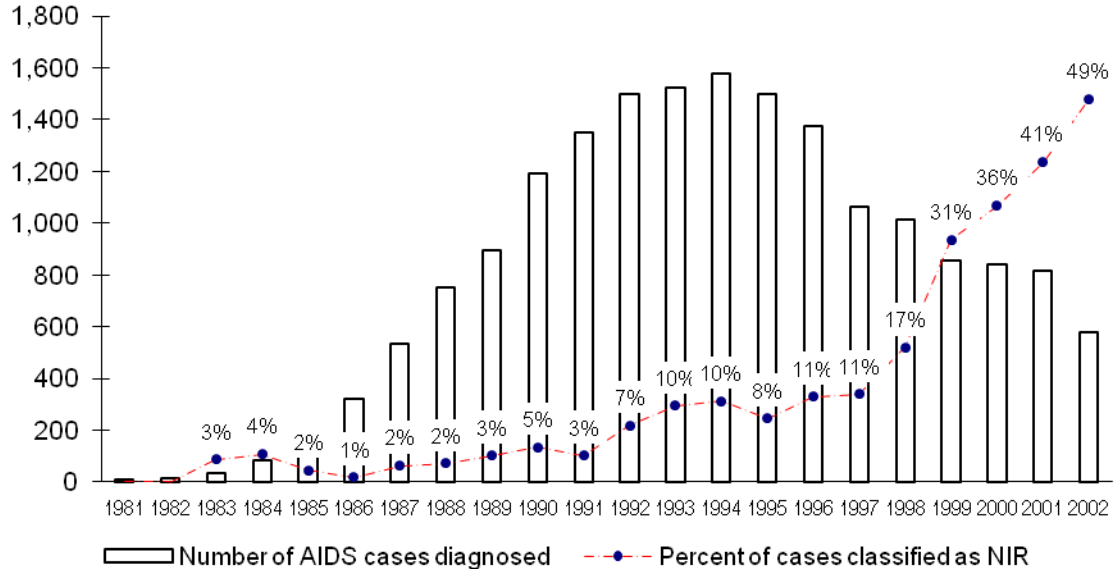


Chart created using data from CDC APIDS dataset. Data http://www.cdc.gov/hiv/topics/surveillance/resources/software/apids/

1.3. Purpose statement

This research will identify factors contributing to the high percentage of HIV/AIDS cases reported in the State of Georgia with missing or incomplete patient risk history information and describe strategies identified in the literature for addressing this problem. The following study questions will frame this research.

1. What factors contribute to the high proportion of incomplete risk factor data reported on HIV/AIDS case forms?
2. What strategies are available to improve the completeness of risk factor data reported on HIV/AIDS case forms?
3. What patient demographic characteristics are associated with having incomplete risk factor data on HIV/AIDS case forms?

A literature view will be conducted to answer the first two study questions. For the third study question, a retrospective study will be implemented using the publicly-available CDC AIDS Public Information Dataset (APIDS) and a regression model will be created in SAS 9.2 to identify APIDS variables that may be associated with incomplete risk factor data on AIDS case forms. The APIDS dataset includes 16 deidentified variables for 859,000 AIDS cases reported[1] to the CDC between 1981 and 2002. In addition, interviews with management-level DPH HIV/AIDS surveillance employees will complement this quantitative data analysis.

[1] The APIDS dataset contains information on AIDS cases reported to the CDC from all fifty states, the District of Columbia, U.S. dependencies and possessions, and independent nations in free association with the United States including Puerto Rico, the U.S. Virgin Islands, Guam, American Samoa, the Republic of Palau, the Republic of the Marshall Islands, the Commonwealth of the Northern Mariana Islands, and the Federated States of Micronesia from 1981 to 2002. The dataset does not contain information on HIV (non-AIDS) cases reported during this time because most states did not initiate name-based HIV reporting until the mid-1990's.

1.4. Significance statement

This research may generate useful information for achieving improvements to the quality of patient data reported to public health HIV/AIDS surveillance programs. Potential benefits from improved data quality include more strategic allocation of resources and improved accuracy in our understanding of the nature of HIV/AIDS transmission throughout the United States. Cost savings may result from the reduced need to implement follow-up investigations to correct problems with missing or incomplete information on HIV/AIDS case report forms.

Also, a significant percentage of persons living with HIV/AIDS are not currently in care (CDC, 2011) and the National HIV/AIDS Strategy calls for increasing the percentage of persons living with HIV/AIDS who are linked to care within three months from 65% to 85%. Using eHARS data, progress toward this goal will be monitored for persons living with HIV/AIDS in each of the various transmission categories as well as by age, race/ethnicity, sex and other demographic variables using eHARS data. Improved acquisition of patient risk factor information in eHARS will help facilitate accurate stratification of patient linkage-to-care data for evaluation of the National HIV/AIDS Strategy. (Fagan, 2010)

1.5 Definition of terms

Note: The following definitions are quoted directly from the CDC Technical Guidance for HIV/AIDS Surveillance Programs, 2005 revision.

Active surveillance: Health department staff regularly contact reporting facilities (hospitals, clinics, physician offices, laboratories) to identify potential/suspect HIV/AIDS cases (or confirm no cases). Health department staff review medical records at provider sites or receive information over the telephone, by fax, e-mail, US mail, etc. to establish an HIV/AIDS

case and to elicit information for HIV/AIDS case report forms. All communication should follow security and confidentiality guidelines.

Complete epidemiologic follow-up: A case reported with HIV/AIDS is considered to have undergone complete epidemiologic follow-up if 1) one or more risk factors (see definition of "risk factors") are identified or confirmed, and 2) all data sources available for a person [see list of follow-up data sources below] have been reviewed and/or contacted for risk factor information; or 3) 12 months have elapsed since the date of the initial case report and no risk factor has been identified.

List of follow-up data sources where risk factor information is most likely to be found:

1. Review of medical charts at health care provider who did not test or report the patient for HIV but for whom the reported person is a patient (according to the ATR project sites, this provider is most likely the current HIV treatment provider for the patient).
2. Review of medical charts at health care provider who tested the patient for HIV.
3. Telephone calls or visits to the health care provider who did not test or report the patient for HIV but for whom the reported person is a patient (most likely the current HIV treatment provider for the patient), but where review of the medical charts was not done.
4. Telephone calls or visits to a social service case manager providing physical and emotional assistance to the patient with HIV.
5. Review of medical charts at health care provider who reported, but did not test, the patient for HIV.

6. Telephone calls or visits to the health care provider who reported, but did not test, the patient for HIV, but where review of the medical charts was not done.
7. Other available data sources or other facilities at which the individual has received care and other reporting sources, such as counseling and testing sites (CTS) and sexually transmitted disease (STD) databases.

Date of initial case report: This is the date on which the public health department receives the first report on a potential case of HIV or AIDS. The document can be a laboratory report, a case report form, a birth certificate, a death certificate, etc. This date is not the date on which the surveillance area enters the reported information into a surveillance system.

Epidemiologic follow-up: This is the investigative process for obtaining additional information on a reported HIV/AIDS case.

Exposure category: The term for a new classification (or any of the categories in it) that summarizes the multiple risk factors that an individual may have had by including combination categories of the three most common ones (MSM, IDU, HTC). The exposure category classification was developed in response to the "Risk Consultation" of December 3-4, 2001, as an alternative to the transmission category classification. The consultants stated that the assumption on which the current hierarchical classification ("transmission category") is based -- that sufficient information is collected to allow accurate selection of the most likely mode of transmission from among multiple possible routes of exposure-- is probably not true, and that the resulting concealment of routes of exposure lower in the hierarchy by those that are higher is therefore unjustified and misleading. They therefore recommended a classification that would be less hierarchical, particularly so that HTC would not be concealed as a risk factor when it occurred in combination with MSM or IDU. The exposure category still is hierarchical with

respect to risk factors other than the primary three groups (e.g., receipt of a blood transfusion), which appear only in single categories ranked lower hierarchically than the combinations of MSM, IDU, and HTC.

No identified risk (factor) (NIR): This is an NRR case for which an HIV risk factor cannot be identified or confirmed 1) although all available data sources have been reviewed or contacted, or 2) epidemiologic follow-up was either not initiated or not completed, but 12 months have elapsed since the date of the initial case report.

No reported risk (factor) (NRR): A case is classified as an NRR if it is reported without any risk factor information or with unconfirmed COPHI risk factor information.

Passive surveillance: The health department receives HIV/AIDS case reports from physicians, laboratories, or other individuals or institutions without regularly contacting the reporting sources.

Risk factors: The collective term for the individual routes of exposure (before the person found out he/she was HIV positive or diagnosed with AIDS) on which data are routinely collected for surveillance of HIV/AIDS cases. They are the following variables:

- Male sexual contact with another male (MSM).
- Receipt of nonprescribed drugs by injection, intravenously, intramuscularly, or subcutaneously ("injection drug use": IDU).
- Heterosexual contact with a person known to have HIV infection or at least with a person at increased risk of HIV infection (based on a history of MSM, IDU, or receipt of blood products).
- Perinatal mother-to-child contact: birth to a woman who was known to have HIV infection or was at least at increased risk of HIV infection.

Risk factor ascertainment: The process of risk factor ascertainment should begin when the surveillance program receives a report of an actual or suspected HIV or AIDS case report without any risk factor information (no risk reported or "NRR"). The risk factor ascertainment should include usual followup activities, such as calling a reporting facility or delegating field staff to inquire on a laboratory or provider report received by the surveillance program. Routine case follow-up should include inquiry about all HIV risk factors for each case, or a sample of cases, for surveillance programs that use the sampling protocol. The investment of time and resources to educate providers/reporters and surveillance staff regarding proper risk factor ascertainment should reduce the number of cases on which followup is needed and help to achieve complete and accurate information on all cases reported to the surveillance system.

Supplemental risk factors: Behaviors or proxies other than "risk factors" that may be associated with various routes of transmission, such as number of sex partners, condom usage, noninjection drug usage, selling sex in exchange for money or drugs, a history of other sexually transmitted diseases, having spent time in prison, and diagnosis of viral hepatitis. Unless specified otherwise, "risk factors" should be assumed to refer only to those listed in the preceding paragraph.

Transmission category: The term for summarizing the multiple risk factors (as defined in "risk factors") that an individual may have had by selecting the one through which HIV was most likely to have been transmitted. The selection of the most likely route of transmission is based on a presumed hierarchical order of transmission that was developed in the early years of the AIDS epidemic, and was based on what was known at the time about how HIV was transmitted. The hierarchy has not changed even though our understanding of the most efficient ways of HIV transmission has changed. The expanded transmission category variable has 5 categories of

heterosexual contact (HTC), which differ by the risk factor of the sex partner, and 8 categories of perinatal (motherto- child) exposure, which differ by the risk factor of the mother. In the transmission category variable (not expanded), the 5 categories of HTC are combined into a single category and the 8 categories of perinatal exposure are combined into a single category. For cases in which there were multiple risk factors, the hierarchical nature of this classification may conceal some risk factors. For example, with a combination of IDU and HTC, only the IDU would be selected and the HTC would be hidden. An exception to the hierarchy is made for the combination of MSM and IDU, in which one of those two risk factors is not selected over the other and both are presented in a combination category. The transmission category and expanded transmission category variables include some categories that come in pairs --one for adults/adolescents and another for children-- for 1) receipt of blood transfusion/transplant, 2) receipt of blood products for treatment of hemophilia, 3) "other" risk factors, and 4) absence of reported/identified risk factors. The list of categories in the transmission category classification is provided below.

1. MSM (male sexual contact with another male)
2. IDU (injection dru use)
3. MSM & IDU
4. Adult hemophiliac
5. HTC (heterosexual contact)
6. Adult transfusion
7. Adult other
8. Adult undetermined
9. Pediatric hemophiliac
10. Mother with HIV
11. Pediatric transfusion
12. Pediatric other
13. Pediatric undetermined

2. Review of the Literature

A literature review was conducted beginning with the following keywords and key phrases entered into the Medline, Psychlit, ERIC and PUBMed databases: HIV surveillance, risk factor acquisition, completeness of HIV case reporting, transmission category, transmission classification, CDC risk factor algorithm, notifiable disease reporting, no identified risk, NIR, no reported risk, NRR, CDC, presumed heterosexual, statistical imputation and risk factor redistribution.

An initial set of articles was gathered and reviewed. This initial round of articles generated an additional set of potential database search terms, which were queried to yield new sources. Additional database queries were implemented following an iterative approach until no additional new articles could be identified. Simultaneously, a set of secondary source documents was compiled for possible inclusion in this literature review including government reports, case report forms currently in use, epidemiological surveillance summaries and HIV surveillance grant progress reports prepared by the Georgia Department of Public Health. Finally, a summary was compiled describing key findings from this literature review with topics organized into the following six subject categories: provider-related factors, patient-related factors, system-related factors, strategies for proactively improving risk factor reporting and strategies for obtaining risk factor information after-the-fact.

Provider-related factors

A broad range of provider-related factors may be associated with incomplete patient risk factor information. Examples include the manner in which the risk history questions are asked,

discomfort discussing sensitive topics, the stigma associated with injection drug use and male-to-male sexual contact, concerns about confidentiality, lack of knowledge of reporting requirements and other factors.

Researchers have noted that some providers discuss risk behavior with their patients more frequently than others. The reasons for such differences are not fully understood but factors such as experience with stigmatized groups, time commitments, number of patients served, area of specialty and years of experience may be factors. In a national sample of 1,096 physicians, researchers observed that as physician experience serving HIV-positive patients increased, provider scores decreased on scales for homophobia and scales for discomfort serving patients who use injection drugs. (Gerbert, 1991)

In a survey of 317 physicians, researchers noted that infectious disease specialists were significantly less likely than family practice physicians to discuss HIV transmission risk reduction (0.4 aOR, 95% CI 0.2-0.9) and condom use (0.5 aOR, 95% CI 0.2-0.9) with their HIV-infected patients even when the analysis controlled for the number of patients living with HIV, years of experience serving patients with HIV and perceived time constraints. (Duffus, 2003)

In the assessment of a patient's risk for HIV/AIDS and other STDs, the validity and reliability of self-reported behavior is affected by how the provider asks the risk history questions. Certain actions may discourage honesty while other approaches may make patients more comfortable disclosing potentially embarrassing experiences. Figure 3 provides a summary of recommendations that have been shown to improve the reliability and validity of sexual behavior assessments. (Weinhardt, 1998)

Figure 3. List of recommendations to improve reliability and validity of HIV-related sexual behaviors reported by patients. (Weinhardt 1998)

1. Use psychometrically evaluated measures
2. Use language that is easily understood
3. Use focus groups, pilot data, and other formative methods to adapt the assessment protocol for sensitivity to cultural issues of the participants
4. Include techniques that improve recall of behavior. Examples include: providing anchor dates for reporting periods, encouraging participants to use appointment books and calendars to recall other memorable events during the reporting period, and recalling extensive periods of abstinence or consistent sexual activities. The "timeline follow-back procedure" utilizes many of these techniques.
5. Establish a working trust with interviewees and questionnaire respondents. The risk assessment should take place after a participant and interviewer have established rapport, and the interviewer has assured the participant of confidentiality.
6. Ask questions in a direct fashion, without apology or hesitancy
7. Adopt default assumptions to gather the most accurate information efficiently. For example: assume minimal understanding so that language is clear and concrete; assume participants will be embarrassed about and have difficulty discussing sexual matters; assume participants will not understand all sexual behavior terms, medical terminology, etc. As the interviewer or investigator learns more about the client, these assumptions can be adjusted.
8. Sequence the inquiry from the least to most threatening questions
9. Place the "burden of denial" on the participant. Consider asking "How many times have you…?" instead of "Have you ever…?"
10. Be sensitive to contextual issues in administration (interviewer's demeanor, physical setting, perceptions of trust regarding personnel)

As mentioned earlier, a significant number of reported HIV/AIDS cases originate from laboratories through automatic electronic reporting. However, laboratories lack the ability to interview patients for risk factor information. If a case is reported by a laboratory with no subsequent reporting from the patient's medical care provider, the case will most likely lack risk factor information. Unfortunately, compliance with HIV/AIDS case reporting varies among medical care providers for a number of reasons. Some providers may be reluctant to report HIV/AIDS case information because of skepticism about the confidentiality of the surveillance system in general. A widely-publicized breach of confidentiality of the Florida AIDS database and the CDC's collaboration with blood banks in the 1980s to help identify transfusion AIDS

cases are compelling examples of historical incidents that may influence provider perceptions of surveillance security and confidentiality. (Colfax 1998)

A survey of a random sample of health care providers in South Carolina revealed broad variation in level of awareness among providers regarding basic aspects of the reportable disease surveillance system. (Jones 1992) Survey respondents were grouped into two categories: those who report AIDS cases "rarely or never" and those who report AIDS cases "usually or always." Forty-three percent of primary care providers and 72% of physician specialists stated they had served at least one patient with AIDS either sometime in the past or currently. "Non-reporters" were more likely than "reporters" to not report a case if they think it has already been reported in another state (78% vs 29%) and to indicate that they do not have responsibility for case reporting (63% vs. 23%). Among "non-reporters" and "reporters," there was no statistically significant difference in the percentage who expressed concerns about the case form being too long (26% vs. 18%), discrimination against patients (46% vs. 39%), confidentiality (61% vs 54%) or liability (32% vs. 34%).

A national survey of 4,223 physicians in the United States found that HIV/AIDS case reporting overall was low and that misperceptions exist regarding reporting requirements. For example, 30% of physicians said they were uncertain if their state required health care providers to report all AIDS cases. Among physicians with a past history of diagnosing AIDS only 53% said they reported AIDS cases "always" while 41% said they reported AIDS cases "never." When asked if they thought laboratories were reporting all HIV cases to the local health department, 77% of physicians agreed. (St. Lawrence 2002)

A survey of 345 physicians at two hospitals identified lack of knowledge of the reporting system as a major contributing factor in the underreporting of disease including lack of

knowledge of which diseases are reportable and procedures for reporting. (Konowitz 1984) When asked to estimate the percentage of cases they had reported based on a reference list of all reportable diseases, 36% of physicians said "none." On average, physicians estimated they had reported 28% of the cases they could remember. The two barriers most frequently cited by respondents were "did not know how to report" and "did not know it was a reportable disease." This demonstrates the importance of promoting ongoing awareness of the critical role of health care providers in public health surveillance and the benefits of reporting notifiable diseases.

A survey of 177 physicians in Georgia found that 52% of providers often did not report because they believed others would report. (McClean 2010) The authors identified a need for further research to improve understanding of health care provider characteristics and beliefs associated with levels of knowledge of the disease reporting system and completeness of reporting. Some characteristics are known, but opportunities remain for further describing other likely barriers and beliefs such as active engagement with public health. Many health care providers stated that reporting was too time-consuming and of low priority. Researchers identified a need for a more comprehensive approach to training and outreach that engages provider staff at all levels given the fact that disease reporting is often implemented by receptionists, social workers and other employees in administrative or support roles.

Patient-related factors

The public health importance of population-level factors is critical given their influence on the social context in which people make personal health decisions and their documented association with health outcomes. Examples of such social factors include the existence of historical and current racism, disparities in poor education, joblessness, racial profiling and the resulting

disparate rates of incarceration, the manner in which drugs and alcohol are marketed and low access to health care, to name a few. (Adimora, 2005; Farley, 2005) There is extensive evidence of disparities among racial and ethnic minorities in terms of negative health outcomes such as rates of sexually transmitted infections, AIDS-related opportunistic infections and lack of retention in care. This social context may exacerbate the strong multiple stigmas faced by men and women of color (Sayles, 2007) and informs our understanding of the existence of lower rates of risk factor ascertainment among non-White populations.

The stigma associated with anal sex, male-to-male transmission and injection drug use may contribute to the denial of such behavior by patients who initially self-report only vaginal sex with a heterosexual partner. Several studies have noted that patient self-reports often yield an underestimate of the proportion of respondents engaging in behavior other than vaginal sex with a heterosexual partner. (Brody, 1995a) Factors influencing patient non-reporting or under-reporting of taboo behavior may include difficulty recalling details, stigma, difficulty comprehending survey questions and the manner in which questions are asked, among others. Clearly, the fact that many patients are uncomfortable self-reporting stigmatized behavior is of concern because of the potential for inflating our understanding of the extent to which HIV/AIDS cases are attributable to heterosexual vaginal sex. In fact, Brody (1995b) stated that "intravenous and anal activities [are] the only clear vectors for HIV transmission" because of the pathology of HIV and the extent to which follow-up investigation often uncovers a history of anal sex and/or injection drug use in cases where such behavior was initially denied and only heterosexual vaginal sex was self-reported.

The use of non-human survey tools has resulted in improved data validity as such tools help assuage the stigma and discomfort associated with discussing anal sex, injection drug use

and male-to-male sexual contact. In a study of 1,268 sexually-active women, researchers noted that respondents were nine times more likely to self-report anal sex if the questionnaire was self-administered in private using an audio-assisted computerized interface, compared to questionnaires administered face-to-face by a human interviewer (OR 9.0, 95% CI 1.14-71.0). (Gross, 2000)

In the "Mode of Transmission Validation Study Group," researchers sought to determine if transmission categories other than heterosexual contact might be applicable to reported cases of HIV which initially only met the criteria for heterosexual contact. Follow-up, in-person investigation was conducted on 1,952 cases drawn from participating research sites in Alabama, California, Florida, New Jersey, New York City and Texas. Using active surveillance, chart reviews and other follow-up investigation activities, a significant number of cases originally reported as having only heterosexual risk were later found to be validated as having a risk other than heterosexual. Specifically, 24% of the male cases that were originally reported as heterosexual were found to meet the criteria for MSM (9%), IDU (12%) and MSM/IDU (2%) after follow-up investigation. For females, 13% of cases originally reported as only heterosexual were found to also meet the criteria for other transmission categories. (Klevens 1999) A similar study conducted in Florida involved follow-up investigation of a total of 168 heterosexually-acquired AIDS cases in Broward and Coastal Palm Beach Counties from January 1, 1989, to March 31, 1990. Through review of patient medical records and interviews, 50 of the 168 AIDS cases were reclassified. Among men, there were 40 cases that were reclassified as either MSM (19), IDU (18) or MSM/IDU (3). Among women, there were 10 cases that were reclassified as either IDU (7) or transfusion recipient (10). After reclassification, the percentage of AIDS cases attributed to heterosexual contact decreased among males (from 10% to 6%) and females (from

33% to 28%). (Nwanyanwu, 1993) These two studies illustrate the critical importance of identifying barriers such as stigma that may make some patients uncomfortable disclosing a history of injection drug use or male-to-male sexual contact. Clearly, collecting and reporting accurate risk history information is critical as errors in risk data may contribute to the appearance of false trends. (Espinoza, 2007)

System-related factors

Within surveillance systems, opportunities exist for improving the systems and processes through which providers receive training on disease reporting requirements. For example, a 1989 study of state reporting requirements for infectious disease and related conditions identified inconsistencies nationwide with regard to case definitions, time frames, persons required to report, methods by which reporting was accepted (such as online, fax or via telephone) and forms required. Lack of uniformity was cited as a possible factor contributing to the nationwide problem of incomplete reporting among health care providers. (Chorba 1989)

In addition, the total volume of cases requiring HIV/AIDS surveillance has grown at a rapid pace each year in part because of the increasing number of persons living with HIV/AIDS (due to improved health outcomes and a significant reduction in AIDS-related mortality rates) and a significant increase in the number of laboratories that automatically submit HIV/AIDS cases electronically. As a result, state health departments have experienced a significant increase in total HIV/AIDS surveillance case volume. (McDavid, 2006) This increase in volume has made it more difficult for state health departments to implement successful follow-up surveillance without increased funding to compensate for higher case volume.

What may be perhaps the most compelling system-related barrier to risk factor ascertainment involves the HIV/AIDS surveillance system itself. Several researchers and advocates have criticized the currently existing CDC hierarchy of transmission categories, illustrated in Figure 4. Critics suggest the hierarchy may discourage reporting of risk factor information because it masks HIV/AIDS cases that resulted from heterosexual contact. (Gollub, 2000; Haverkos, 2003) For example, in the existing hierarchy a heterosexual female with a history of injection drug use would be categorized as IDU even if she knew for a fact that she acquired HIV from a specific heterosexual partner and only used new needles which were never shared. Another hypothetical example involves a 35-year old male self-described as heterosexual with no history of injection drug use who contracted HIV from a female commercial sex worker at age 29. Even if this hypothetical male had sex with another male only once in his lifetime (say, at the age of 18 and never again thereafter) the hierarchy would still categorize his case as MSM, despite the fact that the true source of his infection was through heterosexual contact. In short, advocates argue the hierarchy is inherently flawed because it masks the magnitude of HIV/AIDS incidence and mortality among heterosexual men and women. Advocates for a new classification system have criticized the hierarchy as outdated, no longer meaningful and systematically biased against cases of heterosexual transmission as these are low in the hierarchy and defined restrictively. (Mokotoff, 2001; National Women and AIDS Collective, 2007)

Figure 4. Simplified illustration of CDC algorithm for determining "transmission category" for adult HIV/AIDS cases

	IF sex_male = YES AND idu = YES AND birth_sex = 1 or M	THEN	trans_categ = 03	(MSM&IDU)
ELSE	IF sex_male = YES AND birth_sex = 1 or M	THEN	trans_categ = 01	(MSM)
ELSE	IF idu = YES	THEN	trans_categ = 02	(IDU)
ELSE	IF received clotting factor AND birthdate before March 1985	THEN	trans_categ = 04	(Adult receipt of HIV-contaminated clotting factor)
ELSE	IF birth_sex = 1 or M AND (sex_idu = YES OR sex_hemo = YES OR sex_transfusion = YES OR sex_transplant = YES OR sex_hiv = YES)	THEN	trans_categ = 05	(High-risk heterosexual contact)
ELSE	IF birth_sex = 2 or F AND (sex_idu = YES OR sex_bisexual_male = YES OR sex_hemo = YES OR sex_transfusion = YES OR sex_transplant = YES OR sex_hiv = YES)	THEN	trans_categ = 05	(High-risk heterosexual contact)
ELSE	IF Before March 1985 received transfusion or transplant	THEN	trans_categ = 06	(Adult receipt of transfusion/ transplant)
ELSE	IF 12 months or more since case was initially reported	THEN	trans_categ = 09	(Adult with *No Identified Risk [NIR],* meaning 12 or more months have elapsed since date of initial report and case still does not meet any of the transmission category definitions)
ELSE →			trans_categ = 10	(Adult with *No Reported Risk [NRR],* meaning case does not meet any of the transmission category definitions but less than 12 months have elapsed since date of initial case report. Follow-up case investigation might possibly result in more risk factor information.)

Source: CDC. eHARS Technical Reference Guide, Version 3.2. Chapter 8, pp11-12. Atlanta, Georgia. October 2011.

Strategies for proactively improving risk factor reporting

Numerous examples exist of innovative strategies for improving the completeness of risk information provided on case report forms. These strategies can be considered "proactive" because they seek to minimize the need for active, follow-up surveillance by ensuring providers are aware of the importance of risk factor reporting up-front, thus ensuring minimal non-compliance. Many such proactive strategies also serve to improve the total number of notifiable disease cases ascertained as well as the timeliness of case reporting.

In 2003 the New York State Occupational Lung Disease Registry tested the effectiveness of three randomly-assigned message templates mailed to 368 health care providers with patients with lung disease (determined through hospital records) who had not reported the lung disease cases to the registry. The three message templates included one that emphasized the legal obligation to report, one that emphasized the public health benefits of compliance with reporting and a third template that emphasized both messages. Physicians that received both message types were more likely to provide complete and timely disease reports. (Brissette, 2006) An innovative campaign by the STD program of the Virginia Department of Health piloted a system of reimbursement for mailing costs. The reimbursement incentive was shown to improve the timeliness of local health departments' case reporting to the central STD office. (Vasiliu, 2009)

As part of proactive outreach and training of providers, the CDC recommends delivering to each reporting facility a quarterly statistical report describing the number and percentage of cases reported by that facility with missing risk factors. As a facility improves its percentage, the surveillance team could provide positive recognition for that site. Acknowledgment and praise could be in a newsletter or other surveillance publication, for example. Also, the surveillance

staff could proactively offer tailored technical assistance, training and outreach to facilities that do not experience satisfactory improvement. (CDC, 2005)

General outreach has often been cited as an effective proactive tool for improving risk factor acquisition. Outreach involves health departments building relationships and developing communication and training campaigns aimed at health care providers, laboratories, non-governmental organizations and other stakeholders to promote awareness of the requirements and benefits of notifiable disease reporting. This includes "Dear Colleague" letters and articles describing the importance of notifable disease reporting and risk factor acquisition in professional medical newsletters as well as messaging in brochures and fact sheets at events frequented by medical care providers. For example, in February 2012 the Commissioner of the Georgia Department of Public Health sent a "Dear Colleague" letter to all the health care providers in Georgia reminding them about notifiable disease reporting requirements. (Georgia DPH, 2012)

Also, in 2011 the HIV Prevention Unit of the Georgia Department of Public Health began requiring all of its HIV testing grantees to report all newly-diagnosed, confirmed HIV-positive cases to the health department within seven days. This requirement was instituted in the form of a written standard which was incorporated into all HIV testing contracts and affected eighteen non-profit organizations that receive a grant award to implement HIV testing on behalf of the Department. Justification for this requirement was made based on the fact that the Georgia notifiable disease surveillance law (OCGA §31-12-1) applies to health care providers and laboratories, which technically includes any organizations that implement a diagnostic test. (DPH, 2011) The contract language specifically states, "Each of the patient history questions in section 5 must be asked of all newly-diagnosed, confirmed HIV positive clients. Please go

through the entire patient history section with the patient and select a response for each line. If patient indicates 'unknown,' please make sure to mark that response. If reporting heterosexual contact, make sure all fields under the 'heterosexual relations' section are completed."

Strategies for obtaining risk factor information "after-the-fact"

For risk factor ascertainment, several aspects of active surveillance involve procedures implemented after a case form has already been submitted with missing risk factor information. These strategies can generally be referred to as reactive or "after-the-fact" strategies for risk factor acquisition and include three main options: 1) prioritization of cases for individualized follow-up investigation, 2) data matching and 3) integration of surveillance databases.

For active follow-up investigation of cases, the CDC recommends a strategic prioritization approach whereby staff prioritize newer cases and cases from the reporting facilities that have the highest number of NRRs. Guidelines published by CDC for state and territorial health departments emphasize the importance of conducting follow-up investigation of cases with missing risk factor information and recommend that health departments should conduct more outreach and training with the health care providers and reporting sites identified as submitting large numbers of NIR/NRR cases. The guidelines also set forth quality assurance metrics upon which surveillance systems should be evaluated, including the expectation that at least 85% of all reported cases of HIV (or a representative sample of those cases) must have risk factor information sufficient for meeting one of the CDC transmission category definitions. (CDC, 2005; McDavid, 2006)

As part of active surveillance, health departments may also employ data matching. This strategy involves gathering missing information from two or more data sources accessible to the

state health department. (Newman, 2009) Examples include databases for STD surveillance, HIV testing, HIV care and the AIDS Drug Assistance Program. However, the selection of which variables to match is critical because all datasets might not have the same data elements. In addition, some records may have missing data. The problem of erroneously-matched records is also a practical concern since matching information (such as name and date of birth) may be identical for two or more patients and unique identifiers (such as Social Security numbers) may not be collected in all databases. In general, the likelihood of an erroneous match can be decreased by increasing the number of variables required for a pair of records to be considered a match. However, this increased level of specificity comes at the expense of decreased sensitivity and generally yields fewer matched pairs.

Finally, the most sophisticated and complex level of data matching is database integration. This strategy involves creating a single database for multiple surveillance programs (such as STD, TB and HIV/AIDS) and has the added advantage of eliminating or minimizing duplication of data entry work. However, full integration of surveillance systems is hampered by the lack of compatibility among surveillance databases, differing definitions and values for variables in each dataset and the cost associated with creating and maintaining a new database. (Jennings, 2009)

The degree to which surveillance systems are integrated varies widely among health departments. In a survey mailed to the 58 Federally-funded local and state STD programs, 54% of respondents (n=47) used shared databases for AIDS and STD surveillance. Several challenges were identified including the existence of strict policies regarding limited access to HIV data and logistical challenges such as the procurement and maintenance of appropriate information technology systems. Limited time, capacity and expertise were cited as key challenges along

with issues of data compatibility and the existence of variables with differing definitions. (Dowell, 2009)

In Georgia, health care providers interested in submitting a notifiable disease case report currently face the challenge of maneuvring through a complex set of options depending on the type of disease being reported. The complexity of options is partly due to the fact that the eHARS database is dedicated exclusively to the reporting of HIV/AIDS cases but other cases are tracked using the Georgia State Electronic Notifiable Disease Surveillance System (SENDSS). A unique reporting form exists for HIV/AIDS cases but all other cases may be reported using a generic Georgia notifiable disease case report form.

A significant number of patients in the eHARS database have co-infection case information in the SENDSS database but these two databases are not integrated. As a result, public health employees are often required to check both databases if they need to conduct a case match as part of investigation of missing information such as patient contact information or risk factors. With grant funding from the CDC Care and Prevention in the United States (CAPUS) Demonstration Project, efforts are underway at DPH to integrate SENDSS and eHARS. In addition, the CAPUS grant will make it possible for online HIV/AIDS case reporting to be available to persons who are not employees of the DPH HIV/AIDS Surveillance Unit. When online HIV/AIDS case reporting is implemented in Georgia, DPH will have the ability to configure every aspect of the online case report form to facilitate accurate data entry. Questions can be defined as "mandatory" (so the user cannot advance forward without first providing an answer) or "optional." This feature could be useful for ensuring none of the patient risk factor questions are skipped.

Analysis and interpretation of case data with missing or incomplete risk factor information

Even after implementing the aforementioned types of proactive and reactive strategies, a significant number of HIV/AIDS cases remain coded in surveillance system databases as NRR or NIR. When these efforts are insufficient to yield risk factor data, public health agencies employ a broad range of classification methods for "counting" HIV/AIDS cases in the NRR or NIR categories. Each of these methods has advantages and limitations.

For example, in its nationally-published HIV surveillance reports the CDC deals with NIR/NRR cases by employing risk factor redistribution and multiple imputation. This involves redistribution of cases previously categorized as NIR/NRR for which follow-up investigation was either unsuccessful or not conducted using imputations that generally reflect trends observed in the distribution of risk category percentages from successful follow-up investigations. (CDC, 2005)

As part of the Supplement to HIV-AIDS Surveillance (SHAS) project, researchers implemented a behavioral survey with a random sample of HIV-positive patients age 18 and older who were reported to the eHARS database. (Buehler, 1996) This initiative provided information about the distribution of cases by risk category which has been useful for such multiple imputation purposes. (CDC, 2004) CDC continues to utilize data from the SHAS project for risk factor redistribution, but the data is now 10 years old because the SHAS project ended in 2004.

Researchers Lansky et al have proposed the use of discriminant function analysis as a method for classifying female NIR/NRR cases. Similar to multiple imputation, this method involves interviewing a representative sample of HIV-positive women drawn from the state HIV/AIDS database for whom a transmission category was already documented and a

representative sample of women from the same database with no identified risk in the case record. Follow-up case investigations were conducted with 1,297 women from 1993 to 1996. Women were asked to participate in a survey of demographic variables that extended beyond the core variables collected in the eHARS database. A multivariate logistic regression model was then created to identify demographic variables for the prediction of transmission category. The following variables were found to be the strongest predictors of transmission category: alcohol abuse, non-injection drug use and crack use. Other strong predictors include year of HIV/AIDS diagnosis, age, employment status and region of residence. (Lansky, 2001)

While these types of redistribution approaches could potentially yield significant cost savings, they assume found cases are representative of all cases. Many cases are hard to find, however, and found cases may differ greatly from those lost to follow-up in terms of important social and demographic variables. (Song, 2005; Lansky, 2001)

Also, the Council of State and Territorial Epidemiologists has made two data presentation recommendations aimed at reducing the number of cases categorized as NIR/NRR when multivariate imputation is not employed. These include the "presumed heterosexual" case definition and the use of mutually-exclusive HIV transmission risk classification categories in lieu of the hierarchy of most probable mode of transmission. These recommendations have been described elsewhere. (CSTE, 2007; Lee, 2003a; Lee, 2003b) Researchers applied both of these data presentation methods to AIDS cases from 1999 to 2001 from all fifty states and territories and HIV cases that had not progressed to AIDS in the national database from 29 states with mature HIV reporting systems from the same period. (Lee, 2003b) Neither method was successful at reducing substantially the percentage of cases classified as NIR/NRR. Specifically, of 9,532 NIR/NRR cases among males (32.4% of all male cases) only 545 met the criteria for the

"presumed heterosexual" definition. Also, of 14,668 NIR/NRR cases among females (43.3% of all female cases) only 450 met the criteria for the "presumed heterosexual" definition. Researchers showed the presumed heterosexual definition reduced the percentage of cases classified as NIR/NRR from 32.4% to 30.6% among males (reduction of 1.9 percentage points) and from 43.3% to 40.2% among females (reduction of 3.1 percentage points). This indicates the problem of high percentages of NIR/NRR cases is rooted in data acquisition and not data presentation.

However, when the presumed heterosexual definition was applied to newly-diagnosed cases of HIV/AIDS in Georgia, a larger percentage of cases met the definition. As shown in Table 1, there were 1,691 NIR/NRR cases among males (68% of all male cases) and 763 NIR/NRR cases among females (90% of all female cases) in 2010 in Georgia. Cases that met the criteria for the "presumed heterosexual" definition include 232 cases among males (9% of all male cases) and 113 cases among females (13% of all female cases). This analysis demonstrates use of the presumed heterosexual definition would reduce the percentage of cases classified as NIR/NRR from 68% to 59% among males (reduction of 9 percentage points) and from 90% to 77% among females (reduction of 13 percentage points) if the definition were applied to cases in Georgia in 2010. As shown in Table 2, similar reductions would occur for cases reported in 2011 if the presumed heterosexual definition were utilized as 8% of male cases and 11% of female cases would meet the presumed heterosexual definition. Of course, this option does not completely resolve the NIR/NRR problem because a significant proportion of cases (59% of male cases and 77% of female cases) remained in the NIR/NRR category even after applying the presumed heterosexual definition.

Table 1. Number of HIV cases diagnosed and reported in 2010 in Georgia by exposure category: Comparison of two data presentation options

Data presented using currently-existing CDC exposure category definitions			Data presented using proposed "presumed heterosexual" exposure category definition described in Lee et al (2003)[1]		
Males			Males		
Male-to-male sexual contact (MSM)	712	29%	Male-to-male sexual contact (MSM)	712	29%
Injection drug use (IDU)	19	1%	Injection drug use (IDU)	19	1%
MSM/IDU	17	1%	MSM/IDU	17	1%
Heterosexual[2]	35	1%	Heterosexual[2]	35	1%
NIR/NRR[3]	1,691	68%	NIR/NRR[3]	1,459	59%
			Presumed heterosexual[4]	232	9%
Total male	2,474	100%	Total male	2,474	100%
Females			Females		
Injection drug use (IDU)	11	1%	Injection drug use (IDU)	11	1%
Heterosexual[2]	72	9%	Heterosexual[2]	72	9%
NIR/NRR[3]	763	90%	NIR/NRR[3]	650	77%
			Presumed heterosexual[4]	113	13%
Total female	846	100%	Total female	846	100%

Source: Data request submitted to Georgia Department of Public Health, February 6, 2013.

Note: Figures are not shown for cases of perinatal transmission (fewer than 5) and cases with unknown sex. Number of newly diagnosed HIVinfections are based on data entered through December 31, 2012 and have not been adjusted for reporting delays. Newly diagnosed cases are based on a residence of diagnosis in Georgia. Case counts include incarcerated persons who may artificially inflate the numbers.

[1] Lee, L. M., McKenna, M. T., & Janssen, R. S. (2003). Classification of transmission risk in the national HIV/AIDS surveillance system. Public Health Rep, 118(5), 400-407.

[2] Heterosexual is defined as sexual contact with a person known to have HIV infection or at least with a person at increased risk of HIV infection (based on a history of MSM, IDU, or receipt of blood products).

[3] NIR/NRR: Cases are categorized as "No risk reported" (NRR) if risk information was absent from the initial case report because the information had not been reported by the reporting source, had not been sought, or had not been found by the time the case was reported. Cases may remain NRR until epidemiologic follow-up has been completed and potential risks (exposures) have been identified. If risk has not been identified within 12 months of being reported as NRR, the case may be considered "No identified risk" (NIR) which means epidemiologic follow-up has been conducted, sources of data have been reviewed (which may include an interview with the patient or provider) and no mode of exposure has been identified. All NRR cases are classified as NIR if they continue to have no reported risk 12 or more months after the date of initial diagnosis.

[4] Presumed heterosexual females are those with no known exposure category, who have history of a sexual contact with a male and no history of injection drug use. Presumed heterosexual males are those with no known exposure category who have a history of sexual contact with a female.

Table 2. Number of HIV cases diagnosed and reported in 2011 in Georgia by exposure category: Comparison of two data presentation options

Data presented using currently-existing CDC exposure category definitions			Data presented using proposed "presumed heterosexual" exposure category definition described in Lee et al (2003)[1]		
Males			Males		
Male-to-male sexual contact (MSM)	777	27%	Male-to-male sexual contact (MSM)	777	27%
Injection drug use (IDU)	22	1%	Injection drug use (IDU)	22	1%
MSM/IDU	10	0%	MSM/IDU	10	0%
Heterosexual[2]	35	1%	Heterosexual[2]	35	1%
NIR/NRR[3]	2,002	70%	NIR/NRR[3]	1,778	62%
			Presumed heterosexual[4]	224	8%
Total male	2,846	100%	Total male	2,846	100%
Females			Females		
Injection drug use (IDU)	18	2%	Injection drug use (IDU)	18	2%
Heterosexual[2]	54	6%	Heterosexual[2]	54	6%
NIR/NRR[3]	787	92%	NIR/NRR[3]	692	81%
			Presumed heterosexual[4]	95	11%
Total female	859	100%	Total female	859	100%

Source: Data request submitted to Georgia Department of Public Health, February 6, 2013.

Note: Figures are not shown for cases of perinatal transmission (fewer than 5) and cases with unknown sex. Number of newly diagnosed HIVinfections are based on data entered through December 31, 2012 and have not been adjusted for reporting delays. Newly diagnosed cases are based on a residence of diagnosis in Georgia. Case counts include incarcerated persons who may artificially inflate the numbers.

[1] Lee, L. M., McKenna, M. T., & Janssen, R. S. (2003). Classification of transmission risk in the national HIV/AIDS surveillance system. Public Health Rep, 118(5), 400-407.

[2] Heterosexual is defined as sexual contact with a person known to have HIV infection or at least with a person at increased risk of HIV infection (based on a history of MSM, IDU, or receipt of blood products).

[3] NIR/NRR: Cases are categorized as "No risk reported" (NRR) if risk information was absent from the initial case report because the information had not been reported by the reporting source, had not been sought, or had not been found by the time the case was reported. Cases may remain NRR until epidemiologic follow-up has been completed and potential risks (exposures) have been identified. If risk has not been identified within 12 months of being reported as NRR, the case may be considered "No identified risk" (NIR) which means epidemiologic follow-up has been conducted, sources of data have been reviewed (which may include an interview with the patient or provider) and no mode of exposure has been identified. All NRR cases are classified as NIR if they continue to have no reported risk 12 or more months after the date of initial diagnosis.

[4] Presumed heterosexual females are those with no known exposure category, who have history of a sexual contact with a male and no history of injection drug use. Presumed heterosexual males are those with no known exposure category who have a history of sexual contact with a female.

3. Key Informant Interviews

3.1 Introduction

The following employees of the DPH HIV/AIDS Surveillance Unit were interviewed using the discussion guide shown in Attachment #1.

- Jane Kelly, MD, HIV Surveillance Director
- Roderiques Lambert, Surveillance Coordinator
- Marguerite Camp
- Akilah Spratling
- LaToya Moss
- Eugene Pennisi

Interviews were conducted at the HIV/AIDS Surveillance Unit main offices at 2 Peachtree Street NW in downtown Atlanta. Written notes from the discussions were analyzed for recurring themes and comments were summarized and organized as shown in section 3.2 ("Overall Key Findings") of this report.

3.2 Overall key findings

Reasons for the high percentage of cases with missing or incomplete risk factor information

1. Some providers may be averse to case reporting and/or risk factor reporting because they perceive the process as too time-consuming. Others may have concerns about privacy and confidentiality.
2. Medical schools might not emphasize skills for patient risk assessment interviewing.
3. The instructions on the form might not be as clear as they could be.
4. It is not clear what the option "unknown" means (in the risk history section).
5. In the hierarchy of transmission categories, it is more difficult for a case to meet the criteria for the "heterosexual contact" definition because at least two variables must have

an affirmative “yes” value: 1) sex with an opposite-sex partner who specifically 2) has an HIV risk factor (as shown in Figure 5). The case definition for other transmission categories (such as MSM and IDU) is much more sensitive because it only requires one affirmative “yes” value: sex with another male or use of injection drugs.

6. If a patient reports sex with an opposite-sex partner, the patient is asked about the HIV-related risk behaviors of that partner. However, because such behaviors are taboo the patient often cannot confirm or deny their existence. One key informant recalled a female patient who stated, “You want to know if my boyfriend is bisexual, injects drugs or has HIV. I don’t think so. I could ask him, but no man is going to admit those things to his wife or girlfriend even if they are true -- and I’m not a detective. Plus, what do you think will go through his mind if I ask him those things? Don’t you think he will get angry at me, like I’m accusing him of something? And don’t you think he will wonder if I have HIV?”
7. Some providers might not be aware of the importance of getting a "yes or no" answer to the heterosexual contact questions. Some providers might think an "unknown" answer is acceptable, when in reality the "unknown" option should be discouraged.
8. Some providers might not be aware of the many intricate aspects of notifiable disease reporting in Georgia such as the different types of forms required for various diseases and the requirement that all HIV/AIDS cases must be reported within 7 days.
9. There was a general consensus that medical care providers in the public health sector are more likely to submit complete case reports within 7 days (compared to those who work in theprivate sector) in part because of the network of public health Communicable Disease Specialists. Some providers in the private sector might not be as closely

connected with the public health sector to understand fully the benefits and complex requirements of notifiable disease reporting.

10. Some patients who are MSM or who inject drugs might feel uncomfortable reporting these highly stigmatized behaviors.

Recommendations for improving risk factor ascertainment

1. Scale-up active surveillance efforts.
2. Scale-up provider outreach and education efforts.
3. On a regular basis, generate a series of custom-tailored progress reports for facilities that report the highest number of HIV/AIDS cases. Reports should include performance measures such as percentage of cases with complete risk factor information and percentage of cases reported within 7 days. Also consider perhaps including benchmark figures for comparison, against which each site may compare its performance.
4. Improve the layout of the 13 risk factor questions in the "Risk History" section of the form. Figure 5 provides a proposed alternative layout with clearer instructions.
5. Currently, the Georgia HIV/AIDS case report form and its detailed step-by-step instructions are in two separate files. Consider creating one single file containing both the form and the instructions section. This way, anyone who downloads the form will always have prompt access to the detailed step-by-step instructions.
6. In the instructions, explain that the risk history questions should be asked of the patient. Providers should avoid trying to answer the questions independently of the patient. In most cases, the patient should be able to give a "yes or no" answer to the risk history questions. The instructions for the form should also describe situations in which a

provider is allowed to answer the questions using information that was not obtained directly from the patient. For example, a patient who denies having a past history of drug injection might have visible evidence of injection drug use. Medical care providers may be trained to detect such evidence including needle puncture wounds, needle track marks, skin lesions, granulomas, scars, atrophy, abscesses, and semi-linear pigment changes.

7. Some providers might interpret the "unknown" answer option as "Patient does not know the answer." This should be avoided. If the patient does not know the answer, the provider should ask the patient to give his or her best guess for what is the most likely answer (either "yes" or "no") and go with that answer.
8. For the heterosexual risk factor questions, instead of "yes or no" consider having a "likelihood scale" such as "definitely yes, probably yes, probably no, definitely no." The heterosexual risk factor questions are highly sensitive because they require the patient to make assumptions about their sex partners' behaviors. If a female patient is asked, "Did you ever have heterosexual relations with a bisexual male?" she might feel less discomfort saying "probably yes" as opposed to "Yes." This is because "yes" might imply 100% certainty.
9. Promote awareness of proven strategies for obtaining valid patient information in risk assessment interviews. Disseminate this information to the medical community through distance learning opportunities, recorded (archived) webinars, YouTube videos, conference presentations, printed newsletter articles, "Dear Colleague" letters, quick-reference guides, flyers and posters for display in healthcare settings.
10. Continue collaborating with the HIV Prevention Unit to ensure all funded HIV counseling and testing sites are implementing its HIV/AIDS case reporting requirements.

11. Establish a new collaboration with the Fulton County Department of Health and Wellness to ensure the HIV counseling and testing sites they fund are also complying with HIV/AIDS case reporting requirements.
12. As part of the new CAPUS initiative, online HIV/AIDS case reporting will soon be launched in Georgia. DPH should configure the patient risk history questions as "mandatory" (so the user cannot advance forward without first providing an answer) to ensure none of the questions are skipped or left blank.

Figure 5. Proposed alternative layout for "Patient Risk History" section of the Georgia HIV/AIDS case report form.

CURRENT LAYOUT

☐Female to Male

V. PATIENT HISTORY—*Complete ALL fields*

BEFORE the first positive HIV test or diagnosis, patient EVER had:	Yes	No	Unk
Sex with male			
Sex with female			
Injected drugs			
Received clotting factor			
HETEROsexual relations with the following:			
Injection drug user (IDU)			
Bisexual male **(applies to females only)**			
Person with hemophilia/ coagulation disorder			
Transfusion recipient w/ documented HIV infection			
Person with AIDS or documented HIV infection, risk unspecified			
Received transfusion: Date 1st / Last: /			
Received organ transplant, tissue or artificial insemination			
Worked in healthcare/clinical laboratory If yes, **SPECIFY OCCUPATION:**			
Was patient infected perinatally?			

PROPOSED NEW LAYOUT
With improved instructions

V. PATIENT HISTORY – *Please ask patient all 13 questions*

Note: Please see instructions packet for interview suggestions to help increase the likelihood of obtaining accurate information from patients.

Ask patient: Before your first positive HIV test or diagnosis, did you …	Yes	No
1) …ever have sex with a male?		
2) …ever have sex with a female?		
3) …ever inject drugs?		
4) …ever received clotting factor?		
5) …ever received a blood transfusion? Date 1st ___/_____ Last: ___/_____		
6) …ever received an organ transplant, tissue or artificial insemination?		
7) …ever work in healthcare or clinical lab? If yes, occupation:		
8) Was patient infected perinatally?		

Ask patient: Before your first positive HIV test or diagnosis, did you ever have heterosexual relations with…	Definitely no	Probably no	Probably yes	Definitely yes	Don't know
9) …an injection drug user?					
10) …a bisexual male? (ask females only)					
11) …a person with hemophilia/ coagulation disorder?					
12) …a transfusion recipient with documented HIV infection?					
13) …a person with AIDS or documented HIV infection, risk unspecified?					

4. Methodology

4.1. Introduction

There has been only one previous study identified in the literature which examined eHARS database variables associated with missing or incomplete risk factor using crude odds ratios. In 2005, researchers McDavid and Kajese analyzed HIV cases reported to the national surveillance system from 1994 to 2003 from 32 states with name-based, confidential HIV reporting. The researchers identified patient age, race/ethnicity and sex as variables associated with missing or incomplete risk factor information. (McDavid, 2005)

This study will build upon research conducted by McDavid and Kajese by calculating odds ratios from logistic regression considering multiple predictor variables simultaneously. The APIDS dataset will be utilized, which contains 16 de-identified variables for 859,000 AIDS cases reported between 1981 and 2002 as part of the nationwide system of HIV/AIDS surveillance. No subsequent updates to the dataset have been made since 2002.

4.2. Population and sample

Only adult and adolescent cases from the APIDS dataset (ages 13 and older) were included in this analysis because cases among patients younger than age 13 have typically been associated with a narrow set of possible risk factors (such as mother-to-child transmission) in follow-up investigation. In addition, patients younger than age 13 comprised a relatively small portion of the APIDS dataset (approximately 1%). Observations with a race value of "unknown" were also excluded. Finally, the data was filtered to include only its most recent year (2002) based on the assumption that recent data should resemble the current state of the problem better than older

data. Cases removed included 9,220 younger than 13 years of age; 642 with race unknown and 833,903 not diagnosed in 2002. After applying these three analysis requirements to the dataset a total of 833,978 cases were removed, leaving a total of 25,022 cases for analysis. Category counts do not sum to total removed because categories were not mutually exclusive. Cases were not adjusted for reporting delays because the *adjwgt* weight variable is not reliable for cases diagnosed during the most recent 6 to 9 months of the dataset. (APIDS Manual, Section 1 -- Delay in Reporting).

4.3. Data analysis procedures

A copy of the CDC APIDS dataset was downloaded from the following web address.

http://www.cdc.gov/hiv/topics/surveillance/resources/software/apids/index.htm

The dataset was imported into SAS 9.2 and dataset contents were examined coarsely using the accompanying codebook and *proc contents*. (See Attachment #4 for a copy of the codebook.) An outcome variable called *NIR* was created and each case was assigned an *NIR* value of either "0" (not NIR) or "1" (NIR) depending on whether or not the case met the criteria for a specific CDC-defined transmission category. Specifically, a case was considered "no identified risk" if the "HIV transmission category" variable *(TRANSCAT)* had value of "8" (adult/adolescent, risk not reported or identified).

Variables in the APIDS dataset were selected for inclusion in this analysis primarily based on a review of the literature which identified patient age, race/ethnicity and sex as variables associated with missing or incomplete risk factor information. (McDavid, 2005) These three variables are part of the APIDS dataset and will be included in this analysis.

In addition, the APIDS dataset variable *MSA* describes whether or not a case was reported in a metropolitan statistical area (MSA) with a population of 500,000 persons or greater and the variable *BIRTH* describes whether or not the patient was born in the United States. The *MSA* variable was included in this analysis as a possible proxy measure of urbanity because greater HIV-related stigma has been observed among patients in non-urban areas compared to patients in urban areas. (Hudson, 2001; Moneyham, 1996; Moneyham, 2000) Also, the *BIRTH* variable was included because higher rates of HIV-related stigma have been observed among foreign-born patients with HIV, compared to US-born patients. (Ojikutu, 2013) For this study, HIV-related stigma is of concern because discomfort with HIV-related topics may be a barrier to patient-provider communication. (Mayer, 2004) Clearly, the inability of patients and their health care providers to communicate in an open, honest and frank manner would hamper the ability to obtain and report accurate patient risk factor information.

For each independent variable, the proportion of cases classified as NIR was calculated in bivariate analysis with a Chi-square test of significance. Next, a backward elimination regression model was created in SAS 9.2 with NIR as the outcome of interest and statistical significance at alpha=0.05 for variable entry and retention in the model. Finally, a logistic regression equation was created to express the probability of being classified as NIR in terms of the APIDS variables of interest.

To complement the regression analysis, subject matter experts and key informants were interviewed including employees of the DPH HIV surveillance section, employees of the DPH HIV prevention unit and four health care providers.

4.4. Limitations and delimitations

The APIDS dataset provides the convenience of quick access to a large number of case observations but this dataset is limited because it does not provide data on non-AIDS cases of HIV. This is a significant limitation since the challenge of risk factor ascertainment is more severe for HIV cases compared to cases of AIDS. (Glynn, 2007; McDavid, 2005) Also, the dataset does not provide information about important patient demographic variables identified in the literature review such as poverty status and educational attainment. The literature review also identified other factors potentially associated with incomplete reporting of patient risk factor data including health care provider characteristics (such as amount of training in disease reporting and how the risk assessment questions are asked) but these types of variables are not part of the APIDS dataset. In addition, the APIDS dataset includes an unknown number of non-NIR cases which were initially reported with insufficient risk factor data ("no risk reported" or "NRR") but subsequently met the criteria for a transmission category after health department employees successfully elicited risk factor information during follow-up case investigation. As a result, the proportion of cases categorized as NIR in the APIDS dataset understates the true magnitude of the problem of incomplete reporting of patient risk factor information. It should also be noted that calculation of an adjusted odds ratio from exponentiation of the logistic regression coefficient is best suited for rare outcomes such as less than 20% because the adjusted odds ratio exaggerates the estimate of the magnitude of association with common outcomes. (Szklo, 2007) Finally, the APIDS dataset does not provide data more recent than 2002.

5. Results

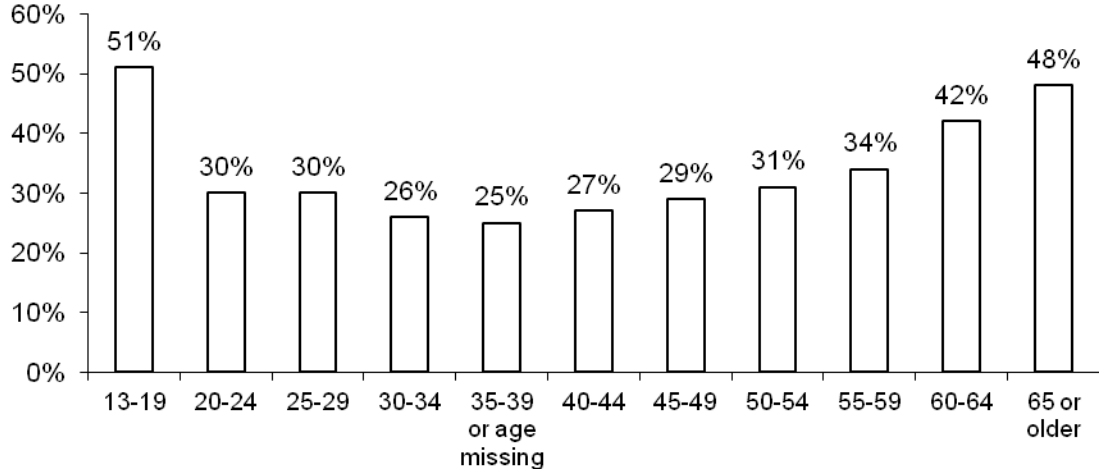
5.1. Findings

A total of 25,022 unweighted observations of adult and adolescent AIDS cases reported in 2002 were included in this retrospective study. Table 1 describes selected characteristics of the sample. Of the total, 17,829 patients were classified into a CDC-defined transmission category while the remaining 7,193 patients were classified as "no identified risk" (29% NIR, 71% nor NIR). Notably, non-Hispanic Blacks comprised the largest racial/ethnic group (12,882 patients; 51% of the sample) followed by non-Hispanic Whites (7,225 patients; 29% of the sample) and Hispanics (4,554 patients, 18% of the sample). Among non-Hispanic Blacks, 35% were classified as NIR compared to 19% of non-Hispanic Whites and 27% of Hispanics.

Males comprised 74% of the sample and had a lower NIR percentage (26% NIR) compared to females (37% NIR). Persons born in the United States comprised 86% of the sample and had a lower NIR percentage (28% NIR) compared to persons born outside the United States (31% NIR). Persons residing in a large metropolitan statistical area (defined as having a population of 500,000 or greater) comprised 79% of the sample and had a lower NIR percentage (28% NIR) compared to persons who resided outside a large MSA (31% NIR).

The two largest age groups were 35-39 (5,434 cases) and 40-44 (4,971 cases) which together comprised 42% of the sample. By age group, the percentage of patients classified as NIR generally followed an asymmetrical upward-opening curve (as shown in Figure 6) which fell from a starting point of 51% for the 13-19 age group to 25% for the 35-39 age group, after which the NIR percentage gradually increased to 48% for the 65 and older age group. The shape of this curve is similar to that observed by researchers McDavid et al.

Figure 6. Among adult and adolescent AIDS cases diagnosed in 2002: Percentage classified as "No Identified Risk" (NIR), by age group – APIDS dataset



In both forward stepwise and backward elimination logistic regression, all five independent variables (race, sex, age group, residence in a large MSA and birth abroad) were retained at alpha=0.05 as predictors significantly and independently associated with an NIR outcome for missing or incomplete risk factor information (p <0.0001, area under the ROC curve, 62.5%).

Table 2 presents a summary of results obtained for the fitted model including logistic regression coefficients, adjusted odds ratios, 95% Wald confidence intervals and p values. Referent groups are designated with an odds ratio of 1.00 and include non-Hispanic Whites, patients age 35-39, males, patients born in the United States and patients with residence outside a large MSA.

For almost every parameter, statistically significant differences (p <0.0001) were observed at the 95% confidence level. Of particular interest were odds ratios for an NIR/NRR outcome greater than 2.00 which include those for non-Hispanic Black patients (2.1 OR, 95% CI 2.0-2.3), patients age 13-19 (2.6 OR, 95% CI 2.0-3.4), patients age 60-64 (2.1 OR, 95% CI 1.8-

2.5) and patients age 65 or older (2.7 OR, 95% CI 2.2-3.3). This suggests these factors make it more than twice as likely a case will be in the NIR/NRR category when all other factors are held constant. Odds ratios were also statistically significant and higher than 1.0 (albeit not greater than 2.0) for females, Hispanics, patients who do not reside in a large MSA and patients born outside the United States.

In model evaluation, the three statistical tests shown in Table 5 suggest the model offers improved prediction of an NIR/NRR outcome compared to the baseline intercept-only approach. Results indicate the overall model is significant at the 0.05 level according to the model chi-square statistics for the likelihood ratio test, the score test and the Wald test (df=17, p <0.0001). For assessment of model fitness, statistics for the Hosmer-Lemeshow test, the Cox and Snell R^2 index and the Nagelkerke R^2 index are provided in Table 5. The Hosmer-Lemeshow test provided a statistically significant value (DF=8, p<0.0001) which would normally disprove the null hypothesis of a good model fit to the data. However, the Hosmer-Lemeshow test gives increasing significance to small differences as sample size increases. Because of the large size of this sample, this statistic is interpreted with caution. (Kramer, 2007) The insignificant values for the Cox and Snell R^2 index (0.0402) and the Nagelkerke R^2 index (0.0575) suggest the null hypothesis of good model fitness is tenable. The model appears to fit the data well.

The Goodman-Kruskal Gamma statistic (0.254) and Sommer's D (0.25) suggest that if the model is utilized to predict whether or not a case will be in the NIR/NRR category, the number of false predictions would be approximately 25% fewer than prediction by chance alone. Both are based on the Tau-a statistic with adjustments made for the existence of ties on outcomes and predicted probabilities. (Peng, 2002)

The *c* statistic (0.625) is calculated by examining all the possible discordant case pairs in the dataset (pairs in which one case has a value of NIR=0 and the other case has a value of NIR=1) and then determining the percentage of such pairs (62.5% in this case) for which the model correctly assigned the higher probability to cases with a value of NIR=1. The lowest possible value for the *c* statistic is 0.50 which would mean the model correctly predicted the outcome only 50% of the time or no better than by chance alone.

Table 6 provides a comparison of the observed (actual) and model-predicted frequencies for the NIR outcome variable. This table displays all 25,022 AIDS cases in terms of whether or not they were actually NIR and whether or not the model predicted them to be NIR using the default 0.50 cutoff point. A predicted probability equal to or greater than 0.50 is interpreted as a prediction of NIR=1 and vice-versa. Correct predictions are displayed in slightly larger, bold font. As shown in Table 6, the model had a high degree of specificity but poor sensitivity. In other words, the model gave more accurate predictions for AIDS cases which actually did have sufficient risk factor information (NIR=0; 98.63% of such non-event observations predicted correctly), compared to cases which had insufficient risk factor information (NIR=1; 3.84% of such event observations predicted correctly). The rate of false positive predictions (number of AIDS cases incorrectly predicted to be NIR, among all AIDS cases predicted to be NIR) was 47.02%. The rate of false negative predictions (number of AIDS cases incorrectly predicted to be *not* NIR, among all AIDS cases predicted to be *not* NIR) was 28.23%. Overall, the model provided correct predictions for 71.38% of the 25,022 AIDS cases in the sample, which is an improvement compared to the 50% level of accuracy that would be expected from chance alone. The most optimal cutoff point for this model (0.60) was obtained by graphing sensitivity and specificity at various cutoff points. This would result in maximum improved overall accuracy,

sensitivity, specificity, false positives and false negatives but statistics at this cutoff are not displayed because the purpose of this model does not involve prediction or diagnosis.

Table 7 provides predicted probabilities for ten hypothetical case scenarios based on the model's five predictor variables sorted from lowest to highest predicted probability. In the first scenario, the model gave a relatively low predicted probability (0.1520; 95%CI: 0.1415 - 0.1631) that an AIDS case would have missing risk factor information for a middle-aged non-Hispanic White male born in the United States who resides in a large MSA. The last scenario in the table shows that for an African-American teenage female born outside the United States who does not reside in a large MSA, the predicted probability than an AIDS case would have missing risk factor information is much higher (0.6739; 95%CI: 0.6168 - 0.7345).

Table 3. Number of AIDS cases diagnosed in 2002 among adults and adolescents and percentage classified as "No Identified Risk" (NIR), by selected characteristics -- APIDS dataset

	No.	Percent NIR
Racial/ethnic group		
White (not Hispanic)	7,225	19%
Black (not Hispanic)	12,882	35%
Hispanic	4,554	27%
Asian/Pacific Islander	243	30%
American Indian/Alaska Native	118	11%
Sex		
Male	18,572	26%
Female	6,450	37%
Country of birth		
United States	21,469	28%
Other country	3,553	31%
Type of area of residence		
Resides in a large MSA*	19,755	28%
Does not reside in a large MSA*	5,267	31%
Age group†		
13-19	214	51%
20-24	914	30%
25-29	2,100	30%
30-34	3,784	26%
35-39	5,434	25%
40-44	4,971	27%
45-49	3,443	29%
50-54	2,065	31%
55-59	1,064	34%
60-64	568	42%
65 or older	465	48%
Total	25,022	29%

*Metropolitan Statistical Area with a population of 500,000 or greater.
†Persons for whom age was missing are included in the 35-39 age category in the APIDS dataset.

Table 4. Logistic regression analysis of the association between various sociodemographic characteristics and classification as "No Identified Risk" (NIR): AIDS cases reported in 2002 among adults and adolescents -- APIDS dataset

Characteristic	Logistic regression coefficient*	Adjusted odds ratio	95% CI†	Pr > ChiSq
Intercept	-1.5276			
Racial/ethnic group				
White (not Hispanic)	--	1.0		
Black (not Hispanic)	0.7496	2.1	2.0 - 2.3	<0.0001
Hispanic	0.3588	1.4	1.3 - 1.6	<0.0001
Asian/Pacific Islander	0.5290	1.7	1.3 - 2.3	0.0003
American Indian/Alaskan Native	-0.6798	0.5	0.3 - 0.9	0.0219
Age group‡				
13-19	0.9547	2.6	2.0 - 3.4	<0.0001
20-24	0.1134	1.1	1.0 - 1.3	0.1555
25-29	0.1403	1.2	1.0 - 1.3	0.0157
30-34	0.0249	1.0	0.9 - 1.1	0.6135
35-39	--	1.0		
40-44	0.1143	1.1	1.0 - 1.2	0.0116
45-49	0.1600	1.2	1.1 - 1.3	0.0013
50-54	0.3057	1.4	1.2 - 1.5	<0.0001
55-59	0.4584	1.6	1.4 - 1.8	<0.0001
60-64	0.7507	2.1	1.8 - 2.5	<0.0001
65 or older	0.9930	2.7	2.2 - 3.3	<0.0001
Sex				
Male	--	1.0		
Female	0.3805	1.5	1.4 - 1.6	<0.0001
Country of birth				
United States	--	1.0		
Other country	0.1684	1.2	1.1 - 1.3	0.0003
Type of area of residence				
Resides in a large MSA¶	-0.1917	0.8	0.8 - 0.9	<0.0001
Does not reside in a large MSA¶	--	1.0		

*Logistic regression coefficients calculated using *proc logistic* in SAS 9.2 with backward elimination.
†95% Wald confidence interval
‡Persons for whom age was missing are included in the 35-39 age category in the APIDS dataset.
¶Metropolitan Statistical Area with a population of 500,000 or greater.
Odds ratios are rounded to the nearest tenth of

Table 5. Evaluation of logistic regression model for the association between various sociodemographic characteristics and classification as "No Identified Risk" (NIR): AIDS cases reported in 2002 among adults and adolescents -- APIDS dataset

Test	ChiSq	*df*	Pr > ChiSq
Overall model evaluation			
Likelihood ratio test	1026.6469	17	<0.0001
Score test	1016.0430	17	<0.0001
Wald test	965.9380	17	<0.0001
Goodness-of-fit test			
Hosmer & Lemeshow	46.0995	8	<0.0001

Cox and Snell R^2= 0.0402. Nagelkerek R^2 (Max rescaled R^2)= 0.0575. Kendall's Tau-*a*= 0.102. Goodman-Kruskal Gamma= 0.254. Somer's D_{xy}= 0.25. *c*-statistic= 0.625.

Table 6. Observed and predicted frequencies for classification as "No Identified Risk" (NIR) in a logistic regression model with a 0.50 prediction cuttoff: AIDS cases reported in 2002 among adults and adolescents -- APIDS dataset

	Predicted		
Observed	Not NIR	NIR	% Correct
Not NIR(NIR=0)	**17,584**	245	98.63%
NIR (NIR=1)	6,917	**276**	3.84%
Overall % correct			71.38%

Sensitivity: 276/(6,917+276)%= 3.84%. Specificity: 17,584/(245+17,584)%= 98.63%.
False positive %: 245/(245+276)= 47.02%. False negative %: 6,917/(17,584+6,917)= 28.23%.

Table 7. Cumulative predicted probability of being classified as "No Identified Risk" (NIR) by logistic regression for hypothetical case examples: AIDS cases reported in 2002 among adults and adolescents -- APIDS dataset

Num	Race	Sex	Age group	Born outside United States	Resides in a large MSA	Intercept	Predicted probability of being classified as NIR	
							estimated probability	95% confidence interval
1	White (β=0)	Male (β=0)	35-39 (β=0)	No (β=0)	Yes (β=-0.1917)	-1.5276	0.1520	0.1415-0.1631
2	White (β=0)	Female (β=0.3805)	35-39 (β=0)	No (β=0)	Yes (β=-0.1917)	-1.5276	0.2078	0.1921-0.2243
3	Hispanic (β=0.3588)	Male (β=0)	35-39 (β=0)	Yes (β=0.1684)	Yes (β=-0.1917)	-1.5276	0.2329	0.2158-0.2510
4	Black (β=0.7496)	Male (β=0)	35-39 (β=0)	No (β=0)	Yes (β=-0.1917)	-1.5276	0.2750	0.2608-0.2895
5	Black (β=0.7496)	Female (β=0.3805)	35-39 (β=0)	No (β=0)	Yes (β=-0.1917)	-1.5276	0.3569	0.3387-0.3754
6	Hispanic (β=0.3588)	Male (β=0)	13-19 (β=0.9547)	Yes (β=0.1684)	Yes (β=-0.1917)	-1.5276	0.4409	0.3725-0.5118
7	Black (β=0.7496)	Male (β=0)	13-19 (β=0.9547)	No (β=0)	Yes (β=-0.1917)	-1.5276	0.4963	0.4280-0.5646
8	Black (β=0.7496)	Male (β=0)	13-19 (β=0.9547)	No (β=0)	No (β=0)	-1.5276	0.5441	0.4744-0.6121
9	Black (β=0.7496)	Female (β=0.3805)	13-19 (β=0.9547)	No (β=0)	Yes (β=-0.1917)	-1.5276	0.5904	0.5227-0.6548
10	Black (β=0.7496)	Female (β=0.3805)	13-19 (β=0.9547)	Yes (β=0.1684)	No (β=0)	-1.5276	0.6739	0.6168-0.7345

Note: Cumulative predicted probabilities were calculated using the following predicted probability equation, which was created based on a logistic regression model. For each of the above hypothetical scenarios, X_1 to X_{17} are the independent predictor variables shown below with possible values of 0 (for "no") or 1 (for "yes") for each variable. The logit was converted to a probability using $p=(e^{logit})/(1+e^{logit})$.

$$\ln\left(\frac{P(NIR=1)}{P(NIR=0)}\right)=\left(\begin{array}{llllll}\beta_0 & +\beta_1x_1 & +\beta_2x_2 & +\beta_3x_3 & +\beta_4x_4 & +\beta_5x_5 \quad +\beta_6x_6\\ & +\beta_7x_7 & +\beta_8x_8 & +\beta_9x_9 & +\beta_{10}x_{10} & +\beta_{11}x_{11} \quad +\beta_{12}x_{12}\\ & +\beta_{13}x_{13} & +\beta_{14}x_{14} & +\beta_{15}x_{15} & +\beta_{16}x_{16} & +\beta_{17}x_{17}\end{array}\right)$$

$$\ln\left(\frac{P(NIR=1)}{P(NIR=0)}\right)=\left(\begin{array}{l}-1.5276 + 0.7496x_1 + 0.3588x_2 + 0.5290x_3 + (-0.6798)x_4 + 0.9547x_5 + 0.1134x_6\\ \quad + 0.1403x_7 + 0.0249x_8 + 0.1143x_9 + 0.1600_0x_{10} + 0.3057x_{11} + 0.4584x_{12}\\ \quad + 0.7507x_{13} + 0.9930x_{14} + 0.3805x_{15} + 0.1684x_{16} + (-0.1917)x_{17}\end{array}\right)$$

x_1=Black (not Hispanic)
x_2=Hispanic
x_3=Asian/Pacific Islander
x_4=American Indian/Alaska Native
x_5=Age 13-19
x_6=Age 20-24
x_7=Age 25-29
x_8=Age 30-34
x_9=Age 40-44
x_{10}=Age 45-49
x_{11}=Age 50-54
x_{12}=Age 55-59
x_{13}=Age 60-64
x_{14}=Age 65 or older
x_{15}=Female
x_{16}=Born outside United States
x_{17}=Resides in a large MSA

5.2. Discussion

Analysis of the APIDS dataset resulted in several key observations. First, the 35-39 age group had the lowest percentage of cases in the NIR category compared to other age groups when other study variables were held constant. For each successive age group upward or downward from the 35-39 age group starting point, the percentage of cases that were in the NIR category gradually increased. These percentages generally formed an upward-opening asymmetrical curve when graphed. A similar curve shape was noted for the adjusted odds ratios of these age groups after applying a logistic regression model. Although little research is available on the topic of patient age and the ability to ascertain patient risk factors for HIV transmission, it is possible health care providers and their patients may feel more comfortable discussing sexual behavior and other sensitive topics when the patient and provider are of a similar age. Being of the same generation, they would presumably have more in common. Also, in some cultures discussing sex with elders is considered offensive. Interestingly, the mean age of nurses in the United States has remained around 39 years of age since 1995 while the mean age of physicians, surgeons and nurse practitioners has remained around 43 years of age. (Bureau of Labor Statistics, 2011) Thus, it is reasonable to assume that when their AIDS case report forms were being completed, many of the patients in this study in the 35-39 age group may have been interviewed by health care providers who were of a similar age. On the other hand, patients in the various age groups older than 35-39 likely were likely interviewed a provider who was much younger than them (and vice versa). If a patient 52 years of age was interviewed by a nurse 39 years of age there would be a substantial age difference (13 years).

When cases were stratified by race/ethnicity and other study variables were held constant, non-Hispanic Blacks had the highest adjusted odds ratio for probability of being in the NIR

category (2.1 OR, 95% CI 2.0-2.3) followed by Hispanics (1.4 OR, 95% CI 1.3-1.6) when compared to non-Hispanic Whites. As noted earlier, patients of color are generally more likely to face barriers to effective patient-provider communication in health care settings including challenges of HIV-related stigma, homophobia, discrimination, health literacy, social norms which discourage self-reporting taboo behavior, low patient empowerment and other barriers. These barriers are partly the result of an extensive history of oppression against persons of color. This context of oppression has directly and indirectly perpetuated disparities in rates of incarceration, drug use, educational attainment, household income, health care and utilization and poor health care outcomes.

The adjusted odds ratio for probability of being in the NIR category was 1.5 for women compared to men (1.5 OR, 95% CI 1.4-1.6) when other study variables were held constant. This finding may be explained in part by the fact that the current CDC algorithm for determining a patient's transmission category places heterosexual contact at a lower point on the hierarchy of likely transmission modes and defines heterosexual contact more restrictively than other categories.

For patients who reside in a large MSA, the adjusted odds ratio for probability of being in the NIR category was 0.8 compared to patients who reside in less-populated areas (0.8 OR, 95% CI 0.9-0.9) when other study variables were held constant. Social norms in urban and rural areas likely differ and this may create barriers to ascertaining non-heterosexual behavior and injection drug use. Because the vast majority of persons living with HIV reside in large MSAs, these areas may be home to more AIDS service organizations implementing programs for HIV prevention through patient education and empowerment. These areas may also have a greater number of HIV-related training opportunities for health care providers because of the higher

number of patients being served. It is also reasonable to assume social marketing campaigns for reducing HIV-related stigma may be targeted with greater emphasis in large MSAs because these areas generally have higher rates of HIV incidence and prevalence.

Finally, for patients born outside the United States, the adjusted odds ratio for probability of being in the NIR category was 1.2 compared to patients born in the United States (1.2 OR, 95% CI 1.1-1.3) when other study variables were held constant. Immigrants living with HIV may face significant cultural differences, social norms, language barriers, fear of government systems and the possibility of deportation, lower levels of health literacy and confusion maneuvering through complex systems of care.

6. Conclusions and recommendations

To identify factors which may be associated with missing or incomplete patient risk factor information on AIDS case report forms, this study involved the use of key informant interviews with subject matter experts and analysis of the APIDS dataset for variables associated. In addition, based on a review of the existing literature a number of potentially helpful strategies were identified for addressing this problem.

In key informant interviews, several themes emerged. In terms of factors contributing to low rates of risk factor ascertainment, highlights from key informant interviews include concerns regarding provider awareness, concerns regarding the clarity of instructions found on the case report form as well as HIV-related stigma and homophobia which together discourage self-reporting of taboo behavior such as male-to-male sexual contact and the use of injection drugs. Also, several key informants felt the current CDC algorithm has more restrictive criteria for the male-to-female sexual contact category, when compared to the criteria for other transmission categories. Key informants offered several helpful suggestions for improving risk factor ascertainment and emphasized the value of scaling-up outreach to health care providers as well as opportunities for improving the layout and instructions found on the HIV/AIDS case report forms.

In logistic regression, factors identified as associated with missing or incomplete risk factor information include patient age (especially younger and older age groups), female sex, non-White race/ethnicity, residence outside a large MSA and birth outside the United States. These findings were consistent a study of crude odds ratios conducted in 2006 by researchers McDavid et al.

It is important to keep in mind the fact that risk factor data alone does not adequately describe the full range of factors that lead to infection with HIV. Apart from sexual behavior and needle-sharing, many additional social factors contribute significantly to racial/ethnic disparities in HIV incidence. Prompt diagnosis and retention in care, for example, facilitates viral load suppression and is associated with reduced likelihood of infection. Thus, while risk behavior data may provide useful clues to inform our understanding of HIV incidence trends and population-specific needs for prevention and care, the need for interventions aimed at changing individual behavior should not overshadow the need for population-level approaches and structural interventions that address inequities in social determinants of health. These include strategies aimed at increasing access to health care services among the uninsured, reducing disparities in rates of incarceration, educational attainment and household income; promoting HIV testing among persons of color as well as strategies for earlier diagnosis and linkage to care. Biomedical approaches with demonstrated evidence of cost-effectiveness (such as viral suppression and non-occupational post-exposure prophylaxis for HIV infection) should also be employed as part of a broad, multi-faceted strategy for HIV prevention. (Hallfors, 2007) As opportunities arise for strengthening the nationwide system of HIV/AIDS surveillance, stakeholders should be open to redefining the purpose of the patient risk history section of the case report form altogether, allowing new questions to be added, outdated questions to be removed and the algorithm to be modified. Questions that are asked of patients should serve a specific and important purpose and information obtained from surveillance should support public health efforts to improve our response to the HIV/AIDS epidemic.

References

Adimora, A. A, et al. (2006). HIV and African Americans in the Southern United States: Sexual networks and social context. Sexually Trans Dis. Pubmed: 16794554.

Brissette, I., Gelberg, K. H., & Grey, A. J. (2006). The effect of message type on physician compliance with disease reporting requirements. Public Health Rep, 121(6), 703-709.

Brody, S. (1995a). Lack of evidence for transmission of human immunodeficiency virus through vaginal intercourse. Arch Sex Behav. 1995 Aug;24(4):383-93.

Brody, S. (1995b). Patients misrepresenting their risk factors for AIDS. Int J STD AIDS, 6(6), 392-398.

Buehler, J. W., Diaz, T., Hersh, B. S., & Chu, S. Y. (1996). The supplement to HIV-AIDS Surveillance project: an approach for monitoring HIV risk behaviors. Public Health Rep, 111 Suppl 1, 133-137.

CDC. (2004). Demographics and Behavioral Data from a Supplemental HIV/AIDS Behavioral Surveillance Project, 1997–2000. Atlanta: US Department of Health and Human Services, Centers for Disease Control and Prevention; 2004; Special Surveillance Report No. 2: 5-6.

CDC. (2005). Technical Guidance for HIV/AIDS Surveillance Programs. Atlanta, Georgia.

CDC. (2011). Vital Signs: HIV Prevention Through Care and Treatment -- United States. MMWR. 60(47);1618-1623.

Chorba, T. L., Berkelman, R. L., Safford, S. K., Gibbs, N. P., & Hull, H. F. (1989). Mandatory reporting of infectious diseases by clinicians. JAMA, 262(21), 3018-3026.

Colfax, G. N., & Bindman, A. B. (1998). Health benefits and risks of reporting HIV-infected individuals by name. Am J Public Health, 88(6), 876-879.

CSTE (Council of State and Territorial Epidemiologists). (2007). Heterosexual HIV Transmission Classification: CSTE Recommendation #07-ID-09. Infectious Disease Committee.

Dowell, D., Gaffga, N., Weinstock, H., & Peterman, T. (2009). Integration of Surveillance for STDs, HIV, Hepatitis, and TB: A Survey of U.S. STD Control Programs. Public Health Rep, 124(Sup 2), 31-38.

Doyle, T. J., Glynn, M. K., & Groseclose, S. L. (2002). Completeness of notifiable infectious disease reporting in the United States: an analytical literature review. Am J Epidemiol, 155(9), 866-874.

Duffus WA et al. Effect of physician specialty on counseling practices and medical referral patterns among physicians caring for disadvantaged human immunodeficiency virus-infected populations. Clin Infect Dis. 2003 Jun 15;36(12):1577-84. Epub 2003 Jun 5.

Espinoza, L., Hall, H. I., Hardnett, F., Selik, R. M., Ling, Q., & Lee, L. M. (2007). Characteristics of persons with heterosexually acquired HIV infection, United States 1999-2004. Am J Public Health, 97(1), 144-149. doi: AJPH.2005.077461 [pii] 10.2105/AJPH.2005.077461

Fagan, J. L., Bertolli, J., & McNaghten, A. D. (2010). Understanding people who have never received HIV medical care: a population-based approach. Public Health Rep, 125(4), 520-527.

Farley, T. A., et al. (2006). Sexually transmitted diseases in the Southeastern United States: Location, race and social context. Sexually Trans Dis. 33(7):S58-64. Pubmed 16432486.

Georgia Department of Public Health (DPH). (2011). Standard #T-28: HIV/AIDS Surveillance and Notifiable Disease Reporting. List of minimum quality standards for HIV prevention

and HIV testing programs. HIV Unit, Infectious Disease and Immunization Program, Georgia Department of Public Health.

Georgia Department of Public Health (DPH). (2013). Data request submitted to the HIV/AIDS Epidemiology Section. Data request submitted February 6, 2013.

Georgia Department of Public Health (DPH). (2012). PHWeek newsletter dated February 27, 2012, accessed January 30, 2013. <http://health.state.ga.us/phweek/022712/>

Gerbert B, Maguire BT, Bleecker T, Coates TJ, McPhee SJ. Primary care physicians and AIDS. Attitudinal and structural barriers to care. JAMA. 1991 Nov 27;266(20):2837-42.

Glynn, M. K., Lee, L. M., & McKenna, M. T. (2007). The status of national HIV case surveillance, United States 2006. Public Health Rep, 122 Suppl 1, 63-71.

Gollub, E. L. (1999). Human rights is a US problem, too: the case of women and HIV. American Journal of Public Health 89.10:1479-1482.

Gross M, Holte SE, Marmor M, Mwatha A, Koblin BA, Mayer KH. Anal sex among HIV-seronegative women at high risk of HIV exposure. The HIVNET Vaccine Preparedness Study 2 Protocol Team. J Acquir Immune Defic Syndr. 2000, 24: 393-8.

Harrison, K. M., Kajese, T., Hall, H. I., & Song, R. (2008). Risk factor redistribution of the national HIV/AIDS surveillance data: an alternative approach. Public Health Rep, 123(5), 618-627.

Haverkos, H. W., Chung, R. C., & Norville Perez, L. C. (2003). Is there an epidemic of HIV/AIDS among heterosexuals in the USA? Postgrad Med J, 79(934), 444-448.

Hudson AL, Lee KA, Miramontes H, Portillo CJ. (2001). Social interactions, perceived support, and level of distress in HIV-positive women. Journal of the Association of Nurses in AIDS Care. 12(4):68–76.

Jennings, J., Stover, J., Bair-Merritt, M., Fichtenberg, C., Munoz, M., Maziad, R., . . . Zenilman, J. (2009). Identifying Challenges to the Integration of Computer-Based Surveillance Information Systems in a Large City Health Department: A Case Study. Public Health Rep, 124(Sup 2), 39-48.

Jones, J. L., Meyer, P., Garrison, C., Kettinger, L., & Hermann, P. (1992). Physician and infection control practitioner HIV/AIDS reporting characteristics. Am J Public Health, 82(6), 889-891.

Kaiser Family Foundation. (2011). HIV Testing in the United States. Accessed March 6, 2013, at: http://www.kff.org/hivaids/upload/6094-11.pdf.

Klevens, R, et al. (1995). Is there really a heterosexual AIDS epidemic in the United States? Findings from a multisite validation study, 1992–1995. American journal of epidemiology 149.1:75-84.

Konowitz, P. M., Petrossian, G. A., & Rose, D. N. (1984). The underreporting of disease and physicians' knowledge of reporting requirements. Public Health Rep, 99(1), 31-35.

Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. Crit Care Med. 2007 Sep;35(9):2052-6.

Lansky, A., Fleming, P. L., Byers, R. H., Karon, J. M., & Wortley, P. M. (2001). A method for classification of HIV exposure category for women without HIV risk information. MMWR Recomm Rep, 50(RR-6), 31-40.

Lee, L. M., Lehman, J. S., Bindman, A. B., & Fleming, P. L. (2003a). Validation of race/ethnicity and transmission mode in the US HIV/AIDS reporting system. Am J Public Health, 93(6), 914-917.

Lee, L. M., McKenna, M. T., & Janssen, R. S. (2003b). Classification of transmission risk in the national HIV/AIDS surveillance system. Public Health Rep, 118(5), 400-407.

Mayer, K. H., Safren, S. A., Gordon, C. M. (2004). HIV care providers and prevention: Opportunities and challenges. J Acquir Immune Defic Syndr. 37(2):130.

McClean CM, S. B., Buehler JW, Berkelman RL. (2010). Disease reporting among Georgia physicians and laboratories. J Public Health Management Practice, 16(6), 535-543.

McDavid, K., & McKenna, M. T. (2006). HIV/AIDS risk factor ascertainment: A critical challenge. AIDS Patient Care STDS, 20(4), 285-292. doi: 10.1089/apc.2006.20.285

Mokotoff, E. D. (2001). Priorities in HIV prevention. Science, 291(5501), 45-46.

Moneyham L, Sowell R, Seals B, Demi A. (2000). Depressive symptoms among African American women with HIV disease. Scholarly Inquiries in Nursing Practice. 14(1):9–39.

Moneyham L, Seals B, Demi A, Sowell R, Cohen L, Guillory J. (1996). Perceptions of stigma in women infected with HIV. AIDS Patient Care & STDS. 10(3):162–167.

Nash, D., Andreopoulos, E., Horowitz, D., Sohler, N., & Vlahov, D. (2007). Differences among U.S. states in estimating the number of people living with HIV/AIDS: impact on allocation of federal Ryan White funding. Public Health Rep, 122(5), 644-656.

National Women and AIDS Collective. (2007). The voices of women: HIV surveillance system position paper and policy proposal. http://foundation.org/File Library/NWAC/ NWACHIVSurveillanceSystemPositionPaper.pdf

Newman, L., Samuel, M., Stenger, M., Gerber, T., Macomber, K., Stover, J., & Wise, W. (2009). Practical Considerations for Matching STD and HIV Surveillance Data with Data from Other Sources. Public Health Rep, 126(Sup 2), 7-17.

Nwanyanwu, O. C., Conti, L. A., Ciesielski, C. A., Stehr-Green, J. K., Berkelman, R. L., Lieb, S., & Witte, J. J. (1993). Increasing frequency of heterosexually transmitted AIDS in southern Florida: artifact or reality? Am J Public Health, 83(4), 571-573.

Ojikutu B, Nnaji C, Sithole J, Schneider K, Higgins-Biddle M, Cranston K, Earls F. (2013). AIDS Patient Care and STDs. 27(1): 45-54. doi:10.1089/apc.2012.0312.

Page, M. J., Harrison, K. M., Wei, X., & Hall, H. I. (2010). Federal funding for reporting cases of HIV infection in the United States, 2006. Public Health Rep, 125(5), 718-727.

Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. The Journal of Educational Research, 96(1), 3-14.

Sayles, et al. (2007). Experiences of Social Stigma and Implications For Healthcare Among a Diverse Population of HIV Positive Adults. J Urban Health. 84(6): 814–828. PMCID: PMC2232034.

Schmidt, M. A., & Mokotoff, E. D. (2003). HIV/AIDS surveillance and prevention: improving the characterization of HIV transmission. Public Health Rep, 118(3), 197-204.

Song, R., Hall, H. I., & Frey, R. (2005). Uncertainties associated with incidence estimates of HIV/AIDS diagnoses adjusted for reporting delay and risk redistribution. Stat Med, 24(3), 453-464. doi: 10.1002/sim.1935

St Lawrence, J. S., Montano, D. E., Kasprzyk, D., Phillips, W. R., Armstrong, K., & Leichliter, J. S. (2002). STD screening, testing, case reporting, and clinical and partner notification practices: a national survey of US physicians. Am J Public Health, 92(11), 1784-1788.

Szklo, M., Nieto, J. (2007) Epidemiology: Beyond the Basics, 2nd Edition. Chapter 3: Measuring Associations Between Exposures and Outcomes. Boston, MA: Jones & Bartlet Publishers.

Valdiserri, R. O., Janssen, R. S., Buehler, J. W., & Fleming, P. L. (2000). The context of HIV/AIDS surveillance. J Acquir Immune Defic Syndr, 25 Suppl 2, S97-104.

Vasiliu, O., Stover, J., E, M., Bissette, J., Dolan, C., & Sirbu, C. (2009). Providing Mailing Cost Reimbursements: The Effect on Reporting Timeliness of Sexually Transmitted Diseases in Virginia. Public Health Rep, 124(Sup 2), 78-86.

Weinhardt, L. S., Forsyth, A. D., Carey, M. P., Jaworski, B. C., & Durant, L. E. (1998). Reliability and validity of self-report measures of HIV-related sexual behavior: progress since 1990 and recommendations for research and practice. Arch Sex Behav, 27(2), 155-180.

Attachment #1: Key Informant Interview Guide

The first questions are about the NIR/NRR challenge in overall and general strategies for addressing it.

- In your experience, does there exist a problem with high percentages of cases with missing or insufficient risk factor information? If so, how severe is this problem?
- Why is this an important problem for public health to address?
- In general, what are some of the factors that contribute to this problem?
- In your experience, is this problem more common with any particular subgroups of patients, such as by race/ethnicity?
- In your experience, is this problem more common with any particular subgroups of reporting facilities, such as health care providers in the private sector, public health providers, hospitals, etc.?
- In your experience, what types of facilities tend to be the most late in reporting cases?
- What are some of the reasons why this problem exists?
- What are some of effective strategies for addressing this problem, in your opinion?

My next questions are about the HIV case report form itself, including the "risk history" section.

- Do you think providers understand the form?
- What are some of the criticisms you've heard about from providers, regarding the case report form?
- Probe: Opposition to case reporting for ideological reasons? Concerns about length, confidentiality, significance, patient backlash? Discomfort asking sensitive questions?
- When health care providers select the "unknown" option for one of the HIV risk factor questions on the case report form, what do they think that means?
- Probe: Do you think the health care provider would interpret this as "unknown to the patient" or "unknown to the health care provider"?
- In your experience, what are some of the most common errors made in the "risk history" section of the case report form? Why?
- If you could improve the form or the form's instructions, what changes would you make?
- What activities did the HIV/AIDS surveillance program implement during the most recent two years to promote awareness of the importance of risk factor acquisition among health care providers. Specifically, Appendix F (from Section 3) of the CDC guidance for HIV/AIDS surveillance programs describes recommended Training and Dissemination Options for training providers regarding risk factor acquisition. The guidance recommends addressing the importance of risk factor acquisition through presentations at professional associations/ organizations, cover memoranda that accompany quarterly statistical reports, publications in professional journals and newsletters, "Dear colleague" letters signed by public health officials and other strategies.
- Are there any other potential key informants you could recommend for my research?

Questions specifically for key informants who are physicians:

- Is the reporting of HIV/AIDS risk history perceived as important?
- What is the perceived purpose of reporting a patient's HIV/AIDS risk history?
- Is it seen as just another form?
- Do you have any recommendations for changing these perceptions?
- What are some of the barriers to complete case reporting, which are faced by medical care providers? Do you have any recommendations for overcoming those barriers?

Attachment #2: Georgia Adult HIV/AIDS Confidential Case Report Form (page 1)

Georgia Adult HIV/AIDS Confidential Case Report Form

(Patients ≥ 13 years of age at time of diagnosis)

If reporting HIV and AIDS on the same patient, please complete a separate form for each diagnosis.

I. STATE HEALTH DEPARTMENT USE ONLY

Document ID	State No
GA00-_____	_____

Date Form Completed: ____/____/_____ (mm/dd/yyyy)

Return completed form to:
Georgia Division of Public Health, Epi Section
P.O. Box 2107
Atlanta, GA 30301
Phone: 1-800-827-9769
http://health.state.ga.us/epi/hivaids

II. PATIENT IDENTIFIER INFORMATION—*Data NOT transmitted to CDC*

Patient Name: _____ (last / first / middle) **Alias:** _____ **Maiden:** _____

Current Address: _____ **City:** _____ **State:** ____

County: _____ **Zip:** _____ **Phone** ( ) _____-_____ **Social Security** _____-____-_____

III. REPORTING FACILITY INFORMATION

Provider Name: _____ (last / first / degree) **Facility:** _____

Address _____ **City:** _____ **State:** _____ **Zip:** _____

Med Rec No: _____ **Person completing form:** _____ **Phone** ( ) _____-_____

Is the reporting facility also the facility of initial diagnosis? ☐Yes ☐No **If no, also complete Section IX on reverse side.**

IV. DEMOGRAPHIC INFORMATION—*Complete ALL fields*

Diagnostic Status:
☐Adult HIV
☐Adult AIDS

Sex at Birth:
☐Male ☐Female

Transgender (if applicable):
☐Male to Female
☐Female to Male

Date of Birth:
____/____/_____ (mm/dd/yyyy)

Country of Birth:
☐US ☐Unknown
☐US Depend/Territory
Specify: _____
☐Other, *Specify:* _____

Vital Status:
☐Alive
☐Deceased

Date of Death:
____/____/_____ (mm/dd/yyyy)

State of Death: _____

Residence at Diagnosis: ☐Same as current
Address: _____
City: _____ State: _____ Zip: _____

Race (check all that apply):
☐Black/African-American ☐White
☐Native Amer/AK Native ☐Asian
☐Hawaiian/Pac Island ☐Unknown
☐Other: _____

Ethnicity:
Hispanic/Latino:
☐Yes ☐No
☐Unknown

V. PATIENT HISTORY—*Complete ALL fields*

BEFORE the first positive HIV test or diagnosis, patient EVER had:	Yes	No	Unk
Sex with male			
Sex with female			
Injected drugs			
Received clotting factor			
HETEROsexual relations with the following:			
Injection drug user (IDU)			
Bisexual male **(applies to females only)**			
Person with hemophilia/ coagulation disorder			
Transfusion recipient w/ documented HIV infection			
Person with AIDS or documented HIV infection, risk unspecified			
Received transfusion: Date 1st / Last: /			
Received organ transplant, tissue or artificial insemination			
Worked in healthcare/clinical laboratory If yes, **SPECIFY OCCUPATION:**			
Was patient infected perinatally?			

VI. DOCUMENTED LABORATORY DATA

TYPE OF TEST	RESULT			TEST DATE	
HIV Antibody Tests at Diagnosis (FIRST known positive test)					
	+	-	Indet.	Mo	Yr
HIV-1 EIA					
HIV-1/HIV-2 EIA					
HIV-1 Western Blot					

Earliest Positive HIV Detection Test	Mo	Yr
☐ Qual PCR DNA ☐ p24 antigen		
☐ Qual PCR RNA ☐ NAT		

CD4 Count	cells/µl	%	Mo	Yr
At or closest to HIV diagnosis				
First <200 or <14% OR at first AIDS OI				

Detectable HIV Viral Load	Type*	Copies/mL	Mo	Yr
Earliest				
Most Recent				

***Specify Type**: **1**-NASBA, **2**-RT-PCR (standard) **3**-RT-PCR (ultrasen), **4**-bDNA-v. 2, **5**-bDNA-v. 3

	Mo	Yr
Physician Diagnosis: If HIV lab tests were not documented, is HIV diagnosis documented by a physician? ☐**Yes** ☐**No**		

Attachment #2: Georgia Adult HIV/AIDS Confidential Case Report Form (page 2)

Georgia Adult HIV/AIDS Confidential Case Report Form (continued)

VII. AIDS INDICATOR DISEASES (ONLY COMPLETE FOR DIAGNOSED DISEASES)

Clinical Record Reviewed? ☐ Yes ☐ No	Initial Date (mo/yr)	Initial Diagnosis: Definitive	Initial Diagnosis: Presumptive	RVCT Case No _____	Initial Date (mo/yr)	Initial Diagnosis: Definitive	Initial Diagnosis: Presumptive
Candidiasis, bronchi, trachea, or lungs	/		n/a	Lymphoma, Burkitt's (or equivalent term)	/		n/a
Candidiasis, esophageal	/			Lymphoma, immunoblastic (or equivalent term)	/		n/a
Carcinoma, invasive cervical	/		n/a	Lymphoma, primary in brain	/		n/a
Coccidioidomycosis, disseminated or extrapulmonary	/		n/a	*Mycobacterium avium* complex or *M. kansasii*, disseminated or extrapulmonary	/		
Cryptococcosis, extrapulmonary	/		n/a	*M. tuberculosis*, pulmonary	/		
Cryptosporidiosis, chronic intestinal (>1mo. duration)	/		n/a	*M. tuberculosis*, disseminated or extrapulmonary	/		
Cytomegalovirus disease (other than in liver, spleen, or nodes)	/		n/a	*Mycobacterium*, of other or unidentified species, disseminated or extrapulmonary	/		
Cytomegalovirus retinitis (with loss of vision)	/			*Pneumocystis* pneumonia	/		
HIV encephalopathy	/		n/a	Pneumonia, recurrent, in 12 month period	/		
Herpes simplex: chronic ulcers (>1 mo. duration), or bronchitis, pneumonitis, or esophagitis	/		n/a	Progressive multifocal leukoencephalopathy	/		n/a
Histoplasmosis, disseminated or extrapulmonary	/		n/a	Salmonella septicemia, recurrent	/		n/a
Isosporiasis, chronic intestinal (>1mo. duration)	/		n/a	Toxoplasmosis of brain	/		
Kaposi's sarcoma	/			Wasting syndrome due to HIV (10% weight loss with diarrhea OR chronic weakness and fatigue for 30 days)	/		n/a

VIII. TREATMENT/SERVICES REFERRALS

Is patient aware of infection? ☐Yes ☐ No ☐ Unknown

PARTNER NOTIFICATION:

The Georgia Division of Public Health (GDPH) offers HIV-positive patients partner notification and linkage to care services. Please indicate if you would like GDPH to contact this patient and offer partner notification.

☐ Yes
☐ No, provider will offer.

This patient's HIV medical treatment is primarily reimbursed by:

☐ Ryan White
☐ Medicare/Medicaid
☐ Private Insurance
☐ No coverage
☐ Other public funding
☐ Clinical trial/program
☐ Unknown

Has patient received or is receiving antiretroviral therapy? ☐ Yes ☐ No ☐ Unknown

Is patient receiving or been referred for:

HIV related medical services? ☐ Yes ☐ No ☐ Unknown

Substance abuse treatment services? ☐ Yes ☐ No ☐ Unknown

IX. FACILITY OF DIAGNOSIS—*if different from Sec. III*

Facility Name _____

Address _____

City _____ State _____ Zip _____

X. WOMEN ONLY

Is patient receiving or been referred for OB/GYN services?

☐ Yes ☐ No ☐ Unknown

If Yes, provider:_____

Is patient currently pregnant?

☐ Yes ☐ No ☐ Unknown

If Yes, list EDC (due date):____/____/_____

Has patient delivered a live-born infant?

☐ Yes ☐ No ☐ Unknown

If Yes, how many times since HIV infection? _____

Date of most RECENT birth: ____/____/_____

Hospital:_____

City_____ State_____Zip_____

Child's name: _____
Last First Middle

XI. COMMENTS

_____

_____

_____

_____

Revised 4/09. This form substitutes CDC 50.42A Rev. 03/07 OMB No. 0920-0573 Exp 2/28/10.

Attachment #3: CDC 2011 Adult HIV Confidential Case Report Form: 4 pages

Patient Identification

*Patient Name *First Name	*Middle Name	*Last Name	Last Name Soundex
*Alternate Name Type (ex Alias, Married)	*First Name	*Middle Name	*Last Name
Address Type ☐ Residential ☐ Bad Address ☐ Correctional Facility ☐ Foster Home ☐ Homeless ☐ Postal ☐ Shelter ☐ Temporary	*Current Street Address		*Phone ( ) _____
City	County	State/Country	*ZIP Code
*Medical Record Number	*Other ID Type:	Number:	

U.S. Department of Health & Human Services

Adult HIV Confidential Case Report Form

(Patients ≥13 Years of Age at Time of Diagnosis) * Information NOT transmitted to CDC

Centers for Disease Control and Prevention

Health Department Use Only

Form approved OMB no 0920-0573 Exp. 01/31/2013

Date Received at Health Department __ __ /__ __ /__ __ __ __	eHARS Document UID _____	State Number _____
Reporting Health Dept - City / County	City/County Number	
Document Source _____	Surveillance Method ☐ Active ☐ Passive ☐ Follow up ☐ Reabstraction ☐ Unknown	
Did this report initiate a new case investigation? ☐ Yes ☐ No ☐ Unknown	Report Medium ☐ 1-Field Visit ☐ 2-Mailed ☐ 3-Faxed ☐ 4-Phone ☐ 5-Electronic Transfer ☐ 6-CD/Disk	

Facility Providing Information (record all dates as mm/dd/yyyy)

Facility Name			*Phone ( ) _____
*Street Address			
City	County	State/Country	Zip Code
Facility Type	*Inpatient*: ☐ Hospital ☐ Other, specify _____	*Outpatient:* ☐ Private Physician's Office ☐ Adult HIV Clinic ☐ Other, specify _____	*Screening, Diagnostic, Referral Agency*: ☐ CTS ☐ STD Clinic ☐ Other, specify _____
	Other Facility: ☐ Emergency Room ☐ Laboratory ☐ Corrections ☐ Unknown ☐ Other, specify _____		
Date Form Completed __ __ /__ __ /__ __ __ __	*Person Completing Form		*Phone ( ) _____

Patient Demographics (record all dates as mm/dd/yyyy)

Sex assigned at Birth ☐ Male ☐ Female ☐ Unknown	Country of Birth ☐ US ☐ Other/ US Dependency (please specify) _____
Date of Birth __ __ /__ __ /__ __ __ __	Alias Date of Birth __ __ /__ __ /__ __ __ __
Vital Status ☐ 1- Alive ☐ 2- Dead	Date of Death __ __ /__ __ /__ __ __ __ / State of Death _____
Current Gender Identity ☐ Male ☐ Female ☐ Transgender Male-to-Female (MTF) ☐ Transgender Female-to-Male (FTM) ☐ Unknown ☐ Additional gender identity (specify) _____	
Ethnicity ☐ Hispanic/Latino ☐ Not Hispanic/Latino ☐ Unknown	*Expanded Ethnicity _____
Race (check all that apply) ☐ American Indian/Alaska Native ☐ Asian ☐ Black/African American ☐ Native Hawaiian/Pacific Islander ☐ White ☐ Unknown	*Expanded Race _____

Residence at Diagnosis (add additional addresses in Comments)

Address Type (Check all that apply to address below) ☐ Residence at HIV diagnosis ☐ Residence at AIDS diagnosis ☐ Check if SAME as Current Address			
*Street Address			
City	County	State/Country	*ZIP Code

This report to the Centers for Disease Control and Prevention (CDC) is authorized by law (Sections 304 and 306 of the Public Health Service Act, 42 USC 242b and 242k). Response in this case is voluntary for federal government purposes, but may be mandatory under state and local statutes. Your cooperation is necessary for the understanding and control of HIV/AIDS. Information in CDC's HIV/AIDS surveillance system that would permit identification of any individual on whom a record is maintained, is collected with a guarantee that it will be held in confidence, will be used only for the purposes stated in the assurance on file at the local health department, and will not otherwise be disclosed or released without the consent of the individual in accordance with Section 308(d) of the Public Health Service Act (42 USC 242m).

Attachment #3: CDC 2011 Adult HIV Confidential Case Report Form: 4 pages (page 2)

STATE/LOCAL USE ONLY	– Patient identifier information is not transmitted to CDC! –

Physician's Name: (Last, First, M.I.) Medical Record

_____ Phone No: ( ) _____ No._____

Hospital/Facility: Person Completing Form:

_____ _____ Phone No: ( ) _____

Facility of Diagnosis (add additional facilities in Comments)

Diagnosis Type ☐ HIV ☐ AIDS (check all that apply to facility below) ☐ Check if SAME as Facility Providing Information

Facility Name	***Phone** ( ) _____
***Street Address**	

City	**County**	**State/Country**	**Zip Code**

Facility Type

Inpatient: ☐ Hospital ☐ Other, specify _____

Outpatient: ☐ Private Physician's Office ☐ Adult HIV Clinic ☐ Other, specify _____

Screening, Diagnostic, Referral Agency: ☐ CTS ☐ STD Clinic ☐ Other, specify _____

Other Facility: ☐ Emergency Room ☐ Laboratory ☐ Corrections ☐ Unknown ☐ Other, specify _____

***Provider Name**	***Provider Phone** ( ) _____	***Specialty**

Patient History (respond to all questions) (record all dates as mm/dd/yyyy) ☐ Pediatric risk (please enter in Comments)

After 1977 and before the earliest known diagnosis of HIV infection, this patient had:	
Sex with male	☐ Yes ☐ No ☐ Unknown
Sex with female	☐ Yes ☐ No ☐ Unknown
Injected non-prescription drugs	☐ Yes ☐ No ☐ Unknown
Received clotting factor for hemophilia/ coagulation disorder — Specify clotting factor: Date received (mm/dd/yyyy):__ __ /__ __ /__ __ __ __	☐ Yes ☐ No ☐ Unknown
HETEROSEXUAL relations with any of the following:	
HETEROSEXUAL contact with intravenous/injection drug user	☐ Yes ☐ No ☐ Unknown
HETEROSEXUAL contact with bisexual male	☐ Yes ☐ No ☐ Unknown
HETEROSEXUAL contact with person with hemophilia / coagulation disorder with documented HIV infection	☐ Yes ☐ No ☐ Unknown
HETEROSEXUAL contact with transfusion recipient with documented HIV infection	☐ Yes ☐ No ☐ Unknown
HETEROSEXUAL contact with transplant recipient with documented HIV infection	☐ Yes ☐ No ☐ Unknown
HETEROSEXUAL contact with person with AIDS or documented HIV Infection, risk not specified	☐ Yes ☐ No ☐ Unknown
Received transfusion of blood/blood components (other than clotting factor) (document reason in Comments section) First date received __ __ /__ __ /__ __ __ __ Last date received __ __ /__ __ /__ __ __ __	☐ Yes ☐ No ☐ Unknown
Received transplant of tissue/organs or artificial insemination	☐ Yes ☐ No ☐ Unknown
Worked in a healthcare or clinical laboratory setting If occupational exposure is being investigated or considered as primary mode of exposure, specify occupation and setting:	☐ Yes ☐ No ☐ Unknown
Other documented risk (please include detail in Comments section)	☐ Yes ☐ No ☐ Unknown

Attachment #3: CDC 2011 Adult HIV Confidential Case Report Form: 4 pages (page 3)

Laboratory Data (record additional tests in Comments section)

HIV Antibody Tests (Non-type differentiating) [HIV-1 vs. HIV-2]

TEST 1: ☐ HIV-1 EIA ☐ HIV-1/2 EIA ☐ HIV-1/2 Ag/Ab ☐ HIV-1 WB ☐ HIV-1 IFA ☐ HIV-2 EIA ☐ HIV-2 WB ☐ Other: Specify Test: _____

RESULT: ☐ Positive/Reactive ☐ Negative/Nonreactive ☐ Indeterminate **RAPID TEST** (*check if rapid*): ☐ **Collection Date:** __ __ /__ __ /__ __ __ __

TEST 2: ☐ HIV-1 EIA ☐ HIV-1/2 EIA ☐ HIV-1/2 Ag/Ab ☐ HIV-1 WB ☐ HIV-1 IFA ☐ HIV-2 EIA ☐ HIV-2 WB ☐ Other: Specify Test: _____

RESULT: ☐ Positive/Reactive ☐ Negative/Nonreactive ☐ Indeterminate **RAPID TEST** (*check if rapid*): ☐ **Collection Date:** __ __ /__ __ /__ __ __ __

HIV Antibody Tests (Type differentiating) [HIV-1 vs. HIV-2]

TEST: ☐ HIV-1/2 Differentiating (e.g., Multispot)

RESULT: ☐ HIV-1 ☐ HIV-2 ☐ Both (undifferentiated) ☐ Neither (negative) **Collection Date:** __ __ /__ __ /__ __ __ __

HIV Detection Tests (Qualitative)

TEST 1: ☐ HIV-1 RNA/DNA NAAT (Qual) ☐ HIV-1 P24 Antigen ☐ HIV-1 Culture ☐ HIV-2 RNA/DNA NAAT (Qual) ☐ HIV-2 Culture

RESULT: ☐ Positive/Reactive ☐ Negative/Nonreactive ☐ Indeterminate **Collection Date:** __ __ /__ __ /__ __ __ __

TEST 2: ☐ HIV-1 RNA/DNA NAAT (Qual) ☐ HIV-1 P24 Antigen ☐ HIV-1 Culture ☐ HIV-2 RNA/DNA NAAT (Qual) ☐ HIV-2 Culture

RESULT: ☐ Positive/Reactive ☐ Negative/Nonreactive ☐ Indeterminate **Collection Date:** __ __ /__ __ /__ __ __ __

HIV Detection Tests (Quantitative viral load) Note: Include earliest test after diagnosis

TEST 1: ☐ HIV-1 RNA/DNA NAAT (Quantitative viral load)

RESULT: ☐ Detectable ☐ Undetectable **Copies/mL:** _____ **Log:** _____ **Collection Date:** __ __ /__ __ /__ __ __ __

TEST 2: ☐ HIV-1 RNA/DNA NAAT (Quantitative viral load)

RESULT: ☐ Detectable ☐ Undetectable **Copies/mL:** _____ **Log:** _____ **Collection Date:** __ __ /__ __ /__ __ __ __

Immunologic Tests (CD4 count and percentage)

CD4 at or closest to current diagnostic status: CD4 count: _____cells/μL **CD4 percentage:** ____% **Collection Date:** __ __ /__ __ /__ __ __ __

First CD4 result <200 cells/μL or <14%: CD4 count: _____cells/μL **CD4 percentage:** ____% **Collection Date:** __ __ /__ __ /__ __ __ __

Documentation of Tests

Date of last documented negative HIV test: __ __ /__ __ /__ __ __ __

Specify type of test: _____

If HIV laboratory tests were not documented, is HIV diagnosis documented by a physician? ☐ Yes ☐ No ☐ Unknown

If YES, provide date of documentation by physician: __ __ /__ __ /__ __ __ __

Clinical (select D for Definitive or P for Presumptive where applicable) (record all dates as mm/dd/yyyy)

	D	P	Date		D	P	Date		D	P	Date
Candidiasis, bronchi, trachea, or lungs				Herpes simplex: chronic ulcers (>1 mo. duration), bronchitis, pneumonitis, or esophagitis				M. tuberculosis, pulmonary†			
Candidiasis, esophageal				Histoplasmosis, disseminated or extrapulmonary				M. tuberculosis, disseminated or extrapulmonary†			
Carcinoma, invasive cervical				Isosporiasis, chronic intestinal (>1 mo. duration)				Mycobacterium, of other/unidentified species, disseminated or extrapulmonary			
Coccidiodomycosis, disseminated or extrapulmonary				Kaposi's sarcoma				Pneumocystis pneumonia			
Cryptococcosis, extrapulmonary				Lymphoma, Burkitt's (or equivalent)				Pneumonia, recurrent, in 12 mo. period			
Cryptosporidiosis, chronic intestinal (>1 mo. duration)				Lymphoma, immunoblastic (or equivalent)				Progressive multifocal leukoencephalopathy			
Cytomegalovirus disease (other than in liver, spleen, or nodes)				Lymphoma, primary in brain				Salmonella septicemia, recurrent			
Cytomegalovirus retinitis (with loss of vision)				Mycobacterium avium complex or M. kansasii, disseminated or extrapulmonary				Toxoplasmosis of brain, onset at >1 mo. of age			
HIV encephalopathy								Wasting syndrome due to HIV			

†If TB selected above, indicate RVCT Case Number:

Attachment #3: CDC 2011 Adult HIV Confidential Case Report Form: 4 pages (page4)

Treatment/Services Referrals (record all dates as mm/dd/yyyy)

Has this patient been informed of his/her HIV infection? ☐ Yes ☐ No ☐ Unknown	This patient's partners will be notified about their HIV exposure and counseled by: ☐ 1-Health Dept ☐ 2-Physician/Provider ☐ 3-Patient ☐ 9-Unknown		
For Female Patient			
This patient is receiving or has been referred for gynecological or obstetrical services: ☐ Yes ☐ No ☐ Unknown	Is this patient currently pregnant? ☐ Yes ☐ No ☐ Unknown	Has this patient delivered live-born infants? ☐ Yes ☐ No ☐ Unknown	
For Children of Patient (record most recent birth in these boxes; record additional or multiple births in the Comments section)			
*Child's Name	Child Soundex	Child's Date of Birth	
*Child's Coded ID	Child's State Number		
Hospital of Birth (if child was born at home, enter "home birth" for hospital name)			
Hospital Name	*Phone	*Zip Code	
*Street Address	City	County	State/Country

HIV Testing and Antiretroviral Use History (if required by Health Department) (record all dates as mm/dd/yyyy)

Main source of testing and treatment history information (select one) ☐ Patient Interview ☐ Medical Record Review ☐ Provider Report ☐ NHM&E/PEMS ☐ Other		Date patient reported information __ __ /__ __ /__ __ __ __	
Ever had previous positive HIV test? ☐ Yes ☐ No ☐ Refused ☐ Don't Know/Unknown		Date of first positive HIV test __ __ /__ __ /__ __ __ __	
Ever had a negative HIV test? ☐ Yes ☐ No ☐ Refused ☐ Don't Know/Unknown		Date of last negative HIV test *(If date is from a lab test with test type, enter in Lab Data section)* __ __ /__ __ /__ __ __ __	
Number of negative HIV tests within 24 months before first positive test # _____ ☐ Refused ☐ Don't Know/Unknown			
Ever taken any antiretrovirals (ARVs)? ☐ Yes ☐ No ☐ Refused ☐ Don't Know/Unknown		If Yes, ARV medications:	
Dates ARVs taken	Date first began: __ __ /__ __ /__ __ __ __	Date of last use: __ __ /__ __ /__ __ __ __	

***Comments**

***Local / Optional Fields**

Attachment #4

Codebook (including variable recodes) for analysis of a sample of adolescent and adult observations extracted from the APIDS dataset

RECODE VARIABLES CONSTRUCTED BY GREG BAUTISTA:

nir **Binomial dependent variable based on transcat**
0 = The AIDS case was not an NIR/NRR case; 1 = The AIDS case was either NIR or NRR

num_age Numeric version of the alphanumeric *age* variable
female Numeric version of the sexclass variable (0 = male; 1 = female)
outsidemsa Categorical variable based on MSA
(0 = Resided in a large MSA; 1 = Patient did not reside in a large MSA)
foreignborn 0 = Patient was born in the USA; 1 = Patient was born outside the USA (based on "birth" variable)

NOTE: THE FOLLOWING PAGES ARE ONLY FOR REFERENCE. THESE PAGES SHOW THE VARIABLES (FROM CDC) FOR THE APIDS DATABASE.

The rectangular data file included in the AIDS Public Information Data Set contains one line of data for each AIDS case reported to CDC. Each line contains 35 columns. The columns contain 16 variables extracted from CDC's national AIDS data set.

Column	Variable	Description
1	*age*	Age group at diagnosis of the first AIDS-indicator opportunistic condition
2	*gender*	Sexual classification of patient
3	*race*	Race of patient
4	*categ*	Indicates which of the CDC AIDS case definition revisions the patient meets
5-10	*dxdate*	Month of diagnosis of first AIDS-indicator opportunistic condition
11-16	*repdate*	Date when CDC first received information about the case
17	*death*	Vital status of patient
18-19	*transcat*	HIV transmission category
20	*multrisk*	Indicates if patient had more than one HIV risk factor
21	*birth*	Country of birth
22	*sexbi*	Sex with a bisexual man (women only)
23	*sexiv*	Sex with an injecting drug user
24	*sexother*	Sex with a person with hemophilia or with a transfusion recipient
25	*sexhiv*	Sex with a person known to be infected with HIV or to have AIDS, but whose HIV risk factor is unknown
26-31	*adjwgt*	Reporting delay adjustment weight
32-35	*msa*	Region of residence at diagnosis of AIDS

Age (column 1)
This variable contains the patient's age when he or she was first diagnosed with an AIDS-indicator disease.

0	=	Less than 1 year old
1	=	1 to 12 years old
2	=	13 to 19 years old
3	=	20 to 24 years old
4	=	25 to 29 years old
5	=	30 to 34 years old
6	=	35 to 39 years old or age is missing
7	=	40 to 44 years old
8	=	45 to 49 years old
9	=	50 to 54 years old
A	=	55 to 59 years old
B	=	60 to 64 years old
C	=	65 years old or older

http://www.cdc.gov/hiv/topics/surveillance/resources/software/apids/manual/section2.htm - top

Sexclass (column 2)

Adult/adolescent males are classified according to their sexual orientation.

1 = Adult/adolescent male who has sex only with other men or sex is missing, or sexual orientation is missing

2 = Adult/adolescent male who has sex with both men and women

3 = Adult/adolescent heterosexual male or pediatric male

4 = Female (both adult/adolescent and pediatric)

http://www.cdc.gov/hiv/topics/surveillance/resources/software/apids/manual/section2.htm - top

Race (column 3)

1 = White (not Hispanic)

2 = Black (not Hispanic)

3 = Hispanic

4 = Asian/Pacific Islander

5 = American Indian/Alaskan Native

9 = Unknown

http://www.cdc.gov/hiv/topics/surveillance/resources/software/apids/manual/section2.htm - top

Categ (column 4)

This variable reflects changes made over time to the CDC surveillance definition for AIDS. Only cases meeting the current (1993) surveillance definition are included in this data set. *Categ*indicates whether the patient also met the pre-1985, 1985, or 1987 surveillance definition, and whether the diagnosis, if it meets the 1987 or 1993 definition, was definitive or presumptive. Cases that meet more than one of these surveillance definitions are classified into the category listed first. For more information about the 1993 definition, see *Morbidity and Mortality Weekly Report, Recommendations and Reports,* December 18, 1992.

1 = Case meets the pre-1985 surveillance definition

2 = Case meets the 1985 surveillance definition

3 = Case meets the 1987 surveillance definition and was diagnosed definitively

4 = Case meets the 1987 surveillance definition and was diagnosed presumptively

5 = Case meets the 1993 surveillance definition: pulmonary tuberculosis, recurrent pneumonia, and/or cervical cancer (definitive diagnosis)

6 = Case meets the 1993 surveillance definition: pulmonary tuberculosis and/or recurrent pneumonia (presumptive diagnosis)

7 = Case meets the 1993 surveillance definition, severe HIV-related immunosuppression

http://www.cdc.gov/hiv/topics/surveillance/resources/software/apids/manual/section2.htm - top

Dxdate (columns 5 through 10)

This variable contains the year and month in which the first AIDS-indicator condition was diagnosed. Columns 5 through 8 contain the year; columns 9 and 10 contain the month. Cases diagnosed before 1982 are coded as “198199.” Cases whose month of diagnosis is unknown are coded as “99” in the month portion of this variable.

http://www.cdc.gov/hiv/topics/surveillance/resources/software/apids/manual/section2.htm - top

Repdate (columns 11 through 16)

This variable contains the year and month in which CDC received the case report. Columns 11 through 14 contain the year; columns 15 and 16 contain the month. Cases reported during 1981 are coded as “198199.”

http://www.cdc.gov/hiv/topics/surveillance/resources/software/apids/manual/section2.htm - top

Death (column 17)

0 = CDC has not received a death notification for this case

1 = CDC has been notified that this patient died

Patients diagnosed during the 2 most recent years are coded as “0” regardless of the patient's vital status. AIDS prevalence rates calculated for the most recent two-year period should be interpreted with caution. The rates calculated will be artificially high because all persons diagnosed in this period are coded with a vital status of “0” (alive), even if a death has been reported to CDC for that person. This is to prevent inadvertent indirect identification of any record by linking a death date inferred from this data set to other publicly available data sets which contain death dates on individuals.

http://www.cdc.gov/hiv/topics/surveillance/resources/software/apids/manual/section2.htm - top

Transcat (columns 18 and 19)

For surveillance purposes, AIDS cases are counted only once in a hierarchy of transmission categories. Persons with more than one reported HIV risk factor are classified in the category listed first in the transmission category hierarchy, except for men with both a history of sexual contact with other men and injecting drug use. They make up a separate category. Persons with multiple reported HIV risk factors are indicated in the variable *multrisk.*

“Men who have sex with men” cases include men who report sexual contact with other men (i.e., homosexual contact) and men who report sexual contact with both men and women (i.e., bisexual contact). “Heterosexual contact” cases are in persons who report specific heterosexual contact with a person with, or at increased risk for, HIV infection (e.g., an injecting drug user).

Adults/adolescents born in, or who had sex with someone born in, a country where heterosexual transmission was believed to be the predominant mode of HIV transmission (formerly classified as Pattern II countries by the World Health Organization) are no longer classified as having heterosexually acquired AIDS unless they meet the criteria stated in the preceding paragraph. Similar to other cases in persons who were reported without information about a behavioral or a transfusion risk factor, these cases are now classified (in the absence of other risk factor information that would classify them in another transmission category) as “no risk factor reported or identified” (see *Morbidity and Mortality Weekly Report,*March 11, 1994). Children whose mother was born in, or whose mother had sex with someone born in, a Pattern II country are now classified (in the absence of other risk factor information that would classify them in another transmission category) as “Mother with/at risk for HIV infection: has HIV infection, risk factor not specified.”

“Risk factor not reported or identified” cases are in persons with no reported history of exposure to HIV through any of the routes listed in the hierarchy of transmission categories. Risk not reported or identified cases include persons who are currently under investigation by local health department officials; persons whose HIV risk factor history is incomplete because they died, declined to be interviewed, or were lost to follow-up; and persons who were interviewed or for whom other follow-up information was available and no HIV risk factor was identified. Persons who have an HIV risk factor identified at the time of follow-up are reclassified into the appropriate transmission category.

Adult/adolescent exposure categories

1 = Men who have sex with men

2 = Injecting drug use

3 = Men who have sex with men and inject drugs

4 = Hemophilia/coagulation disorder

5 = Heterosexual contact with a person with, or at increases risk for, HIV infection

7 = Receipt of blood transfusion, blood components, or tissue

8 = Risk not reported or identified

Pediatric exposure categories

9 = Hemophilia/coagulation disorder

10 = Mother with, or at risk for, HIV infection

11 = Receipt of blood transfusion, blood components, or tissue

12 = Risk not reported or identified

http://www.cdc.gov/hiv/topics/surveillance/resources/software/apids/manual/section2.htm - top

Multrisk (column 20)
Multrisk is coded only for adult/adolescent patients (13 years old or older) and indicates if the patient has HIV risk factors other than the one indicated by *transcat*.

0 = Patient's only HIV risk factor is that indicated by *transcat*

1 = Patient has additional HIV risk factor(s)

2 = Patient's HIV risk factor is not reported or identified

http://www.cdc.gov/hiv/topics/surveillance/resources/software/apids/manual/section2.htm - top

Birth (column 21)

1 = Patient was born in the United States or its dependencies and possessions, or place of birth was not specified

2 = Patient was born outside the United States

http://www.cdc.gov/hiv/topics/surveillance/resources/software/apids/manual/section2.htm - top

Heterosexual risk factor information (columns 22 through 25)
These variables (*sexbi, sexiv, sexother,* and *sexhiv*) contain additional exposure information for patients infected heterosexually. All 4 variables are coded as follows:

0 = no

1 = yes

9 = missing/unknown

The variable sexbi is coded only for women (for men, the variable contains a blank). All four variables contain "9" (missing/unknown) for patients with hemophilia, regardless of whether the HIV risk factor information is in fact unknown. This restriction is necessary in order to comply with the Assurance of Confidentiality on page 5. Of the 4,596 AIDS cases reported through December 1995 among adults/adolescents with hemophilia, less than 4% also reported heterosexual contact with a person with, or at increased risk for, HIV infection.

http://www.cdc.gov/hiv/topics/surveillance/resources/software/apids/manual/section2.htm - top

Adjwgt (columns 26 through 31)
This variable contains an adjustment weight which, when used as a weighting variable in a frequency tabulation, produces tabulations of AIDS cases that are adjusted for delays in case reporting (see page 11 for a discussion of delays in reporting). The weights are based on estimated reporting delay distributions that take into account exposure, geographic, and demographic variations in case reporting. The adjustment weights and the resulting tabulations are not reliable for cases diagnosed during the most recent 6 months. The *Tools* menu contains an adjusted weight option. If you select this option, all subsequent tabulations you request will be weighted accordingly.

http://www.cdc.gov/hiv/topics/surveillance/resources/software/apids/manual/section2.htm - top

MSA (columns 32 through 35)
Metropolitan area of residence at diagnosis of AIDS is identified for adult/adolescent patients residing in MSAs with 500,000 or more population, according to the latest available official U.S. Bureau of Census estimates. Each MSA is identified by a 4-digit code listed in Appendix B.

For adult/adolescent patients residing in an MSA with less than 500,000 population, in a non-metropolitan area, or whose metropolitan area of residence is unknown, and for all pediatric patients, region of residence is identified.

The regional codes are:

0001 = Northeast: Connecticut, ME, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont

0002 = Midwest: Indiana, Illinois, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, OH, South Dakota, and Wisconsin

0003 = South: Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia

0004 = West: Alaska, Arizona, California, Colorado, Idaho, Hawaii, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming

0005 = U.S. dependencies, possessions, and independent nations in free association with the United States: Guam, Puerto Rico, the U.S. Virgin Islands, and the U.S. Pacific Islands listed on page 8.

Attachment #5: Exempt study approval letter (Emory IRB)

EMORY UNIVERSITY

Institutional Review Board

January 31, 2013

Gregory Bautista
Principal Investigator
Public Health

RE: **Exemption of Human Subjects Research**

IRB00063823

Factors contributing to incomplete reporting of patient risk factor information on HIV/AIDS case report forms in Georgia

Dear Principal Investigator:

Thank you for submitting an application to the Emory IRB for the above-referenced project. Based on the information you have provided, we have determined on 01/31/2013 that although it is human subjects research, it is exempt from further IRB review and approval.

This determination is good indefinitely unless substantive revisions to the study design (e.g., population or type of data to be obtained) occur which alter our analysis. Please consult the Emory IRB for clarification in case of such a change. Exempt projects do not require continuing renewal applications.

This project meets the criteria for exemption under 45 CFR 46.101(b)(2). Specifically, you will be conducting research to describe the factors contributing to the high percentage of HIV/AIDS cases that are reported with incomplete or entirely missing patient risk factor information, as well as to identify and describe effective strategies for improving the completeness of reported patient risk behavior data. This study involves a chart review of a publicly available de-identified dataset, with no link to PHI, a literature review, and interviews with subject-matter experts. The following documents were reviewed with this submission:

- Protocol (Date Submitted for IRB Consideration: January 9, 2013)
- Statistical data request tables
- Subject matter expert interview guide
- Informed Consent Form (Version Date: 01/23/2013)

Please note that the Belmont Report principles apply to this research: respect for persons, beneficence, and justice. You should use the informed consent materials reviewed by the IRB unless a waiver of consent was granted. Similarly, if HIPAA applies to this project, you should use the HIPAA patient authorization and revocation materials reviewed by the IRB unless a waiver was granted. CITI certification is required of all personnel conducting this research.

Unanticipated problems involving risk to subjects or others or violations of the HIPAA Privacy Rule must be reported promptly to the Emory IRB and the sponsoring agency (if any).

In future correspondence about this matter, please refer to the study ID shown above. Thank you.

Sincerely,

Leslie Justice
Research Protocol Analyst
This letter has been digitally signed

Attachment #6: SAS Code

```
Options nodate nonumber formdlim="-" nocenter;
Title 'Analysis of APIDS Dataset';

PROC IMPORT OUT= WORK.in_apids
            DATAFILE='e:\logistic_regression\Pids02q4.dbf'
            DBMS=DBF REPLACE;
     GETDELETED=NO;
RUN;

/* Drop unnecessary variables.  */
DATA re_apids;
SET in_apids;
DROP CATEG DEATH REPDATE MULTRISK SEXBI SEXHIV SEXIV SEXOTHER;
dxyear=input(substr(DXDATE,1,4),4.);
/* RACE is imported as a character variable. Make it numeric. */
race = input(substr(race,1,1),1.);
run;

ods graphics on;
ods html;

Title 'Count of observations to be excluded from analysis';
Proc freq data=re_apids;
tables dxyear; label race='Count of observations to be excluded, because DXYear is before 2002'; where dxyear ne
2002; run;
Proc freq data=re_apids;
tables age; label age='Count of observations to be excluded, because age <13'; where age in ("0","1"); run;
Proc freq data=re_apids;
tables race; label race='Count of observations to be excluded, because race=missing/unknown'; where race='9';
run;

/* Filter dataset to include only adult and adolescent cases
   diagnosed in 2002 (most recent year of the dataset).
   Exclude if race is unknown/missing.  */
DATA re_apids;
set re_apids;
IF AGE NOTIN ("2","3", "4", "5", "6", "7", "8", "9", "A", "B", "C") THEN DELETE;
if dxyear=2002;
if race ne '9';
RUN;

DATA re_apids;
set re_apids;

/* Create "NIR" variable using the existing TRANSCAT variable in the dataset. */
IF TRANSCAT IN ("01", "02", "03", "04", "05", "07", "09", "10", "11" ) THEN NIR=0; /*cases with a specific
transmission category */
IF TRANSCAT IN ("08", "12") THEN NIR=1;   /*cases with no specific transmission category */
Label NIR="No Identified Risk";

if age="1" then num_age=1;
if age="2" then num_age=2;
if age="3" then num_age=3;
if age="4" then num_age=4;
if age="5" then num_age=5;
if age="6" then num_age=6;
if age="7" then num_age=7;
if age="8" then num_age=8;
if age="9" then num_age=9;
if age="A" then num_age=10;
if age="B" then num_age=11;
```

```
if age="C" then num_age=12;

IF BIRTH = "1" THEN FOREIGNBORN=0;
IF BIRTH = "2" THEN FOREIGNBORN=1;

IF GENDER IN ("1", "2", "3") THEN FEMALE=0; /* male  */
IF GENDER = "4"  THEN FEMALE=1;    /* female  */

IF MSA in ("0001", "0002", "0003", "0004","0005") THEN LARGEMSA=0;
ELSE  LARGEMSA=1;
RUN;

Title 'Number of cases in cohort and percentage NIR, by selected characteristics';

PROC FREQ DATA=re_apids ;
TABLES race*NIR / missing NOCOL NOCUM NOPERCENT;
RUN;

PROC FREQ DATA=re_apids ;
TABLES FEMALE*NIR / missing NOCOL NOCUM NOPERCENT;
RUN;

PROC FREQ DATA=re_apids ;
TABLES num_age*NIR / missing NOCOL NOCUM NOPERCENT;
RUN;

PROC FREQ DATA=re_apids ;
TABLES FOREIGNBORN*NIR / missing NOCOL NOCUM NOPERCENT;
RUN;

PROC FREQ DATA=re_apids ;
TABLES LARGEMSA*NIR / missing NOCOL NOCUM NOPERCENT;
RUN;

ods output classification=classiftable;
title 'Logisticl regression analysis of the APIDS dataset using backward elimination';
proc logistic data=RE_apids ALPHA=0.05 outest=betas plots(only)=(roc);
/* outest and covout to create dataset of parameter estimates and covariances for final selected model */

OUTPUT OUT=PROBS_backward PREDICTED=PHAT lower=lcl upper=ucl predprobs=(individual crossvalidate);
/* 'probs' dataset will be output showing predicted probabilities (phat) for each
   observation in the entire dataset and their 95% confidence limits */
class   FOREIGNBORN (ref='0') NUM_AGE (ref='6') RACE (ref='1') FEMALE (ref='0') LARGEMSA (ref='0')    /
param=ref;
model nir(EVENT='1')= FOREIGNBORN race FEMALE num_age LARGEMSA
       /   selection=backward
           expb /* For each parameter estimate, also display the odds ratio in the output */
           rsquare
           clparm=wald
           fast
           slentry=0.1  /* For a variable to be allow to enter the model,
                        a significance level of 0.1 was required (SLENTRY=0.1) */
           slstay=0.05  /* For a variable to be allow to stay in the model,
                        a significance level of 0.05 was required (SLSTAY=0.05) */
           details
           lackfit
           CTABLE PPROB=(0.05 TO 0.5 BY 0.05)
           ;
score out=ScoreBackward;
run;
```

```
DATA PROBS_backward;
SET PROBS_backward;
PREDICTION_NIR=0 ;
IF PHAT>0.5 THEN PREDICTION_NIR=1;
RUN;

proc freq data=PROBS_backward;
tables nir*PREDICTION_NIR /nocum nopercent norow nocol;
run;

ods html close;
ods graphics off;

ods listing;
run;


Title 'Stepwise logistic regression analysis of APIDS dataset';
ods output classification=classiftable;
proc logistic data=RE_apids ALPHA=0.05 outest=betas covout plots(only)=(roc);
/* outest and covout to create dataset of parameter estimates and covariances for final selected model */

OUTPUT OUT=PROBS_stepwise PREDICTED=PHAT lower=lcl upper=ucl predprobs=(individual crossvalidate);
/* 'probs' dataset will be output showing predicted probabilities (phat) for each
   observation in the entire dataset and their 95% confidence limits */

class   FOREIGNBORN (ref='0') NUM_AGE (ref='6') RACE (ref='1') FEMALE (ref='0') LARGEMSA (ref='0')    /
param=ref;
model nir(EVENT='1')=  FOREIGNBORN RACE FEMALE num_age LARGEMSA
                       /*  FOREIGNBORN | RACE | FEMALE | num_age | LARGEMSA @2  */
                       /*  @2 symbol to consider all possible two-way interactions */
       /   selection=stepwise
           expb /* For each parameter estimate, also display the odds ratio in the output */
           rsquare
           clparm=wald
           slentry=0.1 /* For a variable to be allow to enter the model,
                        a significance level of 0.1 was required (SLENTRY=0.1) */
           slstay=0.05 /* For a variable to be allow to stay in the model,
                        a significance level of 0.05 was required (SLSTAY=0.05).*/
           details
           lackfit /* To request Hosmer and Lemeshow goodness-of-fit test for the final selected model */
           CTABLE PPROB=(0.05 TO 0.5 BY 0.05)
;
score out=ScoreStepwise;
run;

DATA PROBS_stepwise;
SET PROBS_stepwise;
PREDICTION_NIR=0;
IF PHAT>0.5 THEN PREDICTION_NIR=1;
RUN;

proc freq data=PROBS_stepwise;
tables nir*PREDICTION_NIR /nocum nopercent norow nocol;
run;
```

Attachment #7: SAS Output

Count of observations to be excluded from analysis

The FREQ Procedure

dxyear	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1981	440	0.05	440	0.05
1982	1202	0.14	1642	0.20
1983	3154	0.38	4796	0.58
1984	6371	0.76	11167	1.34
1985	12060	1.45	23227	2.79
1986	19417	2.33	42644	5.11
1987	29134	3.49	71778	8.61
1988	36146	4.33	107924	12.94
1989	43541	5.22	151465	18.16
1990	49629	5.95	201094	24.11
1991	60638	7.27	261732	31.39
1992	79754	9.56	341486	40.95
1993	79965	9.59	421451	50.54
1994	73569	8.82	495020	59.36
1995	70056	8.40	565076	67.76
1996	61538	7.38	626614	75.14
1997	49926	5.99	676540	81.13
1998	42624	5.11	719164	86.24
1999	40083	4.81	759247	91.05
2000	38569	4.63	797816	95.67
2001	36087	4.33	833903	100.00

Count of observations to be excluded from analysis

The FREQ Procedure

Count of observations to be excluded, because age <13				
AGE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	3371	36.56	3371	36.56
1	5849	63.44	9220	100.00

Count of observations to be excluded from analysis

The FREQ Procedure

Count of observations to be excluded, because race=missing/unknown				
RACE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
9	642	100.00	642	100.00

Number of cases in cohort and percentage NIR, by selected characteristics

The FREQ Procedure

Frequency Row Pct	Table of RACE by NIR		
	NIR(No Identified Risk)		
RACE(RACE)	0	1	Total
1	5841 80.84	1384 19.16	7225
2	8378 65.04	4504 34.96	12882
3	3334 73.21	1220 26.79	4554
4	171 70.37	72 29.63	243
5	105 88.98	13 11.02	118
Total	17829	7193	25022

Number of cases in cohort and percentage NIR, by selected characteristics

The FREQ Procedure

Frequency Row Pct	Table of FEMALE by NIR		
	NIR(No Identified Risk)		
FEMALE	0	1	Total
0	13750 74.04	4822 25.96	18572
1	4079 63.24	2371 36.76	6450
Total	17829	7193	25022

Number of cases in cohort and percentage NIR, by selected characteristics

The FREQ Procedure

Frequency Row Pct	Table of num_age by NIR		
	NIR(No Identified Risk)		
num_age	0	1	Total
2	104 48.60	110 51.40	214
3	638 69.80	276 30.20	914
4	1475 70.24	625 29.76	2100
5	2794 73.84	990 26.16	3784
6	4070 74.90	1364 25.10	5434
7	3607 72.56	1364 27.44	4971
8	2454 71.28	989 28.72	3443
9	1418 68.67	647 31.33	2065
10	698 65.60	366 34.40	1064
11	330 58.10	238 41.90	568
12	241 51.83	224 48.17	465
Total	17829	7193	25022

Number of cases in cohort and percentage NIR, by selected characteristics

The FREQ Procedure

Frequency Row Pct	Table of FOREIGNBORN by NIR		
	NIR(No Identified Risk)		
FOREIGNBORN	0	1	Total
0	15388 71.68	6081 28.32	21469
1	2441 68.70	1112 31.30	3553
Total	17829	7193	25022

Number of cases in cohort and percentage NIR, by selected characteristics

The FREQ Procedure

Frequency Row Pct	Table of LARGEMSA by NIR		
	NIR(No Identified Risk)		
LARGEMSA	0	1	Total
0	3645 69.20	1622 30.80	5267
1	14184 71.80	5571 28.20	19755
Total	17829	7193	25022

Logisticl regression analysis of the APIDS dataset using backward elimination

The LOGISTIC Procedure

Model Information		
Data Set	WORK.RE_APIDS	
Response Variable	NIR	No Identified Risk
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	25022
Number of Observations Used	25022

Response Profile		
Ordered Value	**NIR**	**Total Frequency**
1	0	17829
2	1	7193

Probability modeled is NIR=1.

Backward Elimination Procedure

Class Level Information											
Class	Value	Design Variables									
FOREIGNBORN	0	0									
	1	1									
num_age	2	1	0	0	0	0	0	0	0	0	0
	3	0	1	0	0	0	0	0	0	0	0
	4	0	0	1	0	0	0	0	0	0	0
	5	0	0	0	1	0	0	0	0	0	0
	6	0	0	0	0	0	0	0	0	0	0
	7	0	0	0	0	1	0	0	0	0	0
	8	0	0	0	0	0	1	0	0	0	0
	9	0	0	0	0	0	0	1	0	0	0
	10	0	0	0	0	0	0	0	1	0	0
	11	0	0	0	0	0	0	0	0	1	0
	12	0	0	0	0	0	0	0	0	0	1
RACE	1	0	0	0	0						
	2	1	0	0	0						
	3	0	1	0	0						
	4	0	0	1	0						
	5	0	0	0	1						
FEMALE	0	0									
	1	1									
LARGEMSA	0	0									
	1	1									

Step 0. The following effects were entered:

Intercept FOREIGNBORN RACE FEMALE num_age LARGEMSA

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	30021.799	29029.152
SC	30029.926	29175.447
-2 Log L	30019.799	28993.152

R-Square	0.0402	Max-rescaled R-Square	0.0575

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1026.6469	17	<.0001
Score	1016.0430	17	<.0001
Wald	965.9380	17	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
FOREIGNBORN	1	13.2751	0.0003
RACE	4	471.2574	<.0001
FEMALE	1	142.7170	<.0001
num_age	10	227.1342	<.0001
LARGEMSA	1	30.5494	<.0001

Analysis of Maximum Likelihood Estimates							
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept		1	-1.5276	0.0483	1000.4607	<.0001	0.217
FOREIGNBORN	1	1	0.1684	0.0462	13.2751	0.0003	1.183
RACE	2	1	0.7496	0.0361	431.7077	<.0001	2.116
RACE	3	1	0.3588	0.0509	49.6726	<.0001	1.432
RACE	4	1	0.5290	0.1468	12.9911	0.0003	1.697
RACE	5	1	-0.6798	0.2966	5.2530	0.0219	0.507
FEMALE	1	1	0.3805	0.0319	142.7170	<.0001	1.463
num_age	2	1	0.9547	0.1425	44.8929	<.0001	2.598
num_age	3	1	0.1134	0.0798	2.0178	0.1555	1.120
num_age	4	1	0.1403	0.0581	5.8362	0.0157	1.151
num_age	5	1	0.0249	0.0492	0.2552	0.6135	1.025
num_age	7	1	0.1143	0.0453	6.3658	0.0116	1.121
num_age	8	1	0.1600	0.0497	10.3489	0.0013	1.174
num_age	9	1	0.3057	0.0578	27.9978	<.0001	1.358
num_age	10	1	0.4584	0.0730	39.4785	<.0001	1.582
num_age	11	1	0.7507	0.0924	66.0186	<.0001	2.118
num_age	12	1	0.9930	0.0998	99.0881	<.0001	2.699
LARGEMSA	1	1	-0.1917	0.0347	30.5494	<.0001	0.826

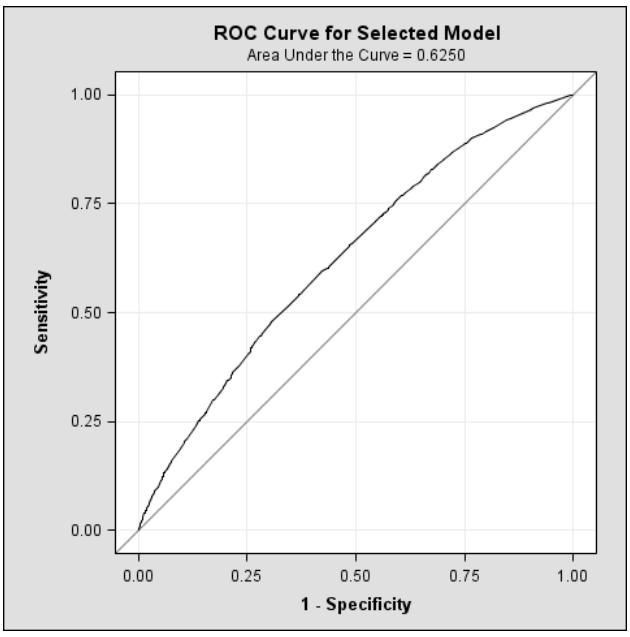
Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
FOREIGNBORN 1 vs 0	1.183	1.081	1.296
RACE 2 vs 1	2.116	1.972	2.271
RACE 3 vs 1	1.432	1.296	1.582
RACE 4 vs 1	1.697	1.273	2.263
RACE 5 vs 1	0.507	0.283	0.906
FEMALE 1 vs 0	1.463	1.375	1.557
num_age 2 vs 6	2.598	1.965	3.435
num_age 3 vs 6	1.120	0.958	1.310
num_age 4 vs 6	1.151	1.027	1.289
num_age 5 vs 6	1.025	0.931	1.129
num_age 7 vs 6	1.121	1.026	1.225
num_age 8 vs 6	1.174	1.065	1.294
num_age 9 vs 6	1.358	1.212	1.520
num_age 10 vs 6	1.582	1.371	1.825
num_age 11 vs 6	2.118	1.768	2.539
num_age 12 vs 6	2.699	2.220	3.282
LARGEMSA 1 vs 0	0.826	0.771	0.884

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	61.6	Somers' D	0.250
Percent Discordant	36.6	Gamma	0.254
Percent Tied	1.7	Tau-a	0.102
Pairs	128243997	c	0.625

Analysis of Effects Eligible for Removal			
Effect	DF	Wald Chi-Square	Pr > ChiSq
FOREIGNBORN	1	13.2751	0.0003
RACE	4	471.2574	<.0001
FEMALE	1	142.7170	<.0001
num_age	10	227.1342	<.0001
LARGEMSA	1	30.5494	<.0001

Note: No (additional) effects met the 0.05 significance level for removal from the model.

Wald Confidence Interval for Parameters				
Parameter		Estimate	95% Confidence Limits	
Intercept		-1.5276	-1.6222	-1.4329
FOREIGNBORN	1	0.1684	0.0778	0.2590
RACE	2	0.7496	0.6789	0.8203
RACE	3	0.3588	0.2590	0.4586
RACE	4	0.5290	0.2413	0.8167
RACE	5	-0.6798	-1.2611	-0.0985
FEMALE	1	0.3805	0.3181	0.4430
num_age	2	0.9547	0.6754	1.2339
num_age	3	0.1134	-0.0430	0.2697
num_age	4	0.1403	0.0265	0.2542
num_age	5	0.0249	-0.0716	0.1213
num_age	7	0.1143	0.0255	0.2031
num_age	8	0.1600	0.0625	0.2575
num_age	9	0.3057	0.1925	0.4190
num_age	10	0.4584	0.3154	0.6014
num_age	11	0.7507	0.5696	0.9317
num_age	12	0.9930	0.7975	1.1885
LARGEMSA	1	-0.1917	-0.2596	-0.1237



Partition for the Hosmer and Lemeshow Test					
		NIR = 1		NIR = 0	
Group	Total	Observed	Expected	Observed	Expected
1	2941	388	460.48	2553	2480.52
2	2580	469	477.63	2111	2102.37
3	2536	649	557.94	1887	1978.06
4	2130	576	542.12	1554	1587.88
5	2484	701	689.99	1783	1794.01
6	2492	704	746.25	1788	1745.75
7	2475	805	783.35	1670	1691.65
8	2636	943	922.03	1693	1713.97
9	2470	899	954.30	1571	1515.70
10	2278	1059	1058.92	1219	1219.08

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
46.0995	8	<.0001

Classification Table									
	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.050	7193	0	17829	0	28.7	100.0	0.0	71.3	.
0.100	7187	52	17777	6	28.9	99.9	0.3	71.2	10.3
0.150	7181	89	17740	12	29.1	99.8	0.5	71.2	11.9
0.200	6383	4490	13339	810	43.5	88.7	25.2	67.6	15.3
0.250	5524	7026	10803	1669	50.2	76.8	39.4	66.2	19.2
0.300	3900	11194	6635	3293	60.3	54.2	62.8	63.0	22.7
0.350	2383	14139	3690	4810	66.0	33.1	79.3	60.8	25.4
0.400	1110	16507	1322	6083	70.4	15.4	92.6	54.4	26.9
0.450	440	17373	456	6753	71.2	6.1	97.4	50.9	28.0
0.500	276	17584	245	6917	71.4	3.8	98.6	47.0	28.2

Logisticl regression analysis of the APIDS dataset using backward elimination

The FREQ Procedure

Frequency	Table of NIR by PREDICTION_NIR		
NIR(No Identified Risk)	PREDICTION_NIR		
	0	1	Total
0	17584	245	17829
1	6917	276	7193
Total	24501	521	25022