ABSTRACT

**Purpose**: This study evaluates the quality and usefulness of data from the current HIV Surveillance Systems in USA.

**Method**: Using the content analysis method, data from the current national HIV surveillance system is evaluated against a “quality” and “usefulness” criteria developed from previous literature.

**Results**: Several shortcomings are identified. These shortcomings weaken the nation’s ability to track, report and respond to the new HIV epidemiological trends. The problems may also have led to a misalignment in the way resources are allocated by failing to properly target the key subpopulations that need the highest resources by virtue of being at the highest risk of HIV infection. Several recommendations are suggested to address the problems.

INTRODUCTION

Public health surveillance systems play a critical role in national health management (German et al., 2001; Thacker & Berkelman, 1988). They provide data that can be used for several purposes. In particular, the analysis and interpretation of data from HIV/AIDS surveillance systems plays a critical role in observing emerging trends and making intervention decisions (Glynn, Lee, & McKenna, 2006; Stoto, 2008). Surveillance data in the US is used “to monitor the spread of HIV infection, to target HIV prevention programs and health-care services, and to allocate funding for HIV prevention and care” (Glynn et al., 2006).

HIV/AIDS continues being a key challenge in the public health arena in the US. The current epidemiological report as well as data reported in the first study using lab technology to make direct estimates puts the number of people living with HIV/AIDS in the US at 1.2 million (Center for Disease Control and Prevention, 2010a; Hall et al., 2008). By the end of 2007, over half a million people had died in America from AIDS (Center for Disease Control, 2007). Further, HIV/AIDS interventions consume a considerable portion of the nation’s budget. For example, President Obama’s fiscal year 2011 federal budget request included a $20.5 billion for domestic HIV/AIDS funding (Kaiser Family Foundation, 2010).

Even more worrying is the direction that the prevalence rates are taking. For example, while new infection rates are declining globally by 19%, and in 22 sub-Saharan Africa countries by 25%, during 1990-2009 (UNAIDS, 2010), the US diagnosis of new infections has increased by 8%
during 2005-2008 (Center for Disease Control and Prevention, 2010a). In the District of Columbia at the end of 2006, at least 3% of residents had HIV or AIDS with the rate reaching almost 6.5% among African American men (Department of Health-Washington DC, 2008). This rate is higher than that of West Africa and at par with the rates in some parts of Uganda and Kenya (UNAIDS, 2010). The US report indicates that the latest data represents a 22% increase in HIV and AIDS cases over the previous data period (Department of Health-Washington DC, 2008). These facts suggest a critical need for increased surveillance of HIV/AIDS epidemiological trends.

Further, it is critical that an evaluation of the quality and usefulness of the data from the HIV/AIDS surveillance systems be done regularly to ensure that the HIV pandemic is being monitored effectively and efficiently (Center for Disease Control, 2001). Such evaluations should also make suggestions on possible improvements, a matter that has been acknowledged in the current HIV surveillance report (Center for Disease Control and Prevention, 2010d). It is also critical to develop the evaluation criteria to be used to measure HIV surveillance systems. Researchers and practitioners can use such a metric to constantly evaluate and improve the current system in response to new trends and demands.

This study evaluates the “quality” and “usefulness” of the current HIV Surveillance Systems in USA. The following section provides the methodological approach and evaluation criteria for HIV surveillance systems from previous literature. After that we comment on current HIV surveillance system’s data in light of the emerging epidemiological trends at the national levels and the shortcomings identified. Lastly we conclude with a few recommendations.

**METHODOLOGY/APPROACH**

This study used the following methodology:

1. **Step 1-** Review the existing literature and develop a yardstick that can be used to evaluate the “quality” and “usefulness” of data from CDC HIV surveillance system as explained in the next subsection.
2. **Step 2 -** Apply the developed yardstick using the content analysis method to evaluate data from the HIV/AIDS surveillance systems in U.S. Content analysis is a "technique for making inferences by objectively and systematically identifying specified characteristics of messages" (Holsti, 1969). The communications in this study were mostly the HIV surveillance reports on diagnosed HIV infections and AIDS in the USA and dependent areas that are periodically issued by CDC. For example, see the following reports (Center for Disease Control and Prevention, 2008b, 2009, 2010a).
3. **Step 3-** Identify the successes and shortcomings of the national HIV surveillance data from the above evaluation and make suggestions on how the shortcomings can be mitigated.

Our evaluation and study focus was led by the following two research questions–:

- **RQ1:** What is the quality of the information from the HIV/AIDS surveillance systems?
- **RQ2:** What is the usefulness of the information from the HIV/AIDS surveillance systems?

**Creating the Yardstick for Evaluating the HIV Surveillance System**
The purpose of evaluating public health surveillance systems is to ensure that problems of health importance are being monitored efficiently and effectively (German et al., 2001). These researchers recommend regular evaluation of the quality and usefulness of data from public health surveillance systems.

Sixteen metrics have been previously suggested for measuring the quality of data by researchers in the data quality field (Wang & Strong, 1996). The most relevant among these for HIV/Surveillance systems were found to be: 1) representativeness, 2) completeness, and 3) accuracy.

Representativeness measures the extent to which the surveillance systems data is a true reflection of the general population distribution (German et al., 2001). For example, representativeness can measure how well the different demographics are represented by the HIV surveillance system’s data.

Completeness refers to the “extent to which data is of sufficient breadth, depth, and scope for the task at hand” (Wang & Strong, 1996). The extent to which an organization can provide outstanding services is a function of its ability to achieve the appropriate level of data completeness and in the process fully understand the consumer needs (Brohman, Watson, Piccoli, & Parasurama, 2003).

Accuracy refers to “the extent to which data is correct, reliable, and certified free of error” (Wang & Strong, 1996). Data accuracy is the foundation dimension of data quality because all the other dimensions matter little if the data is not accurate (Olson, 2003).

A central tenet in data quality literature is that the data must not only be of high quality but it must also be useful. The “fit to use” principle in data quality suggest that beauty is in the eyes of the beholder hence data is only of highly quality if it is useful to the consumer (Juran & Godfrey, 1999). Indeed, the concept of usefulness is built into the very definition of data quality in Wang and Strong’s (1996, p. 6) model in which they define data quality as data that is fit for use by data consumers. This concept has been adopted into the HIV surveillance field where researchers have proposed a guideline for evaluating HIV surveillance systems data with “usefulness” being a central determinant (German et al., 2001).

Thus the second part of the yardstick development was to identify items that would help measure the usefulness of HIV surveillance systems data. Several metrics have been suggested: Previous literature suggest that data generated by HIV surveillance systems can be used in three major ways (Pisani, Lazzari, Walker, & Schwartländer, 2003): First, the data can be used to plan and design for the right HIV interventions and programming decisions by tracking the recent trends, affected geographic areas and the key populations sections that have been adversely affected by HIV/AIDS as well as better understanding of the affected population behavioral patterns. A second and important usage of HIV/AIDS surveillance data is in monitoring and evaluating the success of the national response in the fight against HIV. Such data should be combined with behavioral data to make sure the right interpretations are being made. Third, HIV surveillance systems data can be used to expand awareness and to lobby for increased resources in funding, staffing and the response effort towards the HIV epidemic at the local, national and international level. This can be done from different standpoints like economic, social and political.

Current US HIV/AIDS surveillance data are based on estimates (Hall et al., 2008).

German et al (2001) suggest that data from HIV surveillance system should be useful in
preventing and controlling an outbreak; in revealing a deeper understanding of the implications of the disease and in providing performance measures and indicators that will form the basis of a needs assessment. CDC suggests that the data from public surveillance system should help in detecting changes in the burden of the disease on the society (Center for Disease Control, 2001). The data should also help in identifying and focusing resources in the subpopulations at the highest risk of infection (UNAIDS & World Health Organization, 2010)

RESULTS AND DISCUSSION: EVALUATING THE NATIONAL HIV/AIDS SURVEILLANCE SYSTEM’s data

Brief Description of CDC HIV/AIDS Surveillance System

![Diagram of the National HIV/AIDS Surveillance System](image)

**Figure 1: How the National HIV/AIDS Surveillance System Works (Center for Disease Control and Prevention, 2000)**
Comprehensive HIV surveillance guidelines were issued in 1999 (Center for Disease Control and Prevention, 1999). According to the established norms, CDC employs both passive and active strategies in collecting national HIV/AIDS surveillance data as shown in Figure 1. In the passive strategy, an individual will go to health provider in a local practice, hospitals or clinics and have a HIV test taken. The related laboratory sends a negative or positive test result to the ordering health provider. HIV is a reportable disease hence both the laboratory and the ordering health provider will have to report any positive HIV case to the local health department. In addition, the local health department surveillance personnel actively solicit for information by contacting health care practitioners and reviewing medical records in hospitals and clinics (Center for Disease Control and Prevention, 2000). The local health department then reports this information to the state health department who in turn transmit the information to the national center for disease control after removing the duplicates.

Earlier, we identified two key metrics for assessing the value of data from HIV surveillance systems: The first was the quality of data from the surveillance systems which for this study will look at the data representativeness, completeness and accuracy. The second one is in the usefulness of the data surveillance systems. In the following, we discuss these in two areas, namely quality and usefulness of HIV surveillance systems.

**Usefulness of Data from the National HIV Surveillance System**

The data from the current CDC system has played a crucial role in helping the government monitor, control and publicizes the HIV pandemic. Several achievements have been made: The first measure concerns integrating different data sources. CDC working with state and other government agencies planned for a National Electronic Disease Surveillance System (NEDSS). The primary goal of the system was to connect different states surveillance systems, expand, and allow the government to respond more quickly to public health threats (Center for Disease Control, 2001). The base module is already working allowing states to manage more than 140 diseases and conditions (Center for Disease Control and Prevention, 2010c). This will go a long way in addressing the lack of integration among the different systems.

The second measure involves creation of HIV/AIDS definition guideline. Qualifying cases must be diagnosed with the one or more of the "AIDS indicators" diseases or be HIV-infected with a CD4 T-lymphocyte count of less than 200 cells/ml as defined by the CDC guidelines (Center for Disease Control and Prevention, 2010a). This creates a uniform baseline for determining the qualifying case for all the physicians.

Third, CDC has moved to HIV based reporting (Center for Disease Control and Prevention, 2010d) rather than reporting based on AIDS infections cases as done since previous guidelines (Center for Disease Control and Prevention, 1999). The use of antiretroviral medicines has slowed down the progression of HIV to AIDS meaning AIDS based reporting can no longer be relied upon to give updated movements in the disease trends which is required for timely response. Thus HIV based reporting gives a more complete picture of the HIV/AIDS epidemic and can also help predict the expected level of AIDS in the future based on the average progression rates.

Fourth, as of April, 2008, CDC requires that all the fifty states and the five dependent areas report the AIDS/HIV cases by individual names which are kept confidential by CDC. This gives
higher accuracy and can be easily compared with other related infections which are reported by names.

Fifth, since January 2008, CDC is encouraging and funding states to use the latest technology to estimate the infection time for all the HIV cases being detected. This is because a new HIV diagnosis incident does not necessarily mean that the infection is recent. The person could have been with HIV for several years undetected. A significant number of people with HIV in the US have never been tested. Indeed, a previous study noted that there are 25% of people with undiagnosed HIV who are responsible for 54-70% of new infections (Marks, Crepaz, & Janssen, 2006)

The recommended method developed by CDC is called Serologic Testing Algorithm for Recent HIV Seroconversion (Janssen et al., 1998). This method uses two enzymes immunoassays (EIAs) with different sensitivities to the levels of HIV antibodies. A specimen reaction to the standard EIA but not to the less sensitive EIA, suggest a new infection (Center for Disease Control and Prevention, 2005). This measure will give more accurate infection data rates and hence lead to better evaluation of the decrease or increase of year by year new infections. In summary, data from the HIV/AIDS surveillance systems have helped meet several milestones.

**HIV Surveillance System Data Quality Challenges**

There are several shortcomings with the quality of data from the current HIV reporting and surveillance systems being used by CDC. The first shortcoming is that the data is not representative of the general population. The data only include the cases of those individuals who have been tested (Center for Disease Control and Prevention, 2000). This means that the data does not include those cases of people with HIV who have not been tested. There are several reasons why people may not have gone for testing. It could simply be that the individual is healthy and has not had a reason for being tested. Other reasons include lack of information about availability of testing, cost of testing, and confidentiality of results, stigma and misconceptions (Kaiser Family Foundation, 2010). The health provider may also not have recommended a test. For example in Washington DC while 79.7% in a 2006-7 survey said they had seen a healthcare provider, only 49.4% were offered a test (Department of Health-Washington DC & George Washington University School of Public Health, 2009). Further, access to and type of health insurance may also influence HIV testing and services utilization (Tao, Kassler, & Paterman, 1998).

Despite the fact that CDC now recommend routine testing for all patients age 13 to 64 years (Center for Disease Control and Prevention, 2006), nationally only about 53 % of the population above 18 years reports having been tested (Kaiser Family Foundation, 2010). To make matters worse, public awareness of HIV is going down. For example, the share of those reporting seeing advertisement and hearing about the domestic HIV epidemic today has declined to about half that of five years ago (Kaiser Family Foundation, 2010). This would suggest that there is need for more awareness campaigns to increase the number of people being tested which studies show can result in reduced new infections (Pinkerton, Holtgrave, & Galletly, 2008) and is cost effective (Bartlett et al., 2008; Holtgrave, 2007 ).

The second shortcoming is that the data is neither complete nor accurate. There are several shortcomings that weaken the completeness and accuracy of the CDC data. CDC advised states to move from code-based to name-based HIV case reporting in 1999 (Center for Disease Control and Prevention, 1999). This was later strengthened to a recommendation in 2005 and into a
requirement in 2008 (Center for Disease Control and Prevention, 2008c). Name based reporting was found to be more accurate and can be easily compared with other related infections which are reported by names. One unintended consequence is that the CDC directive resulted at least in the short term in incomplete and inaccurate data. We will use the state of California to demonstrate this point. California was using a code-based reporting method from 1983 but changed to the name-based system in 2006. However as of April 2009, only about 36,000 cases of HIV had been reported by local health departments to the state by name despite the fact it was estimated in 2008 that California has between 68,000 and 106,000 HIV cases (Legislative Analyst's Office- California, 2010). This suggest that more than half of the HIV cases are yet to be captured by the name based reporting system. This may result in California getting less state funding which is based on the number of reported name-based HIV cases ("The Ryan White HIV/AIDS Treatment Modernization Act of 2006," 2006). Thus the inaccuracy of the system may skew the allocation of resources and may weaken the nation’s capability of predicting new disease trends in the state of California.

Starting to report the data using confidential name reporting does not guarantee inclusion of all the data either. Rather, there is a delay period because CDC requires that states must have been in the confidential name-based reporting for two to three years before the data is considered stable enough to be included in the national data reports (Center for Disease Control, 2008). As of the 2010 report which was the last most updated report, only 37 states and five dependent areas had been doing confidential name-based reporting long enough to be considered stable as per this requirement. This means that the data from the remaining 13 states is missing which adds to the incompleteness of the current national data. This represent approximately 32% of the data which is substantial by any count (Center for Disease Control and Prevention, 2010a). Additionally, CDC expects that the name-based reporting system matures in four years when trends could be regarded as reliable (Centers for Disease Control & Health Resources and Services Administration, 2010).

Another issue of inaccuracy may result from adjustments in data reported to the CDC and to the Health Resources Services Administration (HRSA) based on gaps in code-based and name-based reporting. HRSA continually monitors the care and treatment of groups and areas funded under the Ryan White. As the main instrument of federal funding for HIV countrywide, under the 1987 Ryan White Act, living non-AIDS cases reported directly to the Health Resources Services Administration (HRSA) from the eligible code-based reporting areas are adjusted by 5% to cater for duplicative reporting ("The Ryan White HIV/AIDS Treatment Modernization Act of 2006," 2006). The rule is applied unevenly – it is not applied to areas with name-based reporting – and while controlling funding, may have political effects and affect the tallying of cases.

Emanating from the late 1970s, the US government tracks the progress of the health of Americans among in several areas including HIV. Current updates show that some of the goals such as condom use among unmarried sexually active men, testing according to guidelines, HIV diagnosis prior to AIDS, and HIV/AIDS acquired perinatally, as well as heterosexually transmitted HIV/AIDS in women could not be assessed because there was no available data (Centers for Disease Control & Health Resources and Services Administration, 2010).

**Resulting Usefulness Challenges**

7
The current HIV surveillance system has served the country well as explained earlier. However the system has not been as usefulness as it potentially can be especially in meeting second generation HIV surveillance systems goals (World Health Organization & United Nations Programmes on HIV/AIDS, 2000). The ideal system should concentrate resources where they will yield the most useful information in helping reduce HIV spread and taking care of those affected (Committee on the Ryan White CARE Act, 2004; UNAIDS & World Health Organization, 2010). This means the systems should concentrate its data collection in population most at risk of becoming newly infected with HIV (World Health Organization & United Nations Programmes on HIV/AIDS, 2000).

To understand the above shortcoming, it is important to investigate how well the HIV system has helped in concentrating resources and focus on the subpopulations at the highest risk of infection. This question cannot be answered without identifying the subpopulations that are at the highest risk of HIV infection. In 2008, 52% of the Americans diagnosed with HIV were blacks/African American despite the fact that they only consist of 12% of the US population (Center for Disease Control and Prevention, 2010a). Further, the rate of HIV diagnosis among blacks increased from 68/100,000 persons to 74/100,000 between 2005 and 2008 which is the largest increase in rates of HIV diagnoses by race or ethnicity (Center for Disease Control and Prevention, 2010b). This suggests that blacks/African American is the leading subpopulation at the highest risk of infection.

What is CDC doing about this? For some three decades, CDC has been working with various groups on several initiatives to combat HIV especially among African Americans (Sutton MY et al., 2009). The doubling by CDC of its investment in 2008 to $70 million to increase HIV testing among African American and the launching of a $10 million “Act against AIDS campaign” with 14 of the nation’s leading African American organizations are two examples of such initiatives (Center for Disease Control and Prevention, 2009). The above responses are commendable efforts by the government. However they are a drop in the ocean especially when calculated as a fraction of the whole CDC prevention budget allocated to states. For example CDC allocated a total of $515,000,000 for HIV prevention program for the 50 states and dependent areas for the 2008 fiscal year (Center for Disease Control and Prevention, 2008a). This suggests that the data has not helped as much as it should in focusing resources on the community at the highest risk of infection, a matter that is the focus of a recent modeling exercise for more efficient resource allocation (Lasry, Sansom, Hicks, & V., 2011).

The second implication of the shortcoming is that the criterion used by the surveillance systems to capture data is not specific enough to customize subpopulation specific programs. For example, CDC uses “black“ to include all people black, as does the recently released National HIV Strategy (US White House, 2010). However they fail to recognize the diversity of the
blacks/African American label and may be missing some important pockets of these subpopulations. This fact was recently demonstrated by a study done by analyzing HIV data in states with high African born communities (Kerani et al., 2008). These were California, Georgia, Massachusetts, Minnesota, and New Jersey and in King County, Washington; New York City; and the portion of Virginia included in the Washington, DC, metropolitan area. The researchers found that African born individuals accounted for only 0.6% of the population yet had 3.8% of the HIV cases diagnosed in the participating areas. Even more alarming, they consisted up to 50% of the HIV cases diagnosed in blacks in some areas like Minnesota. This suggest that classifying HIV cases occurring in African-born population as ‘African American’ or ‘black’ can at times be misleading and could lead to neglecting of this subpopulation in the allocation of resources.

A related issue arises when considering immigrant data especially in those states with high immigrant population. For example, immigrants and refugees born outside the US compromise 20% of the total population of Massachusetts residents living with HIV/AIDS yet they only consist of 12% of Massachusetts general population (Massachusetts Department of Public Health, 2008). The report further indicates that the infection rate of the non-US born population rose from 19% in 1999 to 34% in 2007 showing that the infection rate in this section of the population is rising rapidly. The leading nine countries of origin for the non US population in descending order are Haiti, Brazil, Dominican Republic, Uganda, Cape Verde Island, Kenya, Cameroon, Honduras, El Salvador and Ghana. The largest percentage (34%) of the non-US born population diagnosed with HIV/AIDS within the three year period 2005 to 2007 was from Sub-Saharan Africa. This suggests that the country of birth should be included in all cases to have a more effective surveillance system. It is highly likely that the world global HIV prevalence for the immigrants from these countries follow the immigrants as they come in and settle in their specific preferred neighborhoods (Takougang & Tidjani, 2009). Hence, studying the respective countries prevalence rates and behavior would be a key step in understanding and overcoming the spread of infection among these groups respectively (Yewoubdar, 2000).

Conclusions and Recommendations

This study assessed the data from current HIV surveillance system in the US on the basis of prior established criteria for evaluating public health surveillance systems (German et al., 2001; Wang & Strong, 1996). Data and sources of data are of critical importance in measuring the HIV/AIDS epidemic (Brookmeyer, 2010). Here we have assessed the quality of the current US HIV/AIDS data and the usefulness of the surveillance system. The results suggest that the system has achieved certain worthy goals. For example, the system has integrated different data sources at the national and state levels. Further, a standardized HIV case definition, testing and reporting guideline has been established.

However there are several shortcomings with the current surveillance system primarily due to the type of methods used to collect the data and the sources of that data. First the data is not representative of the general population but rather only reflect the population that has undergone testing. Second, the data is incomplete and hence inaccurate due to exclusion of states that do not meet certain stability requirement. For example, states need to have done confidential name reporting for at least three years to be included. Third, the systems use of “catchall labels” like
“blacks” without inclusion of country of origin or other unique identifiers has led to overgeneralization. These shortcomings affect the quality of the collected data: This in turn reduces the overall data usefulness. For example, the system has been less useful in helping concentrate resources and focus on the subpopulations at the highest risk of infection than expected due to its failure to recognize the diversity of some groups like the blacks/African American subgroup.

In view of the shortcoming in the state-based national data collection and reporting system, we recommend that certain changes be made to bolster the usefulness of the data. As HIV transitions into a more heterosexual transmission, more specification in population-based capture and reporting, not less, will be needed. Reporting by the current race groups does not adequately capture the increasingly complex transmission risks due to the cultural and ethnic diversity. Hence, it can be recommended that future strategies should involve not just race-based but also country-of-birth-based reporting to account for the role of migration in transmission. This recommendation will help to boost the successful implementation of the National HIV/AIDS Strategy, whose focus is reducing new infections, improving access to care and reducing disparities (Yehia & Frank, 2011).

References


