



Technical Report  
2004-6

---

The Validity of Self-  
Reported Clinical  
Markers and Medication  
Regimens: A Pilot  
Study

---

Marjorie M. Mariller, MD  
David Abramson  
Magdalena Sobieszczyk, MD

Columbia University  
Mailman School of Public Health  
In collaboration with the Medical and Health  
Research Association of New York,  
the NYC Department of Health and Mental  
Hygiene, the Westchester Department of  
Health, and the NY Health & Human Services  
HIV Planning Council

---

Originally submitted Sep 15, 2004  
Revision submitted Nov 17, 2004  
HRSA Grant # H89 HA 0015-13

---

C.H.A.I.N. REPORT

## **ACKNOWLEDGMENTS**

A Technical Review Team (TRT) provides oversight for the CHAIN Project in New York City and in the Tri-County area of Putnam, Rockland, and Westchester counties. In addition to Peter Messeri, PhD, Angela Aidala, PhD, and David Abramson of Columbia University's Mailman School of Public Health, TRT members include Mary Ann Chiasson, DrPH, MHRA (chair); Ken Butler, PWA Advisory Group; Susan Forlenza, MD MPH, NYCDOHMH; JoAnn Hilger, NYCDOHMH; Julie Lehane, PhD, Westchester County DOH; Grace Moon, NYCDOHMH Office of AIDS Policy; and Jennifer Nelson, MHRA.

This research was made possible by grant number H89 HA 0015-13 from the US Health Resources and Services Administration (HRSA) HIV/AIDS Bureau. The CHAIN study is supported by the HIV Health and Human Services Planning Council of New York under a Title I grant of the Ryan White Comprehensive AIDS Resource Emergency Act of 1990 through the New York City Department of Health. It is conducted under the auspices of the Medical and Health Research Association of New York City, Inc, and in Tri-County under the Westchester Department of Health. Its contents are solely the responsibility of the researchers and do not necessarily represent the official views of the U.S. Health Resources and Services Administration, the City of New York, Putnam, Rockland, or Westchester counties, or the Medical and Health Research Association of New York.

## SUMMARY OF KEY FINDINGS

There is considerable interest in assessing the validity of self-reported data among patients, since the cost of collecting patient-level data on clinical health markers is significantly less burdensome than conducting medical record reviews. Additionally, data collected from patients using survey methods provides the opportunity for a far greater range of items of interest than does a review of a circumscribed medical record. If client self-reported clinical data may be regarded as sound and reliable data, then they may be used in intricate correlational analyses of social and behavioral characteristics and their association with clinical health outcomes.

In this pilot study we compared 34 clients' self-reported measures of CD4 counts, viral loads, HIV medication regimens, and co-occurring medical conditions with data recorded in their medical charts at an HIV/AIDS clinic located in an academic medical center. For the purposes of this analysis we assumed that clinical measures noted in the medical charts were the "gold standard," and assessed how accurately patients reported this information. Key findings from the analyses include the following:

- Overall, our comparison of self-reports by CHAIN participants with their medical records suggest that self-reported data may be considered a valid measure of clinical markers, particularly those related to serious disease status.
- Self-reports of CD4 and VL cell counts agreed substantially with the information in the medical charts. When patients were asked to characterize their CD4 and VL categorically (e.g., less than 200 t-cells, greater than 10,000 copies) there was greater than 84% agreement between self-reported and chart data.
- There was substantial agreement regarding the number of medications by class of antiretroviral therapy. Twenty-five patients out of thirty-four correctly reported their exact medication regimen.
- The accuracy of self-reports of opportunistic infections ranged from substantial agreement to an almost perfect agreement depending on the specific opportunistic disease.
- When respondents' reported HIV medication regimens were analyzed as to whether they were HAART or non-HAART, they were 95% sensitive (only 5% false positive reports of HAART) and 77% specific (23% false negatives).
- Several of the CHAIN respondents were also participants in a treatment adherence program at the medical facility at which the chart reviews were conducted. Nevertheless, patients in the treatment adherence program did not report their CD4 and VL cell counts any more accurately than other patients at the clinic. This may be explained by the fact that the patients enrolled in the program are more likely to be non-adherent to medications and thus less likely to know their clinical markers. Furthermore, as only 6 persons were enrolled in the program in our sample, any discrepancies between self-reports and medical records strongly influence the statistics.

- Clients who reported CD4 counts, VL counts, and medication regimens that did not match the information in the medical records did not differ from those who did report accurately in terms of income, education, gender, or ethnicity. Those who reported the exact CD4 and VL counts accurately tended to be younger than patients who did not report these counts as accurately.

## INTRODUCTION

The Community Health Advisory and Information Network (CHAIN) study is based on the self-report of health conditions and other personal characteristics of a random sample of HIV positive adults in New York City and the surrounding suburban counties of Westchester, Rockland, and Putnam. The accuracy of such self-reported information as clinical markers and medication regimens is critical for any clinically relevant evaluations.

In 2003, a validity analysis was conducted among HIV positive adults outside the CHAIN study to assess the accuracy of self-reported medication use. Fifty HIV positive persons, demographically similar to the CHAIN cohort, were interviewed using the CHAIN survey, and their responses were compared to their medical records. The self-reported medications matched the medications reported in the medical records in 48 cases out of 50. The self-report of medication use was valid, suggesting that self-reports of medication by CHAIN participants may also be valid. This study, however, focused on medication use only.

This report summarizes the findings of a validity analysis conducted in the Spring and Summer of 2004. The self-reported data of 34 CHAIN participants were compared with the data recorded in their medical record. All the participants had taken part in the first wave of the New York City new cohort that began in 2002 and are patients of an HIV clinic located in an academic medical center. We analyzed the agreement of the self-reports of demographic characteristics (birthday & gender), existing health conditions (opportunistic diseases, other health disorders), medication use, HIV history (date of diagnosis), and lab results (CD4 and viral load) with medical records.

The CHAIN study takes several steps to optimize the reliability and accuracy of self-report. Interviewers are trained to use a standardized approach in questioning respondents, which is intended to minimize systematic errors in collecting the data. Interviews are conducted in a place chosen by the participants, where, as much as possible, participants can freely answer the questions. No other family member or friend is permitted to take part in the interview. In order to diminish the risk of recall bias when asked about HIV medications, the participants are shown a list of generic and trade names of all the antiretroviral therapies currently on the market. Pictures of the drugs complement the list of names to further help the participants. Furthermore, most of the questions refer to conditions that took place in the previous 6 months of the interview, thus limiting the risk of inaccurate recall. Inaccuracies in the data processing are minimized as much as possible with data cleaning and checking, and a standardized protocol of data management.

We did not find articles on the accuracy of the self-report of the medication regimen or on the accuracy of the self-report of HIV linked diseases in the international literature. However, 2 studies have been published so far on the accuracy of the report of clinical markers (CD4 and viral load). Cunningham et al (1997) interviewed 209 patients hospitalized in different hospitals

asking for their CD4 and compared the self-reported counts with the appropriate medical record. The self-reported CD4 did not differ significantly from the CD4 reported in the medical record. Kalichman et al (2000) surveyed 172 HIV positive adults asking them to report their most recent CD4 and viral load (VL) counts and compared their responses with their medical record. Self-report of CD4 was reliable and valid. Self-report of VL was, however, at the limit of the accuracy. Participants tended to underestimate their viral load. CD4 count/ VL count were considered as reliable/valid if they were within a range of +/- 50 cells and +/-1log of the VL, respectively. Clinically important categories of CD4 (greater than 500, between 200 and 500, less than 200) and VL (undetectable vs. detectable viral load) were reliable. A study of the factors associated with disagreement has shown that low education, low income, and low health literacy were associated with higher disagreements between self-reported HIV markers and the actual counts reported in the medical record. In a multivariate analysis, however, only poor health literacy was still associated with an increased risk of disagreement (Kalichman, 2000).

## **BACKGROUND & METHODOLOGY**

### **I. Description of CHAIN study**

The CHAIN study is a multiwave longitudinal study of HIV positives people living in New York City. Its aims are to identify the participants' needs for health and human services, to quantify their use of health care and social services, and to measure their satisfaction with care. It has been implemented by researchers at Columbia University's Mailman School of Public Health under contract to the New York City HIV Health and Human Services Planning Council.

The first cohort was established in 1994 and followed until 2002. Seven hundreds participants, representative of the population of New York City HIV-positive adults in the system of care, were recruited at baseline and followed until 2002 (or death). A refresher cohort of 268 HIV positive adults was added in 1998. A second cohort of 693 individuals has been recruited from 2002 to 2004, and the second wave of interviews is currently underway. Another cohort of 398 individuals was recruited in Westchester, Rockland and Putnam counties in 2001-2002 . A similar panel of questions has been asked of the different cohorts. Although the validity analysis presented here is based on the self-report of participants to the new NYC cohort (second cohort, wave 1), it should be applicable to the other cohorts.

#### *CHAIN sampling strategy and study participants recruitment*

The new cohort is composed of 693 participants drawn in a random sample from among a variety of HIV-related health and social service settings in New York City [see CHAIN Reports 2004\_3 and 2004\_4 for greater detail on sampling and cohort analyses]. In order to be eligible for the study, participants had to be at least 20 years old, residents of New York City, and HIV-positive for at least six months. Recruitment took place in health and social agencies in New York, and also included agencies involved in street outreach, needle exchange programs, as well as outside agency settings.

A two-stage stratified sampling method was used to sample participants. In the first stage, a sample frame of New York City health and social service agencies known to have at least 20

HIV positive adult clients was enumerated. This initial sampling frame included over 800 sites of service. Within each borough, the sample frame was further divided into four strata, based on whether they were health or social service agencies, and recipients or not of CARE Act Title I grants. Forty agencies were randomly sampled from these stratified lists. At the end of the first stage, 32 out of the 40 selected agencies have agreed to participate in the study. In the second stage, clients at these agencies were selected using either a random or a sequential sampling method. The random sampling strategy involved CHAIN staff conducting a simple random sample of clients from a list of an agency's client ID numbers. Agency staff then contacted those sampled individuals and invited them to speak with a member of the CHAIN staff. The sequential sampling strategy involved recruiting all eligible clients who presented at a medical clinic. CHAIN Report 2004\_4 describes the sampling in greater detail and provides a bias analysis as well. In addition to the randomly sampled clients, nine HIV positive "unconnected" adults were recruited through outreach agencies and in street settings. In order to be eligible, these participants needed to be aware of their HIV status and to have had no regular contact with health or social service providers in the prior 6 months.

### *CHAIN Data collection*

Data were collected through in person interviews conducted by trained community-based interviewers. Interviews took place either at the participant's home or at a place chosen by the participant. They lasted 2 to 3 hours, and were conducted either in English or in Spanish. The interviewer either directly typed the answers in a computer (computer-assisted data collection), or filled in a regular paper form of the questionnaire. Participants were remunerated for their participation in the study.

The questionnaire contains open- and closed-ended questions on perceived health, use of health services and satisfaction, social environment and housing, as well as risky behaviors. Use of health and social services are measured with questions previously used by a federally-funded study of AIDS service utilization. Questions with well-established psychometric properties (i.e., Medical Outcome Survey scale, indices for health locus of control and self-efficacy) have been used to assess health status.

## II. Medical record review and lab results

Thirty-four participants were recruited into the new NYC cohort at an HIV clinic located in an academic medical center, using a sequential enrollment strategy. All the patients who visited the clinic on four specific days were invited to participate. Among 133 patients present on these four days, 103 met the eligibility criteria, 50 of the eligible persons were approached, and 38 consented to participate. Of those 38 persons, 34 were interviewed. A physician at the clinic abstracted the following data from medical records using a standard form: service utilization; CD4 and VL cell counts; medical history; hospitalizations for opportunistic infections; history of antiretroviral use and other medications; and drug resistance. Laboratory results (CD4 and VL) were extracted from the medical center's electronic database.

## III. Analysis

We conducted a number of validity analyses, using several tests of correlation between self-

report and medical record. These include Kappa statistic, percentage of agreement, intraclass coefficient, t-test or non-parametric equivalent, Pearson correlation coefficient, Spearman correlation coefficient, sensitivity, and specificity. The self-reports of demographic characteristics, current medications, opportunistic infections, other non-HIV diseases, and date of HIV diagnosis have been compared with information recorded on the medical chart. The self-reports of the CD4 and VL cell counts have been compared with the extracted laboratory results. The sensitivity and specificity of the definition of highly active antiretroviral therapy (HAART) used in the CHAIN reports and analyses, as well as of the clinically important threshold of CD4 and VL counts have been measured. In certain instances, the patients had suffered from a disease before their first visit at the clinic. In this case, the medical record's accuracy depends on the self-report of the patient and on whether the physician has asked about the disease.

### *Variables of interest*

#### a) Demographic characteristics

In the medical record, only the month and the year of HIV diagnosis are reported, in contrast to the CHAIN interview where participants are asked about the exact date of diagnosis (for specific CHAIN questions wording, see Table 1). The analysis has been, thus, limited to the report of the exact month and year of HIV diagnosis. We have also tested the accuracy of the self-report of the year of diagnosis alone.

#### b) CD4 and VL

The self-reports of CD4 and VL counts were compared with the electronically-extracted laboratory results. During the interview, patients are asked about their most recent CD4 and VL. If the patient cannot give an exact CD4, he/she is asked if he/she knows if the count is lower than 200, between 201 and 349, between 350 and 499, or more than 500 cells. The inquiry about the VL is limited to the exact count. Self-reports of exact counts and clinically-important categories of CD4 counts (CD4 lower than 500 cells; CD4 lower or equal to 200 cells) and VL counts (detectable vs. undetectable) have been studied, as well as the report of the exact count with a admitted 10% error range. The last CD4 and VL count had to have been done days before the interview in order to allow the patient sufficient time to receive the results. We kept for the analysis only the information in the medical record that predated the CHAIN interview, and for the CD4 and VL only the measures done less than one year prior to the interview. All the participants had a CD4 and VL load in the preceding year. For the VL, both the results of the viral load standard method and of the viral load ultrasensitive method were included in the analysis.

Six patients in our sample were enrolled in a treatment adherence program. As part of this program, these patients agreed to take their medication as prescribed and to take part in sessions with peer educators about the importance of CD4 and VL cell counts. It was hypothesized that these patients might be more aware of their lab results than those not in the program. Therefore, we have also measured the accuracy of the self-report of CD4 and VL in both groups separately.

#### c) Medications

CHAIN participants are asked to report their current exact treatment regimen. As noted earlier, respondents are shown a list of market and generic names to help them as well as pictures of the medications. We analyzed the accuracy of the self-report of the number of medications by class

of antiretroviral therapy. The studied classes are : I) Protease inhibitors (PI), ii) Non-nucleoside reverse transcriptase inhibitors (NNRTI), iii) Nucleoside reverse transcriptase inhibitors (NRTI), and nucleotide reverse transcriptase inhibitor (NtRTI), iv) Fusion inhibitor (FI). The latter concerns only one medication, Fuzeon, which no participant is taking. First, we analyzed the accuracy of self-report of the exact number of overall medications. Then, we studied the accuracy of the self-report of the exact PI regimen, the exact NNRTI regimen, and the exact NRTI/NtRTI regimen, as well as the exact complete regimen. The NRTI and NtRTI have been considered together, as has been done in previous CHAIN analyses, and hydroxyurea, a drug sometimes used in conjunction with other antiretroviral therapy, has been analyzed separately.

Table 1: Exact wording of the CHAIN questions of interest

Variables of interest	Exact question wording	Format of the answer
Month and year of HIV diagnosis	When did you first become aware that you were HIV positive?	Exact date
CD4 cell count	“What was your most recent CD-4 or T-cell count?” If the participant is unable to reply than he/she is asked: “Was it above or below 350?” Depending on the answer, he/she was then ask: “Was your CD-4 or T-cell count {500 or more, or between 250 and 500? 200 or less, or between 200 and 350}?”	exact count or categories
VL cell count	“What was the result of your most recent viral load test?”	Exact count or undetectable, good, bad, do not recall
Medications	“Please look at this card and tell me which of these drugs, if any, you are taking right now.”	List of current medications
Opportunistic infections	“Has a doctor or other medical provider ever told you that you had {list of diseases}?” For each diseases that the participant reported, he/she was asked: “Was that in the last 6 months?”	Yes/no
Sexually-transmitted diseases (STD)	“Has a doctor or other medical provider ever told you that you had an STD (such as syphilis, gonorrhea)?” If it was the case, the participant was asked: “Was that in the last 6 months?”	Yes/no
Other diseases	“Has a doctor or other medical provider ever told you that you had {list of diseases}?” For each diseases that the participant reported, he/she was asked: “Are you currently having any problems with or are you being treated for...?”	Yes/no

d) Opportunistic infections, sexually-transmitted diseases, and other diseases

The participants were asked if they have ever experienced certain medical conditions or diseases, and if so, if they have experienced them in the last 6 months. In the latter case, we considered the disease as a current condition. When the date of diagnosis was missing in the medical record, we

considered that the medical condition was not currently affecting the patient at the time of the interview. This is a reliable assumption since all the participants visited the clinic for more than 6 months, and thus probably that when the date of diagnosis was missing, the conditions occurred before the patient began attending the clinic. Finally, we were not able to study the accuracy of self-report of specific sexually transmitted diseases because of non-existing information in the CHAIN database. The accuracy of self-report of each disease has been measured, as well as the self-report of opportunistic infections in overall.

### *Statistical analyses*

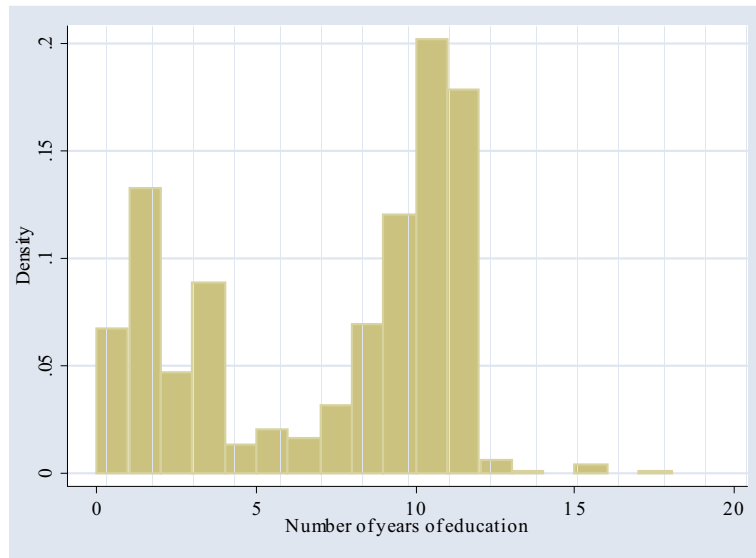
For dichotomous categorical variables or for polytomous categorical variables without natural ordering, the unweighted kappa statistic was used. For polytomous ordinal categorical variables, the weighted kappa statistic and the intraclass correlation coefficient were used (MacLure & Willett, 1987). Finally, for discrete or continuous variables, a scatter plot, a Pearson's correlation coefficient, a Spearman's correlation coefficient, an intraclass correlation coefficient, and a paired t-test or the equivalent non-parametric statistics were used (Szklo & Nieto, 2000). Despite the fact that the Pearson's correlation coefficient is one of the most widely used measures of accuracy, it is less appropriate than the other measures, because it is insensitive to systematic bias between two measures and because of the sensitivity of the coefficient to the range of values (Szklo & Nieto, 2000). In this study, the Pearson's coefficient is reported for the CD4 and VL self-reported counts for comparison with existing literature. A Pearson correlation coefficient makes sense only if the data are normally-distributed, and thus was not measured for all the discrete variables. We have used the scale by Landis & Koch (1977) to describe the degree of agreement corresponding to a given kappa. Although not the only existing scale, this is the most widely used. In general, any kappa above .75 is accepted as a good agreement among the various scales (Szklo & Nieto, 2000). A description of the different tests used in this report with their advantages and disadvantages, as well as of the Landis & Koch scale, can be found in the Appendix A. When information was missing, it was dropped for the analysis. For this reason, some analyses rely on fewer than 34 persons.

## **RESULTS**

### *I. Demographic characteristics of the sample of 34 participants included in the analysis: comparison with the entire cohort*

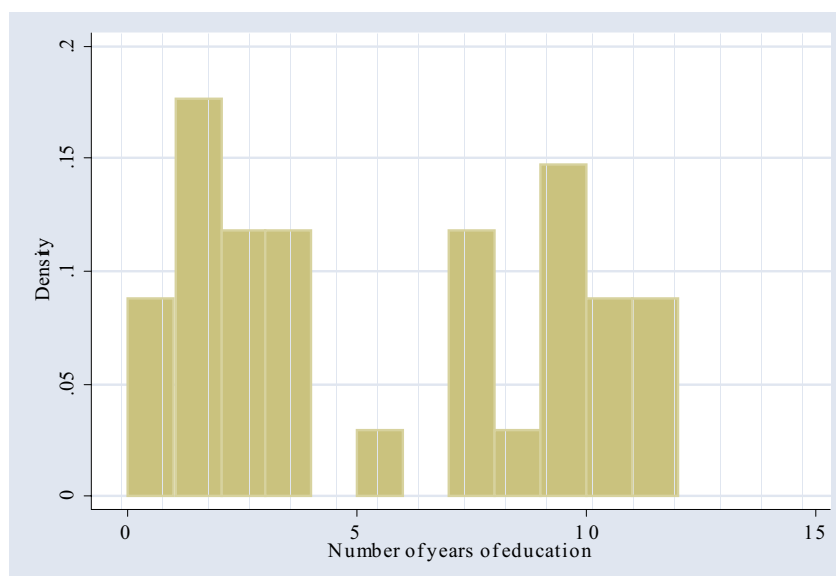
Our sample is composed of 16 women (47%) and 18 men (53%). The age of the participants range from 29 to 75, with a mean age of 49. The majority of the individuals in the sample are black (56%) or Latino (41%). Only one person is white. Past or current drug use is reported by 14% of the participants. About 56% of the participants has an individual annual income below \$7,500, and about 45% has an household income level below \$7,500. Figure 1 reports the distribution of the number of years of education in our sample. The mean number of years of education is 6 years.

Figure 1: Histogram of the number of year of education of the participants collected in our sample



In comparison, 39% of the CHAIN cohort participants are women, 60% are men and 1% are transgender. The mean age is 45 (range from 20 to 75). The cohort is composed in majority of Black people (52%), followed by Latino people (37%), White people (9%), and people from other races (2%). Having used or using drugs is reported by 47% of the participants. About 56% of the respondents have an annual individual income below \$7500, and about 48% have an annual household income below \$7500. Figure 2 reports the distribution of the number of years of education of the CHAIN participants (whole study). The average number of education is 10.5 years.

Figure 2: Histogram of the number of years of education of the CHAIN participants (whole cohort)



## *II. Validity analysis*

A summary of the results presented in this section can be found under the Appendix B.

### I) Accuracy of the self-report of demographic characteristics

Gender, as noted in the medical records, corresponded perfectly with respondents' self-reported gender. Five persons report a birthday that differs from the date of birth in the medical record. The corresponding weighted kappa coefficient is 0.98 (Discordance between the self-reported date of birth and the date of birth reported in the medical record concern only part of the date, that is only the day of the month, the month or the year differ).

#### ii) Accuracy of the self-report of the date of diagnosis

The weighted kappa for the report of the exact month and year of diagnosis is 0.76 (92% of agreement), corresponding to a substantial agreement, and the intra-class coefficient equals 0.86. The date reported differs by an average of 244 days. People tend to report a date of diagnosis before the date of diagnosis reported in their medical record. The year of diagnosis is comparable in 94% of the case, a substantial agreement (weighted kappa= 0.78). This agreement is confirmed by a substantial intra-class correlation coefficient of 0.86.

#### iii) Accuracy of the self-report of the last CD4 and VL counts

##### *a) Self-report of CD4 counts*

When subdividing the CD4 counts into categories ( $\leq 200$ , 201-350, 351-499, and  $\geq 500$  cells), we obtained a weighted kappa of 0.76, a substantial agreement. The clients' self-report and medical record data agree on the categorical classification of CD4 count 91% of the time. Table 2 illustrates the kappa statistics by CD4 count category. It shows that participants have greater difficulty accurately reporting their last CD4 count if it was above or below 500 cells than for other categories. Report of counts above or below 200 cells is reliable (substantial agreement). Note that two people were not included in the analysis since they did not report any CD4 count (neither a range nor an exact count). Four participants out of thirty-four were unable to give an exact count, but provided a range of values.

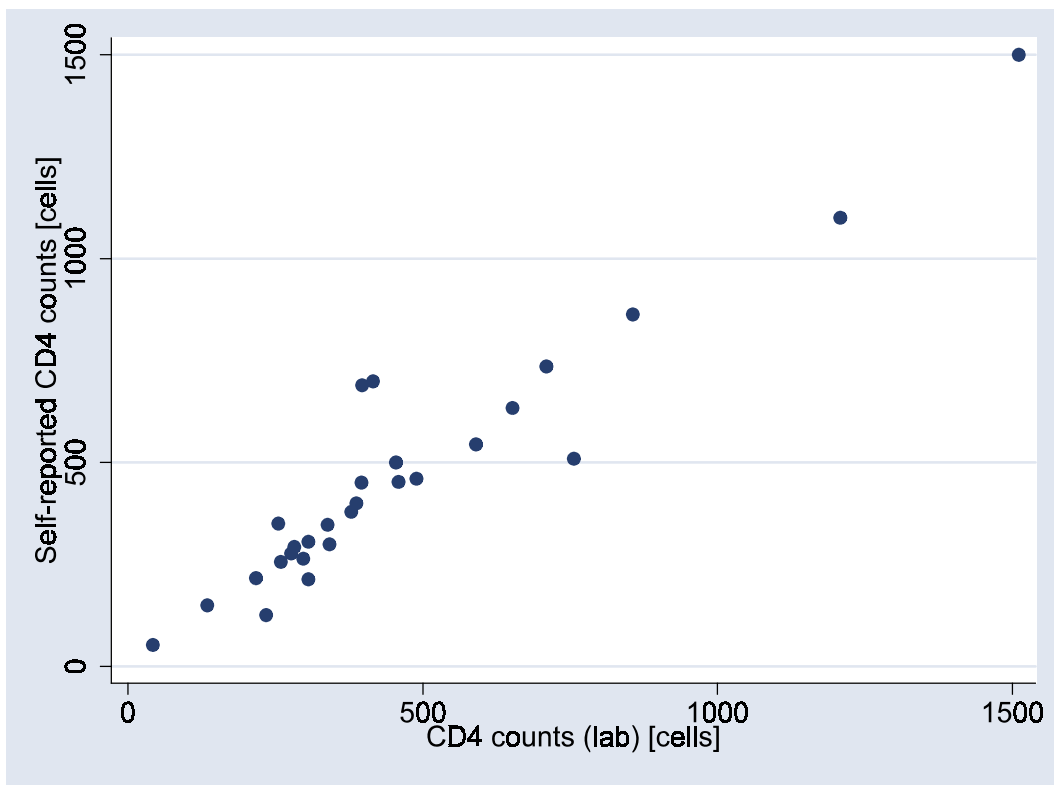
Half of the participants who were able to report an exact count have reported a CD4 count within a 10% range of the real value. A Wilcoxon signed-rank test failed to reject the hypothesis that self-report of CD4 differed from the value reported by the laboratory ( $z=0$ ,  $P=1$ ). Regarding the exact count, the intraclass coefficient equals 0.95 and the Spearman coefficient 0.94. Figure 3 plots the self-reported last CD4 count against the more recent laboratory result prior to the interview.

Table 2: Accuracy of CD4 self-report by category of CD4: positive agreement and kappa coefficient.

Variable of interest	n	% agreement	Kappa	Standard error of kappa	Probability
CD4 ≥ 500 vs. CD4 < 500	32	84%	0.64	0.16	0.0001
CD4 between 351 and 499 vs. other counts	32	91%	0.71	0.17	< 0.0001
CD4 between 201 and 350 vs. other counts	32	94%	0.86	0.17	< 0.0001
CD4 > 200 vs. CD4 ≤ 200	32	94%	0.76	0.18	< 0.0001

Legend: n= number of persons included in the analysis

Figure 3: Scatterplot of the last reported CD4 counts vs. the corresponding CD4 laboratory results



Regarding the impact of participation in the treatment adherence program to the accuracy of self-reported CD4, one of the six patients who took part in this program did not report an exact count or a range, two were not able to report the exact count, but gave a CD4 range. Four out of five people who reported a range did report it in the correct range. This corresponds to a weighted kappa of 0.57, which represents a moderate agreement. One out of the twenty-eight patients who was not in the treatment adherence program did not report a range, three were unable to report the exact value, but reported a range. The percentage of agreement while dividing the CD4 count into categories is 93%, corresponding to a weighted kappa of 0.81, an almost perfect agreement.

Fischer exact tests show that people who were not able to classify their CD4 count in the right category and people who were able to do so did not differ in term of years of education, ethnicity, income, or risky behaviors (results not reported here). Although these differences were not statistically significant, men were less likely to report inaccurate ranges of CD4 than women.

#### b) Self-report of VL count

With respect to the self-report of the most recent viral load, nine participants were unable to recall their most recent VL count. The kappa analysis is thus based on the 25 participants who gave an exact count. When dividing the counts into categories (0; 1-5,000; 5,000-9,999; 10,000-29,999; 30,000-99,999; 100,000-499,999;  $\geq 500,000$ ), we get a weighted kappa of 0.62, which is regarded as a substantial agreement. In 88% of the cases, clients' self-report and medical record data are in agreement on the VL categories. Table 3 reports the percentage agreement, as well as the kappa coefficient measured by category of VL counts. It shows that 84% of the participants are able to say if their VL is detectable or not (unweighted kappa coefficient=0.68). Forty-seven percent of the respondents provided a VL count that was within a 10%-range of the laboratory value. The intraclass coefficient measured using categories of VL equals 0.58. A paired t-test did not show any significant difference between the medical record and the self-reported count. The Pearson coefficient could not be calculated, because the data were not normally-distributed. The intraclass coefficient equals 0.61 and the Spearman coefficient equals 0.63. People are able to correctly classify their VL within +/- 1log of the laboratory value 84% of the time. Figure 4 shows a scatterplot of the self-reported VL counts and the corresponding laboratory results.

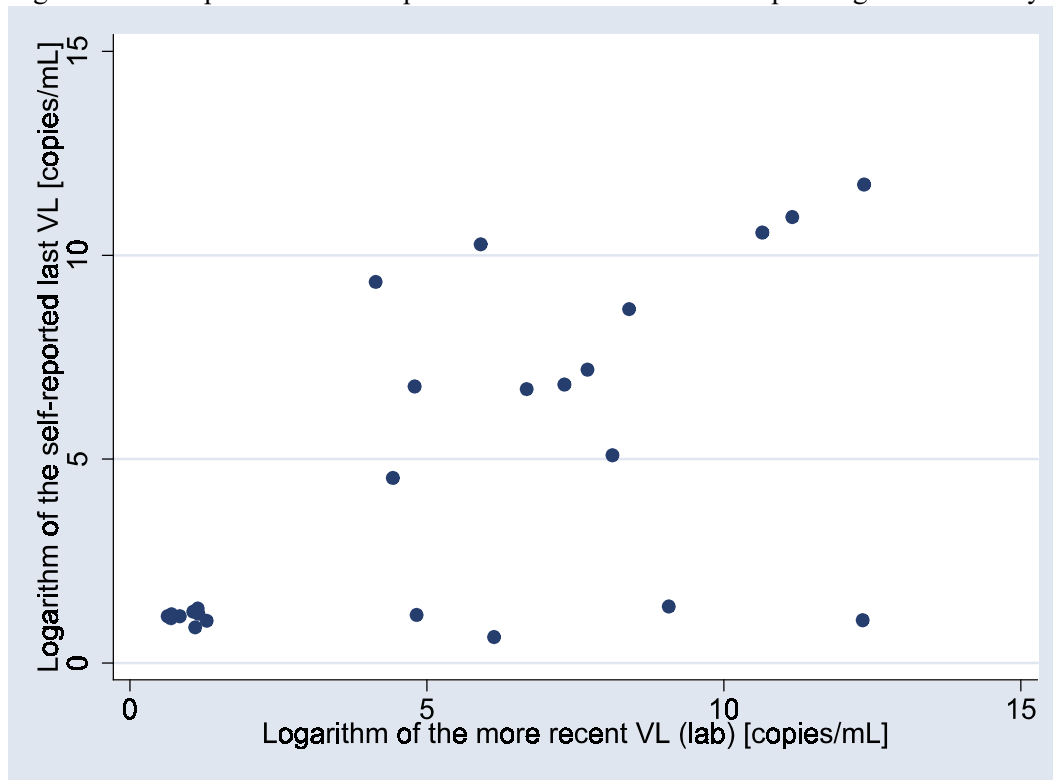
Table 3: Accuracy of VL self-report by category of VL: % agreement and kappa coefficient

Variable of interest	n	% agreement	Kappa	Standard error of kappa	Probability
VL=0 vs. VL $\neq$ 0	25	84%	0.68	0.19	0.0002
VL between 0 and 4,999 vs. other counts	25	84%	0.64	0.19	0.0003
VL between 5000 and 9,999 vs. other counts	25	100%	1.00	0.2	<0.0001
VL between 10,000 and 29,999 vs. other counts	25	92%	-0.04	0.2	0.5825
VL between 30000 and 99,999 vs. other counts	25	96%	0.78	0.20	<0.0001
VL between 100,000 and 499,999 vs. other counts	25	96%	0.65	0.19	0.0003
VL $\geq 500,000^*$	25	0%	0	---	---

\* note: no respondent has reported a VL count greater than 500,000. One respondent, however, has a VL count greater than 500,000.

Legend: n= number of persons included in the analysis

Figure 4: Scatterplot of the last reported VL counts vs. the corresponding VL laboratory results (jitter)



Of the nine participants who did not report any value for the VL, three were enrolled in the treatment adherence program. Thus, half of the people in the Jump Start Program were not able (or not willing) to report their VL. Of the three who did, all were able to report a VL that was in the same category as the VL reported on Webcis (100% agreement; Kappa=1,  $P=0.014$ ). The 22 patients not taking part in the program who report a VL agreed in 86% of the time with the value reported in the medical center's electronic lab database. The corresponding weighted kappa is 0.55, a moderate agreement.

Those who reported the same category of VL as did the electronic database did not differ from those who reported a different category in terms of gender, income, education, and risky behaviors. The group which reported another VL was however slightly older than the other (52 years old vs. 48 yrs old). The 4 persons who have reported near zero VL count although they have significant VL did not differ in terms of gender, personal income, and risk of exposure. However, they have a significantly lower degree of education ( $p=0.007$ ), and they are slightly older (53 vs. 49) than the rest of the population.

The persons who were not able to report the exact range of VL were not the same persons who were not able to report the exact range of CD4. Although 4 persons did not report correctly their VL range and their CD4 range, four did report their exact VL range, but not their CD4 range, and 11 persons did report the correct CD4, but not the correct VL range.

c) Sensitivity and specificity of some clinically-important CD4 and VL thresholds.

Table 4 reports the results of validity analysis of clinically important CD4 and VL thresholds. Note that the headings represent the findings considered as positive in our analyses. The specificity is almost 100% for all the thresholds and the sensitivity is good. The validity of these measures are thus good.

Table 4: Findings of the validity analysis: summary (CHAIN data compared to medical record (gold standard), CPMC HIV Clinic, 2003-2004)

	CD4 <500	CD4 ≤200	VL detectable
Sensitivity	80%	80%	75%
Specificity	100%	96%	100%
True positives	20	4	12
True negatives	7	26	9
False positives	5	1	4
False negatives	0	1	0

d) Comparison of the results with the existing literature

As illustrated in Table 5, which compares the findings of our validity analysis with those of other studies in the literature, our findings are comparable regarding VL and suggesting greater validity concerning CD4.

Table 5: Comparison of CHAIN results with comparable studies in the medical literature

	Sample of CHAIN participants (n=34)	Cunningham et al., 1997 (n=120)	Kalichman et al., 2000 (n=135 for CD4 & 94 for VL)
<b>Most recent CD4 count</b>			
- Mean difference (t-test) or equivalent	not significant	not significant	not significant
- Pearson r	0.94	0.84	0.82
- Intraclass coefficient	0.95	0.82	0.82
- Spearman r	0.63	0.74	0.89
- Kappa (categories)	0.76	0.51 (different cat.)	
- % agreement (categories)	91%	67% (different cat.)	
- Kappa CD4 < 500	0.64		0.71
- Kappa CD4 < 200	0.76		0.82

	Sample of CHAIN participants (n=34)	Cunningham et al., 1997 (n=120)	Kalichman et al., 2000 (n=135 for CD4 & 94 for VL))
<b>Most recent VL count</b>			
- Pearson r (log value)	--		0.73
- Spearman r (log value)	0.63		0.74
- Intraclass coefficient	0.61		0.68
- Mean difference (t-test) or equivalent	not significant		significant
- Kappa: undetectable vs. detectable	0.68		0.72

iv) Accuracy of the self-report of the current medication regimen

a) Self-report of the number of medications by medication class

Table 6 reports the kappa coefficient for each class of medications, and for the number of medications overall. As shown by Table 6, the agreement is almost perfect regarding the number of NNRTIs and the number of NRTIs/NtRTIs, and substantial agreement regarding the number of PIs and the total number of medications. Some medications are listed in the medical record and not reported by the participants. In contrast, some medications are not listed in the medical record, but reported by the participants. Both scenarios happen with the same frequency.

Table 6: Accuracy of self-report of the number of medications

Variable of interest	n	% agreement	Kappa	Standard error of kappa	Probability
Number of PIs	34	93%	0.75	0.15	<0.0001
Number of NNRTIs	34	97%	0.89	0.17	<0.0001
Number of NRTIs/NtRTIs	34	94%	0.81	0.12	<0.0001
Number of medications in total	34	93%	0.75	0.12	<0.0001

Legend: n= number of persons included in the analysis

b) Self-report of the exact medication regimen

Twenty-five participants report the exact medication regimen listed in their charts. This corresponds to an unweighted kappa of 0.70, a substantial agreement. Table 7 reports the percentage of exact regimen by category of medications. The percentage of agreement is high for each of the class of medications, and the agreement is either substantial or almost perfect. Table 8 reports the percentage agreement and kappa statistics for medications that have been taken or are reported to have been taken by the participants.

Table 7: Accuracy of the self-report of exact regimen by class of medication

Variable of interest	n	% agreement	Kappa	Kappa standard error	P-value
PI exact regimen	34	79	0.71	0.10	<0.0001
NNRTI exact regimen	34	97	0.90	0.14	<0.0001
NRTI/ NtRTI exact regimen	34	82	0.78	0.07	<0.0001
Exact regimen overall	34	74	0.70	0.05	<0.0001

Legend: n= number of persons included in the analysis

Table 8: Accuracy of the self-report of specific HIV antiretroviral medication

Variable of interest	n	self-report: # of people who report taking the drug	medical record: # of people with the drug given the medical record	# of agreement	% agreement	Kappa	Kappa standard error	P-value
Crixivan	34	0	1	33	97%	---	---	---
Invirase/For-tovase	34	2	3	33	97%	0.78	0.17	<0.0001
Kaletra	34	11	9	32	94%	0.86	0.17	<0.0001
Norvir	34	1	2	31	91%	-0.04	0.16	0.6
Viracept	34	6	6	32	94%	0.80	0.17	<0.0001
Combivir	34	7	5	32	94%	0.80	0.17	<0.0001
Epivir	34	2	2	34	100%	---	---	---
Retrovir	34	2	2	32	94%	0.47	0.17	0.0031
Videx/Dida-nosine	34	9	7	32	94%	0.84	0.17	<0.0001
Zerit	34	2	2	34	100%	1.00	0.17	<0.0001
Trizivir	34	7	7	32	94%	0.82	0.17	<0.0001
Viread	34	4	6	30	88%	0.53	0.17	0.0007
Sustiva	34	4	4	34	100%	1.00	0.17	<0.0001
Viramune	34	1	2	33	97%	0.65	0.16	<0.0001

Legend: n= number of persons included in the analysis, # = number.

Those who reported the exact medication regimen did not differ from those who did not report the exact regimen in terms of ethnicity, income, education, risky behaviors and age (results not shown). Although not significant, men tend to report their medication regimen less accurately than women.

c) Sensitivity and specificity analysis of the definition used in CHAIN studies for highly active antiretroviral therapy

The sensitivity of the definition of HAART is 95%. The specificity is 77%. In one case out of thirty-four the regimen reported by the participant was not considered as HAART by our definition, while the regimens reported by the medical record was considered as HAART (one false-negative). In 3 cases, the regimen reported by the participants were considered HAART while the regimens reported in the medical record were not considered as HAART by the definition of HAART used in CHAIN studies (three false positives).

Table 9 reports the characteristics of medication regimens based on the medical record against the characteristics found when relying on self-report. In 30 cases out of 34, self-report regimen is classified in the same category of treatment as the regimen reported in the medical record.

Table 9: Characteristics of medication regimen reported in the medical record vs. characteristics of medication regimen as self-reported

		HAART given the medication regimen reported in the medical record				
		no HIV therapy	HAART: 1st choice	HAART: 2 <sup>nd</sup> choice	ART (not HAART)	not recommended
HAART given the self-reported medication regimen	no HIV therapy	9	0	0	1	0
	HAART: 1 <sup>st</sup> choice	1	11	0	0	1
	HAART: 2 <sup>nd</sup> choice	0	0	9	1	0
	ART (not HAART)	0	0	0	0	0
	not recommended	0	0	1	0	0

v) Accuracy of the self-report of the opportunistic diseases

Tables 10 and 11 summarize the percentage of agreement and the kappa coefficient for different opportunistic infections that have occurred in the participant's lifetime and in the last six months, respectively. The tables report only the opportunistic infections that have affected at least one participant according to either the medical chart or the participant's self-report. In general, the agreement is almost perfect or substantial. The participants accurately reported if they have suffered from an opportunistic infection since their infection by the HIV 93% of the time. This seems to be high, but corresponds to a kappa of 0.46, a moderate agreement.

Table 10: Accuracy of the self-report of lifetime opportunistic infections

Variable of interest- EVER	n	# of people who reports having suffer from the disease ever	# of people who has had or has the disease given the medical record	# of agreement	% agreement	Kappa	Kappa standard error	P-value
Thrush	34	12	14	28	82%	0.63	0.17	0.0001
Mycobacterium avium-intracellular infection	34	5	1	30	88%	0.29	0.12	0.0073
Kaposi Sarcoma	34	1	0	33	97%	---	---	---
PCP	34	6	8	28	82%	0.46	0.17	0.0030
Recurrent bacterial pneumoniae	34	9	1	26	76%	0.40	0.17	0.0105
Wasting syndrom	34	12	0	22	65%	---	---	---
Herpes simplex	34	8	9	31	91%	0.77	0.17	<0.0001
Tuberculosis	34	6	7	29	85%	0.52	0.17	0.0010
Recurrent salmonella septicemia	34	1	1	34	100%	1.00	0.17	<0.0001
HIV dementia	34	2	0	32	94%	---	---	---
CNS toxoplasmosis	34	2	2	32	94%	0.47	0.17	0.0031
Extrapulmonary cryptococosis	34	2	1	33	97%	0.65	0.16	<0.0001
Chronic diarrhea	34	8	1	27	79%	0.18	0.10	0.034
Cervical dysplasia	16	5	3	14	88%	0.67	0.24	0.0022
Any opportunistic infection	34	30	27	29	85%	0.47	0.16	0.0021

Legend: n= number of persons included in the analysis, # = number.

Table 11: Accuracy of the self-report of opportunistic infections in the last 6 months.

Variable of interest- in the last 6 months	n	# of people who reports having suffer from the disease in the last 6 months	# of people who has had in the last 6 months or has the disease given the medical record	# of agree-ment	% agree-ment	Kappa	Kappa standard error	P-value
Thrush	34	3	2	33	97%	0.78	0.17	<0.0001
PCP	34	2	1	33	97%	0.65	0.16	<0.0001
Wasting syndrome	34	3	0	31	91%	---	---	---
Herpes simplex	34	1	1	34	100%	1.00	0.17	<0.0001
Recurrent salmonella septicemia	34	1	1	34	100%	1.00	0.17	<0.0001
HIV dementia	34	1	0	33	97%	---	---	---
Chronic diarrhea	34	3	1	32	94%	0.48	0.15	0.0006
Cervical dysplasia	34	1	1	16	100%	1.00	0.25	<0.0001
Any opportunistic infection	34	12	5	27	79%	0.48	0.15	0.0005

Legend: n= number of persons included in the analysis, # = number.

vi) Accuracy of the self-report of non-opportunistic diseases.

a) Sexually-transmitted diseases

The self-report and the medical record agree in 71% of the cases on the presence or absence of sexually-transmitted diseases (STD). Twelve participants report having suffered from a STD. The medical records show that 16 participants have had a STD ever (Tab. 12). The corresponding unweighted kappa is  $0.40 \pm 0.17$  ( $P= 0.0080$ ). According to the medical record and to self-report, one person has suffered a STD in the last 6 months. They agree on all the 34 cases (100% agreement).

Table 12: Self-report of any lifetime STD vs. medical record report of any lifetime STD

	No STD ever given medical record	STD ever given medical record	Total
No self-reported STD ever	15	7	22
Self-reported STD ever	3	9	12
Total	18	16	34

## b) Other medical conditions

Eight medical conditions are studied here. Three participants did not reply to the questions and were thus eliminated from the comparison, except for hepatitis and heart diseases where all the participants could be included. Table 13 summarizes the findings for these 8 medical conditions.

Table 13: Accuracy of the self-report of several non-opportunistic diseases

Variable of interest-ever	n	# of people who reports having suffer from the disease ever	# of people who has had or have the disease given the medical record	# of agreement	% agreement	Kappa	Kappa standard error	P-value
Diabetes mellitus	31	5	6	28	90%	0.67	0.18	0.0001
Hypertension	31	9	7	29	94%	0.83	0.18	<0.0001
Asthma	31	5	4	30	97%	0.87	0.18	<0.0001
High cholesterol	31	9	5	27	87%	0.64	0.17	0.0001
Chronic sinusitis	31	5	0	26	84%	---	---	---
Hepatitis	34	12	19	21	62%	0.26	0.16	0.0486
Hepatitis A	34	0	8	26	76%	---	---	---
Hepatitis B	34	3	4	31	91%	0.52	0.17	0.0010
Hepatitis C	34	9	10	27	79%	0.49	0.17	0.0021
Heart diseases	31	7	3	25	81%	0.31	0.16	0.0273

Legend: n= number of persons included in the analysis, # = number.

## DISCUSSION

Our findings suggest that self-reported data may be considered a valid measure of clinical markers, particularly those related to serious disease status. Self-report of demographic characteristics (gender and birthday) is valid with kappas that show almost perfect agreement. Five persons have reported a date of birth different from the date reported in their medical record, but differences concern only part of the date of birth. There are three possible explanations for these differences: first, the true dates of birth of some persons are not known; second, the date of birth is wrong on the official documents; third, some people cannot remember their date of birth. Overall, the difference in the report of birth day should not affect the analysis, since the average difference is only 42 days.

Self-report of the year and month of diagnosis is also reliable (substantial agreement). We note a tendency to report a diagnosis before the clinical diagnosis. A hypothesis is that people affected by HIV report when they think they have been infected, rather than when they have been diagnosed with HIV.

Furthermore, self-report of CD4 and VL cell counts agree substantially. Accuracy is particularly high when respondents are asked to classify their CD4 count as being above or below 200 cells, and between 201 and 350 cells and other counts. This is not surprising, since the threshold of 200 cells has both clinical significance and is often related to benefit eligibility. Self-report of VL counts is less accurate than CD4. The agreement is, however, still substantial. Eighty-four percent of the time, the patient are able to say whether they have a VL detectable or not. The better reliability of self-reported CD4 count over self-reported VL is consistent with the findings of Kalichman et al (2000).

Overall, our findings are comparable with the findings of Kalichman et al (2000) and Cunningham et al (1997). Moreover, we found a higher reliability of CD4 counts, and VL than did Kalichman et al (2000).

The agreement between the self-report and the laboratory results for CD4 is worse among those who have taken part in an intensive adherence program compared to those who did not. A possible explanation is that any discrepancy between self-report and the laboratory results may influence the kappa widely as only 6 people in our sample took part in the program. Indeed, only one person out of the 5 who reported a CD4 range disagrees with the lab result. Regarding the VL, however, the agreement is perfect among those enrolled in the Program compared to those not enrolled in the Program. Furthermore, the Program only enrolls patients likely to be non-adherent to medications, and those patients may be less likely to know their clinical markers.

Concerning the self-report of medications, in 74% of the time participants reported the exact medication regimen, which is considered to be a substantial agreement. Exact report of medication by class of medications and by the number of medications was also substantial. Given that CHAIN analyses predicate their definition of HAART therapy on self-reported medications, these findings support the validity of these measures.

Self-report of opportunistic infection differs from the medical record. The agreement is moderate. Surprisingly, several CHAIN participants have reported opportunistic infections that were not recorded in the medical record. This finding raises the question of the accuracy of medical records for the reports of previous medical conditions (some of the information in the medical records may have been self-reported by the patients). The fact that five people have reported having suffered from mycobacterium avium-intracellular infection, and only one medical record has reported such a disease is particularly surprising, as well as the finding that nine people have reported recurrent bacterial pneumonia and only one has been diagnosed with this condition according to the medical record, and one person reports having suffered from a Kaposi sarcoma and this was not mentioned in the medical record.

Twelve patients have reported having suffered from wasting syndrome and no medical record reports such a condition. This suggests two possible explanations: either the medical records are not accurate and reliable, or the people think they have been affected by the condition because they have lost weight, but in fact do not fulfill the medical criteria.

Self-report of opportunistic infections in the last 12 months is not more reliable than self-report of lifetime opportunistic infections. This is surprising and unexplained. Clients were more likely to report recent bouts of wasting syndrome and chronic diarrhea that were not documented in the medical chart. This suggests an over report of opportunistic infections by CHAIN participants. As with lifetime occurrences, people may report they suffer from a particular condition although they have not been diagnosed as such. One likely explanation for the over report of chronic diarrhea may be its consequence as a side effect of specific HIV medications.

Self-report of sexually-transmitted diseases is fair to moderate according to the kappa coefficient, despite a percentage of agreement of 71%. Concerning the self-report of some other medical conditions, people seem to report more often other diseases, such as diabetes mellitus, hypertension, etc, except for hepatitis. It is probable that difference between self-report and medical record report are explained by both inaccuracies in the medical record, and patient self-diagnosis.

Those who reported CD4 counts, VL counts, and medication regimen that differ from the reports in the medical records did not differ in terms of income, education, and ethnicity. Those who reported the exact CD4 and VL counts tend to be younger, and although not significant, women tend to more accurately report their CD4 and medication regimen.

The main limitation of this pilot study is the small sample of participants, and the non-representativeness of the sample in comparison to the overall cohort. Another limitation is the use of medical record extraction as the gold standard. Medical records may be inaccurate and may have missing information, depending on how well the physicians assess the history of the patient and on how well he or she has reported the information. Furthermore, medical history often relies on self-report.

Overall, the validity analyses illustrate that self-reported CD4 and VL counts are highly reliable. The definition of highly active antiretroviral therapy is also reliable, and thus analyses based on this definition can be confident of the definition that have been used.

## CONCLUSION

Self-report of demographic characteristics, CD4 and VL counts, and HIV medications are valid in our analysis. It implies that i) one can confidently use such variables for analyses, and ii) people are well informed of their disease status by their physician. Self-report of opportunistic infections and other diseases, however, differs from the medical record, possibly due to inaccuracies in the medical record and self-diagnosis among the patients. Variables based on these self-reports should thus take this into account.

If the medical record underestimates other diseases such as diabetes mellitus, hypertension, etc, this does not affect CHAIN analyses, nevertheless, this may have important repercussion for the health of the patients. Indeed, patients taking antiretroviral therapy are particularly at risk for cardiovascular diseases. It is more likely, however, that the over report of diabetes and other chronic diseases is not adversely affecting patient, since physical examinations and blood examinations are frequently conducted.

Despite the fact that this study is based on a small sample of people and thus lacks broader generalizability, it does demonstrate interesting findings that would be worth further investigation. For the CHAIN Project, it shows that most of the information reported by participants is reliable and valid. The only area where caveats are in order involves the self-report of medical conditions. Another limitation of this study is the fact that the analysis has been conducted among a selected sample of patients attending an academic clinic, and thus the patients may receive more information than elsewhere. The population of this clinic is also mainly Latino and thus it is not representative of the entire population of HIV positive adults in New York City.

APPENDIX A: Description of the different tests used in the reliability/validity analysis

Test name	Description	Significance of findings	Advantages and disadvantage of the test to access reliability/validity
% agreement	% of the cases in which both the self-report and the medical record agree	The higher the % of agreement, the higher the reliability.	- Chance not taken into account
Kappa	Fraction of the observed agreement not due to chance in relation to the maximum nonchance agreement, or:  $k = \frac{P(\text{observed}) - P(\text{expected by chance})}{1 - P(\text{expected by chance})}$	Landis & Koch agreement scale: 0.8-1.0: almost perfect 0.6-0.8: substantial 0.4-0.6: moderate 0.2-0.4: fair 0.0-0.2: slight -1.0-0.0: poor	- Chance taken into account - Dependent on the prevalence of the people who are truly positive.
Intraclass coefficient	Proportion of the total variability in the measured factor that is due to the variability between individual.	The higher the coefficient, the better the reliability.	- Systematic differences between self-report /medical record report taken into account. - Affected by the range of values.
Paired t-test	Measures the mean difference between self-report and medical record report, and test if this difference is statistically different from 0.	P<0.05: fails to reject the hypothesis that self-report differs from medical record report	- Dependent on extreme values
Wilcoxon signed-rank test	Non-parametric equivalent to paired t-test: ranks each values separately for the self-report and the medical record, measures the average difference in rank, and tests if this difference is statistically different to 0.	P<0.05: fails to reject the hypothesis that self-report differs from medical record report	- Lack of power
Pearson r	Measures the strength of the linear relationship between self-report and medical record report.	The higher the coefficient, the higher the reliability.	- Not a good measure of agreement, but widely used. - Insensible to systematic diff. - Sensitive to the range of values
Spearman r	Measures the degree to which the values have the same rank order	The higher the coefficient, the higher the reliability.	- Less influenced by outlying values than the Pearson r. - Insensible to systematic difference

APPENDIX B. Findings of the validity analysis: summary

Variable of interest	n	% agreement	Kappa	Degree of agreement (Landis / Koch)	Intraclass coefficient	t-test or equivalent	Pearson r	Spearman r
<b>Birthday and date of diagnosis</b>								
Birthday	34	99	0.98*	Almost perfect	0.997			
Month and year of diagnosis	34	92	0.76*	Substantial	0.86			
<b>Last CD4 count</b>								
CD4 divided in 4 categories ( $\leq 200$ , 201-350, 351-499, $\geq 500$ )	32	91	0.76*	Substantial	0.77			
CD4 $\geq 500$ vs. CD4 < 500	32	84	0.64	Substantial				
CD4 $\geq 200$ vs. CD4 < 200	32	94	0.76	Substantial				
CD4-exact count	26				0.95	No diff.	0.94	0.74
<b>Last VL count</b>								
VL categories (0, 1-4999, 5000-9999, 10000-29999, 30000-99999, 100000-499999, $\geq 500000$ )	25	88	0.62*	Substantial	0.58			
VL detectable vs. undetectable	25	84	0.68	Substantial				
VL-exact count	25				0.61	No diff.	-**	0.63
Log VL-exact count	25				0.64	No diff.	-**	0.63
<b>Current HIV medication</b>								
PI exact regimen	34	79	0.71	Substantial				
NNRTI exact regimen	34	97	0.90	Almost perfect				
NRTI/ NtRTI exact regimen	34	82	0.78	Substantial				
Number of medication	34	93	0.75*	Substantial				
Exact regimen	34	74	0.70	Substantial				
<b>Opportunistic infections</b>								
Opportunistic infection- ever	34	85	0.47	Moderate				
Opportunistic infection- in the last 6 months	34	79	0.48	Moderate				
<b>Other medical conditions</b>								
STD-ever	34	71	0.40	Fair to Moderate				
STD in the last 6 months	34	100	1.00	100% agreement				
Other medical conditions (range)	31	62 to 97	0.26 to 0.87	Fair to almost perfect				

Legend: \* weighted kappa, \*\* not-normally-distributed, n= number of persons included in the analyses

## **REFERENCES**

Cunningham WE, Rana HM, Shapiro MF, Hays RD. Reliability and validity of self-report CD4 counts in persons hospitalized with HIV disease. *J Clin Epidemiol* 1997;50:829-835

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

Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159-174

MacLure M, Willett WC. Reviews and commentary: misinterpretation and misuse of the kappa statistic. *Am J Epidemiol* 1987; 126: 161-169

Szklo M, Nieto FJ. *Epidemiology: Beyond the basics*. 2000, Aspen Publishers, Inc; Maryland, United States