

**Comparison of Selected Outcomes Reported from Ryan White-funded Agencies
to Clinic-based Data from Medical Chart Review
(2003-2004)**

JSI Research and Training Institute, Inc.

Laureen M. Kunches ANP, PhD

Nancy E. Reinhalter RN, CCRC

Joseph Musolino

Lisa R. Hirschhorn MD, MPH

For the Boston Public Health Commission

2007



Table of Contents

Background	3
Methods	4
Results	6
A. CD4 Counts.....	6
B. HIV Viral Load.....	9
C. Ability to Adhere to Medical Therapies.....	10
D. Level of Depression.....	14
Conclusions	17
Acknowledgements	19
References	19

This project is supported by funding provided by the Boston Public Health Commission through the Ryan White Title I Comprehensive AIDS Resource Emergency Act of 1990, as amended by the Ryan White CARE Act in 1996, 2000.

For more information:

Laureen Kunches PhD
JSI Research and Training Institute, Inc.
44 Farnsworth Street
Boston, MA 02210
617-482-9485
lkunches@jsi.com

Background

Two of the activities that have been funded over the past several years by the Boston Public Health Commission's AIDS Program involve assessment of client outcomes by Suffolk University's Center for Public Management and clinical care quality management by JSI Research and Training Institute. Data from a range of Title I funded agencies are used for these two activities, and some of the factors assessed are similar. However, the methods of ascertainment are different. Suffolk's outcome data are gathered by providers using a variety of secondary sources, including client self-report and records in agency documents, while JSI's quality data come directly from the laboratory test reports and notes in the medical records. Research has shown that memory factors and "social desirability" can affect the accuracy of clients' self-reported measures¹. Others have suggested that self-report should be used with caution in populations with limited education or cognitive impairment². Another potential source of bias is the possibility that agency staff might translate this subjective information inaccurately in their reports. Based on these concerns, we initiated this exploratory analysis.

The recently completed comparison of Boston Public Health Commission client services data to JSI chart review data involved matching clients from the two sources on their date of birth, gender and last four digits of social security number. For this analysis, we used the group of clients found in the JSI and BPHC data and matched them to the Suffolk database for three comparable 6-month periods between March 2003 through August 2004. The individual outcome factors that are common to the two sources include three "Health Outcomes" (*CD4 counts, HIV Viral Load, Ability to adhere to medical therapies*) and one "Quality of Life Outcome" (*Level of Depression*). Assuming that the clinic from which JSI's data was obtained represented the primary site of medical care for the client, the medical record data would be considered the most accurate source of the first two of these outcomes (i.e., CD4 and viral load levels). In contrast, because adherence and depression are defined and determined somewhat differently, each source is equally valid. Consequently our comparison of these findings is offered as background and context, rather than an accuracy assessment.

This report describes the results of this matching and comparison process and addresses the following important questions for the first time:

- How accurate are the secondary sources of outcomes when compared to primary data from medical record review?
- What patient factors predict accuracy of the self-report data?
- How do agencies differ in rates of accuracy?
- For instances of disagreement, what patterns are seen in the extent and direction of differences?

Methods

Comparison of databases was accomplished in two steps:

Step 1 --- To identify clients who are part of the two foundational data sources (i.e., BPHC service utilization and JSI clinic chart review), we matched the two complete FY 2004 databases on the basis of 3 specific variables shared across the systems:

- client date of birth
- last 4 digits of Social Security Number
- gender

Clients had to match on all 3 variables to be included in the final dataset. If a client was an exact match, this was considered confirmation that they used at least one of the Title I supported services in the Boston EMA network during FY 2004.

Step 2 --- JSI shared the client codes for the matching group with Suffolk, who searched for outcomes data in their system from these individuals. For matching clients, Suffolk shared the outcomes reports submitted on the matching clients for three 6-month periods --- March to August 2003, September 2003 to February 2004, and March to August 2004. These intervals were the ones that overlapped with information in the JSI database. Since individual clients have outcomes reported from multiple agencies during the same period, we used outcomes reports (rather than individual clients) as the unit of analysis. Approximately two-thirds of clients had a single agency reporting outcomes for the 6-month period; for the other third, there were multiple reports and these are included as separate outcome episodes. The following table summarizes the relationship of individual clients to the frequency of outcome reports for the 3 time intervals analyzed here:

Number of reports	3/03 to 8/03	9/03 to 2/04	3/04 to 8/04
1	265	277	318
2	92	89	90
3	28	46	40
> 3	40	40	55
TOTAL CLIENTS	425	452	503

We looked for a matching value for the Suffolk outcome measure in the JSI chart review data during the corresponding time interval. If there were multiple laboratory values in the JSI data, having any matching

value was considered “agreement”. If no value was found in the JSI data for the timeframe, the earlier time period was searched, to accommodate potential delays in clients receiving lab results from their clinic. Clients who had information in the Suffolk outcomes data but no comparable data in the JSI chart reviews were deleted from the analysis. For the agency-level analysis, only agencies that reported on 25 or more clients for the outcome were included, since smaller numbers do not provide an adequate base of comparison.

To illustrate agency specific rates, we classified the agencies by type into the following 3 groups, based on the agency services, but not limited to Ryan White funded services:

Medical --- sites where medical care is provided

Mixed support --- sites with multiple services, but not medical

Single --- only one type of service is provided (e.g., food and meals, complementary therapies)

Since case management is a pivotal service, the medical and mixed support categories that also include case management are noted by an asterisk.

Chi-square statistics were used to determine relationships between categorical level variables including patient demographics and agreement on outcome factors.

Results

A. CD4 Counts

The Suffolk outcomes measurement groups the most recent CD4 count into 4 levels:

- Excellent --- greater than 500
- Good --- 200-500
- Fair --- 50-199
- Poor/In Crisis --- Less than 50

We searched the laboratory values gathered in the JSI chart reviews over the same time period for a matching value (according to the 4 Suffolk categories). Having any value that matched during the period was considered “agreement”, even if there were other values that fell into a different category.

The following describes the extent of agreement we found per time period and summarizes the observed relationship of agreement to patient and site factors.

1. Overall accuracy

Overall, 76% (748/989) of the CD4 outcome reports were consistent with values found in the medical records for the same time periods. As shown in Table 1, the rates for the three 6-month reporting periods showed a trend toward improvement over time ($p=0.09$); the number of individual CD4 reports were very similar for each period.

Table 1. Rates of CD4 Outcome Agreement by Time Period

Period	Total CD4 Outcome reports	No. in Agreement with chart reviews	% Agreement
Mar-Aug 2003	332	238	72%
Sept 03-Feb 04	324	247	76%
Mar-Aug 2004	333	263	79%

When differences in CD4 levels were found, Suffolk CD4 values were lower than chart review information 70% of the time. Nearly 60% of the discrepancies were situations where the chart review result was 1 category better (i.e., higher CD4 count) than the outcome value reported to Suffolk, resulting in an under-estimate by Suffolk data; this was found in 142/241 instances of disagreement. The second most common difference was 1 category lower in the chart review than outcomes data

(66/241 or 27%). Ten percent (24/241) were a two category difference where Suffolk data was lower. Only 8 instances (3%) found chart review data to be 2 or 3 categories lower than Suffolk data. There were no significant differences in these trends by time period.

2. Patient factors

We evaluated a range of patient demographic characteristics as well as co-morbidities and clinical factors to identify significant relationships with CD4 discrepancies ($p < 0.05$ by chi square). Rates of agreement were significantly lower for Blacks (70% vs. 77% for Whites and 79% for Hispanics) and US-born persons (74% vs. 80%). A trend ($p = 0.07$) was noted for persons who had risk behaviors other than MSM, where the agreement rate is 74% compared to MSMs (80%).

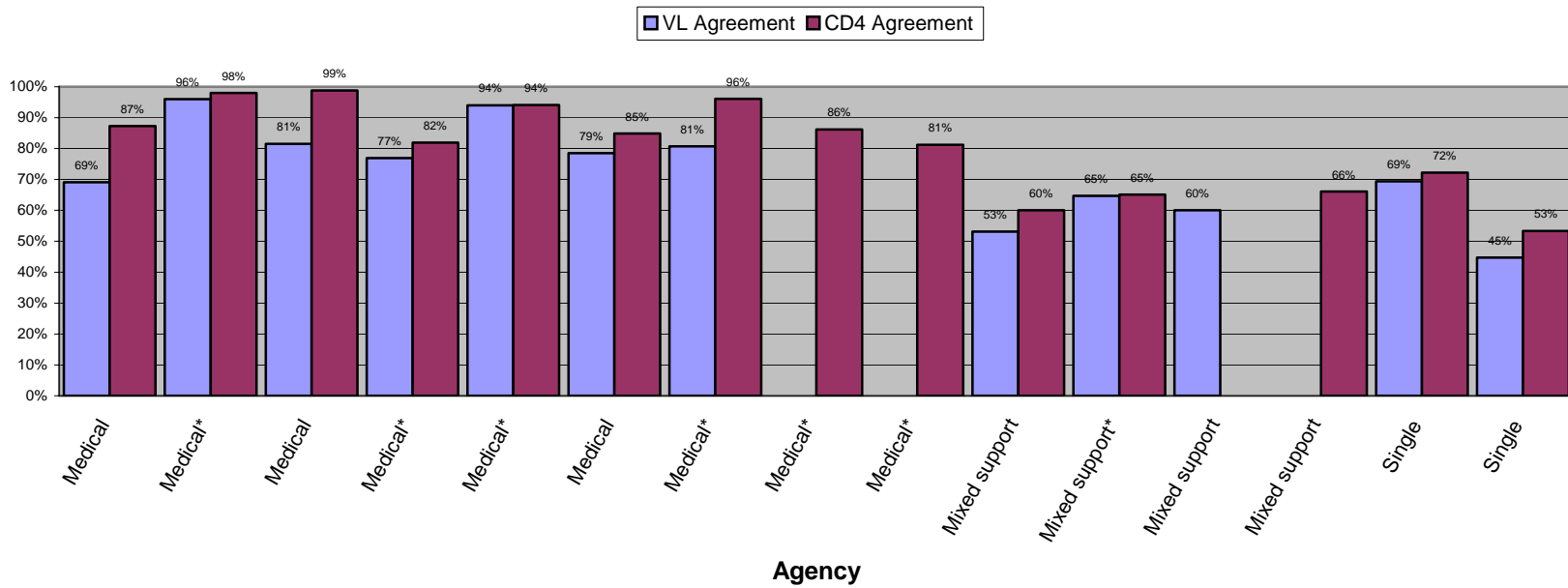
Looking at the direction of these differences, there was a suggestion that Black clients were less likely to be under-estimated by Suffolk data (64%) than Whites (79%), with Hispanics falling in between (68%); however, this difference was not statistically significant ($p = 0.09$). No other significant patterns were noted for the direction of differences.

3. Individual Agencies

The medical programs reported CD4 counts that were in close agreement to JSI chart data. Ninety-two percent of the instances were in agreement, compared to only 65% from the non-medical sites ($p < 0.0001$). The following figure illustrates the rates of agreement on laboratory values for each agency that reported on at least 25 clients, sorted by agency type. Individual agencies are not identified, since there was no agreement in advance concerning this analysis and the findings are considered exploratory. Note that sites could select among outcomes so not all reported on both CD4 counts and HIV viral load levels.

The direction of the differences was similar between types of agencies.

Figure 1. Agreement between Laboratory Outcomes from Secondary Sources and Medical Chart Reviews



*indicates agency provides case management service

B. HIV Viral Load

The 4 categories for the most recent viral load values (used in the Suffolk outcomes measurement) were:

- Excellent --- less than 75/undetectable
- Good --- 75 to 9,999
- Fair --- 10,000 to 100,000
- Poor/In Crisis --- greater than 100,000

As with the CD4 values, agreement was defined as having any value from the chart review during the period that matched the category, even if there were other values that fell into a different category. The results of this matching process are described below.

1. Overall accuracy

Across the 18-month period, 70% of the viral load outcomes were in agreement with the chart reviews (624/891). In contrast to the CD4 outcomes, there was no difference in accuracy between the reporting periods (see Table 2).

Table 2. Rates of HIV Viral Load Outcome Agreement by Time Period

Period	Total Viral Load Outcome reports	No. in Agreement with chart reviews	% Agreement
Mar-Aug 2003	279	196	70%
Sept 03-Feb 04	300	213	71%
Mar-Aug 2004	312	215	69%

Reviewing the direction of the differences, we found 60% were cases in which Suffolk viral loads were higher (e.g., clinically worse) than the chart review reports. As with CD4 counts, 50% of the discrepancies were situations where the chart review result was 1 category better (i.e., reflecting a lower viral load) than the reported outcome value, resulting in an overly negative report in the Suffolk data; this was found in 134/267 instances of disagreement and the majority of these came from non-medical sites (80 or 60). The second most common difference was 1 category worse in the chart review than outcomes data (84/267 or 31%). Two category differences (higher or lower) each accounted for approximately 8% of the discrepancies, with only 7 others being 3 category differences.

2. Patient factors

The significant predictors of having less agreement of viral load outcomes with chart review levels were female gender (67% vs. 73%, $p < 0.03$), being US-born (68% vs. 76%, $p = 0.01$), and having active mental illness (67% vs. 73%, $p = 0.05$). No relationship was seen with age, race/ethnicity, risk behaviors, or active substance abuse. Direction of difference did not vary significantly on the basis of gender, country of birth or mental illness.

3. Individual Agencies

As shown in Figure 1, there was a wide range of disagreement rates across the 16 agencies and the highest rates of agreement were found in the 6 medical sites reporting this measure (4% to 31%). Patterns seen in the direction of difference were not statistically significant.

C. Ability to Adhere to Medical Therapies

The Suffolk outcomes measurement tool categorizes this information as follows:

- Excellent --- routinely adheres to medical therapies
- Good --- frequently adheres to medical therapies
- Fair --- erratically adheres to medical therapies
- Poor/In Crisis --- rarely adheres to medical therapies

In the JSI chart review process, antiretroviral (ARV) adherence problems recorded by the medical providers are noted, creating a “yes/no” status rather than an estimate of the degree. Because JSI uses calendar years as the basis for the ARV information, the comparison data is shown for 2003 and 2004. Since Suffolk uses all medical therapies (not just HIV medications) and neither group defined the levels of adherence quantitatively, our comparison of the two information sources is exploratory and descriptive.

We first determined the ARV status of the clients for who had their medical therapy adherence outcome reported. The results are shown in Table 3 for the two years. Approximately 15-20% of the clients with Suffolk adherence reports were not taking ARVs during the year, and so their information may be reflecting other medication adherence.

Table 3. Antiretroviral Status of Clients for whom Adherence to Medical Therapies was reported

	2003		2004	
	#	%	#	%
On ARVs	84	79%	112	85%
Not eligible for ARVs	14	13%	14	11%
Eligible but not taking	8	8%	5	4%
Total	106		131	

As shown, adherence outcomes were reported to Suffolk on 106 clients (with matching JSI data) in 2003 and 131 in 2004. Based on JSI clinical data, 22 of 106 (21%) and 19 of 131 (14%) were not on ARV's in these years; the majority of these clients were not eligible for ARV treatment based on their laboratory markers and clinical status.

To make the reported data from the two sources most comparable, we focused the comparison on the subgroup of clients on ARVs (since JSI data is restricted to this group). We looked at medical record evidence of ARV adherence problems (classified as yes or no in JSI's data) within each of the four Suffolk adherence categories. The following two figures depict the rates of medical chart adherence problems in each group, by year.

Approximately half of the clients for whom adherence was reported as an outcome were rated as "excellent" --- 46/84 in 2003 and 55/112 in 2004. The Suffolk adherence rates for each year of "good" (26% and 37%), "fair" (18% and 13%) and "poor" (1% and 2%) suggested a trend toward improvement in the second year. However, in both years, nearly ¼ of the clients reported to be "excellent" adherers had evidence in their medical record of ARV adherence problems. As the degree of adherence declined in the Suffolk outcomes data, the proportion who were also listed as having adherence problems in the medical record increased, but the correlation was still relatively weak and not statistically significant. In other words, we found indications that adherence reports were inconsistent with medical chart data across all levels of Suffolk adherence outcomes. The data also suggest that medical providers may under-recognize adherence issues, based on the fact that ARV adherence problems were noted in only 43 and 60% of clients rated as "fair" on adherence outcomes during the 2 years studied.

Figure 2. ARV adherence problems (JSI) by Adherence to medical therapies outcome (Suffolk) 2003

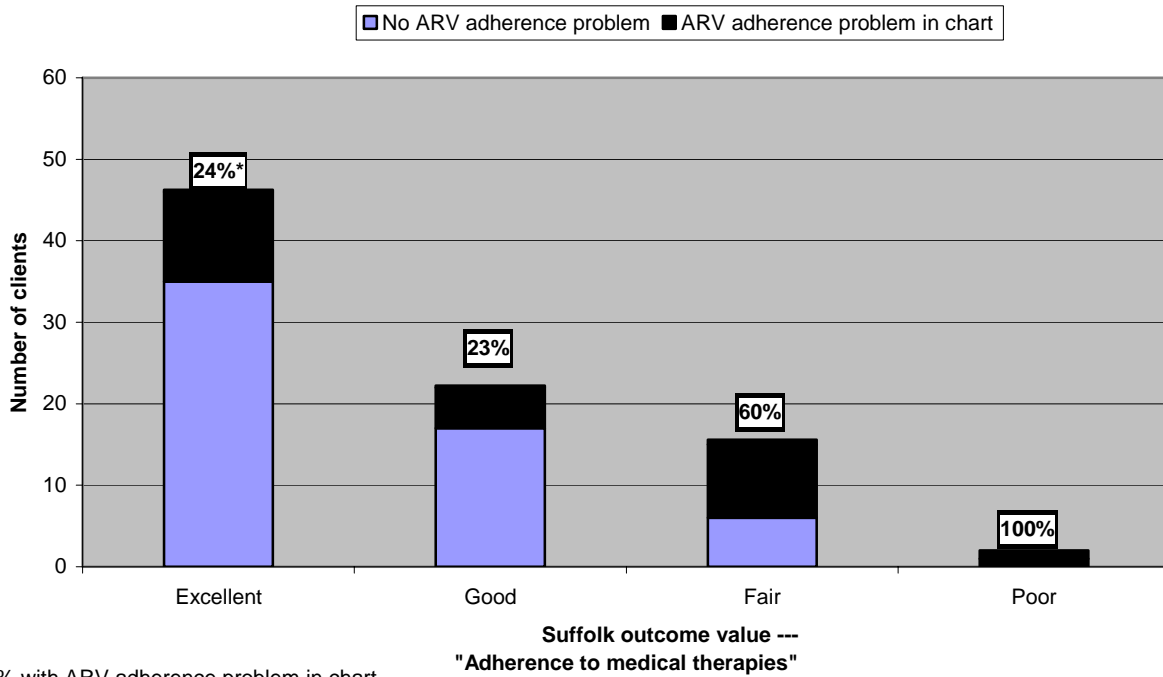
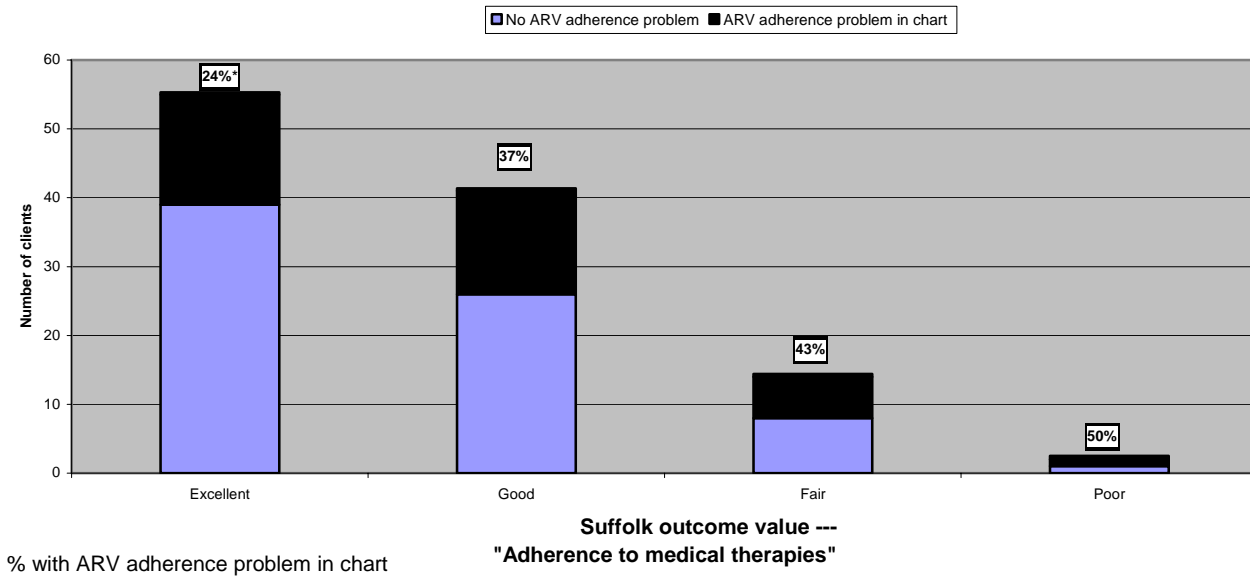


Figure 3. ARV adherence problems (JSI) by Adherence to medical therapies outcome (Suffolk) 2004



Next we focused on agreement in clients with "excellent" rating on Suffolk measure, as defined by having no identified adherence problems noted in the chart review. This was the group of adequate size to evaluate patient factors that might be related to concurrence of the sources. [Note: Fair or poor adherence levels accounted for only 16 clients each year, of whom 7 and 10 had adherence problems in the chart review data.]

Overall, agreement between the sources (for the group rated as "excellent") was found in 35/46 (76%) in 2003 and 39/55 (71%) in 2004. This represented 42% and 35% of the group with adherence data in both sources for 2003 and 2004, respectively. There was no significant difference between medical and non-medical sites in the rates of this factor.

Patient characteristics related to more frequent agreement between the sources on high-level adherence included:

- Being born outside the US (60% vs. 36% in 2003, $p=0.05$)
- First CD4 count above 200 (50% vs. 19% in 2003, $p=0.01$)
- First viral load less than 400 (51% vs. 27% in 2003, $p=0.03$)
- No incarceration between 1999-2004 (46% vs. 0% in 2003, $p=0.01$)
- No IDU risk behavior (44% vs. 23% in 2004, $p=0.02$)
- No active mental illness (50% vs. 27% in 2004, $p=0.02$)

Our next step was to create a variable that captured situations where the Suffolk outcome reports over-estimated adherence (compared to the JSI chart review data). We defined this situation as clients where JSI found ARV adherence problems but their outcome rating was "excellent or good". The over-estimated outcome was found in 19% of the 2003 cases and 28% of the 2004 cases. The only factor associated with increased likelihood that Suffolk outcomes data over-estimated adherence compared to the chart review data was active substance abuse in 2004 (46% vs. 22%, $p=0.01$). All other patient factors (age, gender, risks, race/ethnicity, country of birth, incarceration, active mental illness, clinical status) were not significantly associated with over-estimating adherence. Medical sites were less likely to over-estimate (14% vs. 30% for non-medical in 2003 and 22% vs. 35% for 2004), but this trend did not reach statistical significance ($p=0.10$).

Due to the small number of cases in which Suffolk under-estimated adherence, no specific analysis of patient or site factors was possible.

D. Level of Depression

Four categories are used to track this outcome in the Suffolk measurement tool, according to the following specification:

- Excellent --- Not depressed
- Good --- Light depression
- Fair --- Moderate depression
- Poor/In Crisis --- Severe depression

In the JSI chart review process, mental health diagnoses and psychotropic medications are documented for each calendar year; however, this was a “yes”/“no” assessment and the severity of depression was not recorded. The following summarizes the comparison of information from these different but related sources.

To examine the agreement of the two data sources, we looked at 2003 and 2004 using the following criteria:

- 1) “agreement client is depressed” included cases where JSI had depression as a diagnosis and Suffolk had light, moderate or severe depression.
- 2) “agreement client is NOT depressed” captured situations where JSI had no depression diagnosis and Suffolk had “not depressed”.
- 3) “Suffolk over-estimated depression” designated situations where JSI had no depression diagnosis and Suffolk had light, moderate or severe depression.
- 4) “JSI over-estimated depression” included cases where JSI had a depression diagnosis and Suffolk indicated “not depressed”.

The following Table illustrates these 4 situations:

Table 4. Relationship of JSI and Suffolk Depression information (2003-2004)

	2003		2004	
	#		#	
Agree client is depressed	43	41%	44	31%
Agree client is not depressed	14	13%	33	23%
Suffolk over-estimated depression	42	40%	63	44%
JSI over-estimated depression	5	5%	4	3%
Total	104		144	

Agencies reporting to Suffolk on depression appeared to choose this outcome in clients with some level of depression symptoms, since only 18% (19/104) in 2003 and 26% (37/144) in 2004 were rated as having

“none”. In clients with depression, the level (i.e., light, moderate or severe) showed improvement between 2003 and 2004. The first year, 31% were light and 40% were moderate (32 and 42 of 104, respectively); in 2004 38% were light and 26% were moderate (55 and 37 of 144, respectively).

There was overall agreement on the presence or absence of depression for 54% of the clients in each year (57/104 for 2003 and 77/144 in 2004). Most of the remaining cases (42/47 or 89% in 2003 and 63/67 or 94% in 2004) were situations in which Suffolk data alone was found to indicate the presence of depression.

Next we looked at the rates of depression diagnoses in each of the 4 Suffolk depression categories, by year; the tables on the next page illustrate these relationships.

The medical chart data for the two years showed depression was present in 26% and 11% of those rated as “none” in Suffolk’s data (5/19 and 4/37, respectively). For those reported to Suffolk as having some depression, agreement in the medical chart was highest in the “moderate” (60% and 43%) and “severe” (73% and 53%) categories. Roughly 1/3 of the “light” group in each year were noted to have depression by chart review.

Figure 4. Depression in medical chart (JSI) by Level of Depression outcome (Suffolk) 2003

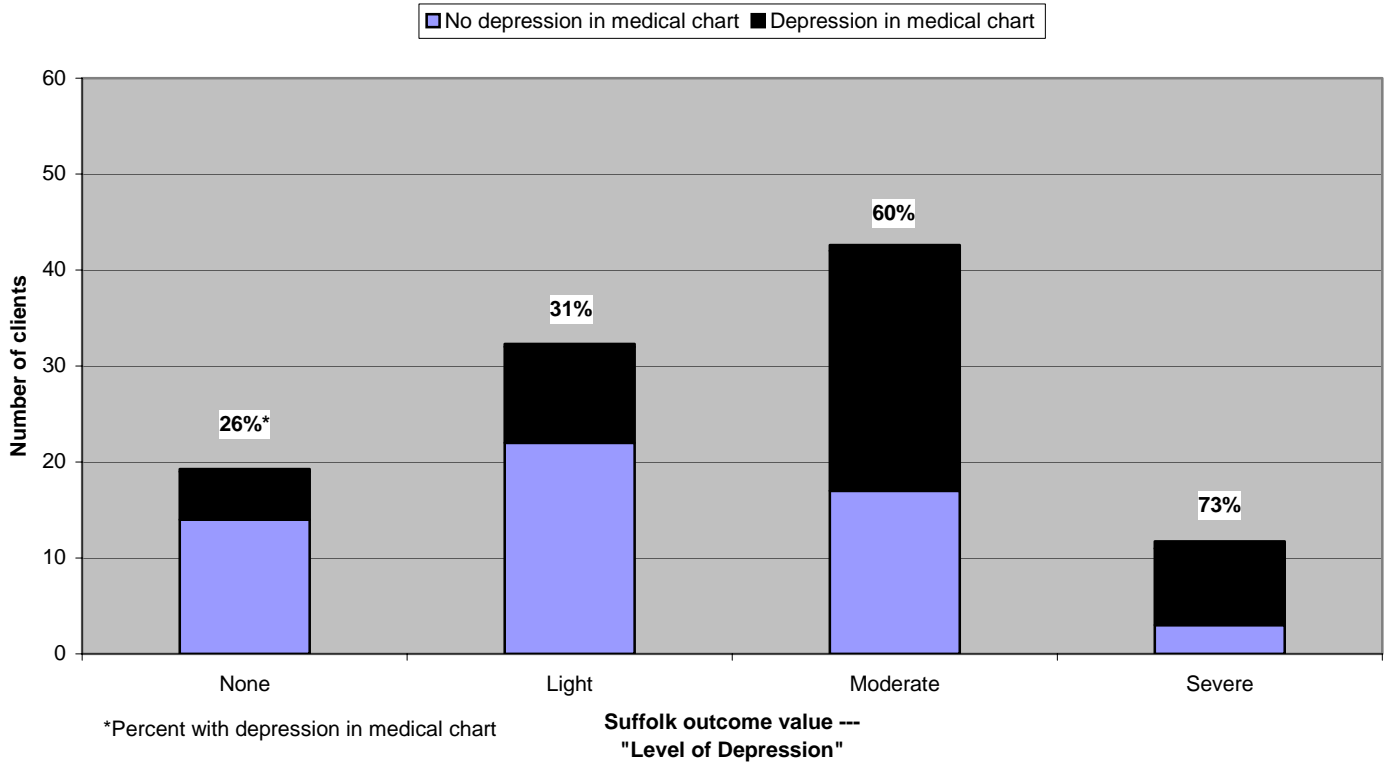
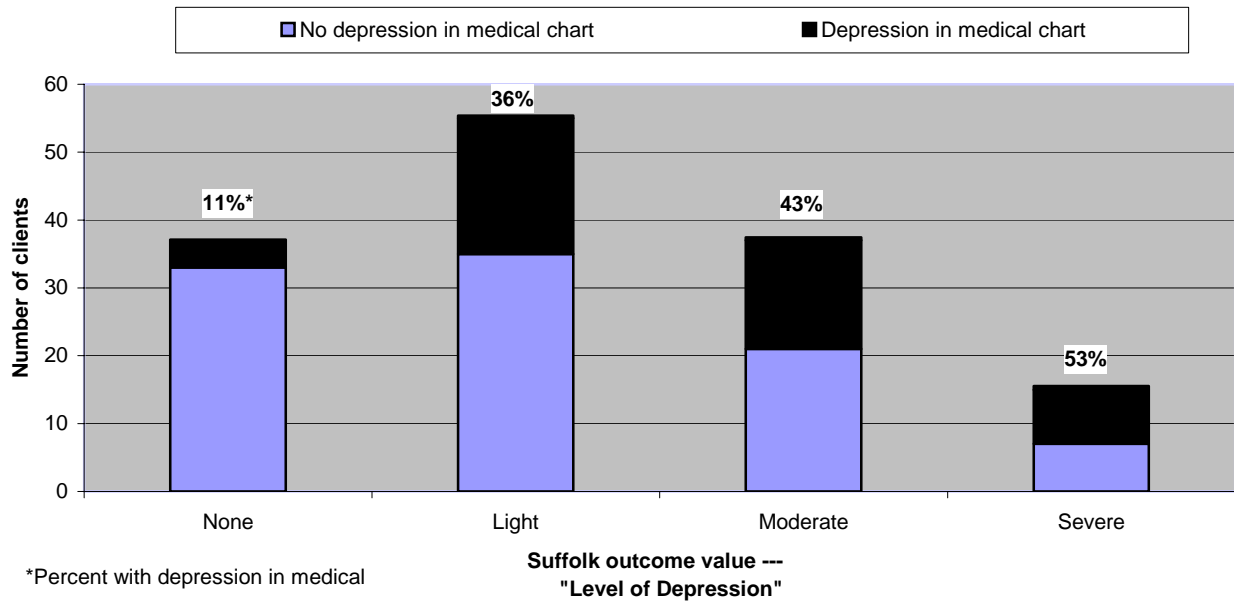


Figure 5. Depression in medical chart (JSI) by Level of Depression outcome (Suffolk) 2004



Small numbers of cases with data for this outcome limits our ability to assess relationships to patient factors. However, we noted that medical sites had higher rates of agreement that client was depressed in 2003 (47% vs. 30% non-medical, $p=0.06$) but not in 2004. Also, in the small number of cases where JSI over-estimated depression (9 cases in 2 years), the outcome data that was inconsistent was reported by non-medical sites in all but 1 instance, substantiating the overall accuracy of the medical site outcomes.

Conclusions

This analysis represents the first comparison of client outcome data submitted by Ryan White-funded agencies (Part A) with a relatively objective standard --- data from medical records in the clinic where they receive their HIV treatment. Unfortunately, differences in definitions and measures pose limitations to the adherence and depression information, but the laboratory measures (CD4 and viral load) are generally comparable. Our definition of agreement was generous in the laboratory measures, counting any matching value as a consistent finding, even when other values from the period might be different. Although the time intervals overlap, exact dates are not part of the available data so there is a possibility that outcome status is changing during the period and the reported information captures different points on the trajectory. Nevertheless, several interesting findings described here should prompt further discussion and may be helpful in the outcome measurement process going forward.

The CD4 counts were relatively accurate at 75%, and showed improvement during the 3 time intervals. Viral load reports were somewhat less consistent (approximately 70%) and not improving over the period. Not surprisingly, medical sites reported with significantly greater agreement than non-medical agencies for both CD4 counts (91% vs. 65%) and viral loads (82% vs. 62%). It is likely that outcome assessments in the medical sites may tap directly into clinical data rather than relying on secondary sources such as patient self-report. This confirms the importance of restricting the analysis of individual changes in these factors over time to reports from the same agency or type of site.

Differences of one category in either direction were the most common, accounting for 87% of the discrepant CD4 counts and 82% of the discrepant viral loads. Surprisingly, the direction of difference in the laboratory outcomes tended to be that medical chart data was better (i.e., clinically more favorable) than Suffolk data. This is contrary to published studies of patient self-report where individuals tend recall their numbers in a favorable light.

Patient factors associated with lower agreement on laboratory measures included being US born for both CD4 and viral load. Blacks had lower agreement on CD4 counts, while women and clients with active mental illness had lower agreement on viral load reports.

In the case of treatment adherence, we were surprised to find that 1 in 5 of the clients reporting this outcome was not receiving antiretroviral therapy during the timeframe. Although the outcome was defined broadly as “medical therapies”, measures have traditionally emphasized adherence to the HIV medications since treatment can be challenging and demands near-perfect observance of the multidrug regimen. The subsequent comparison of data sources was limited to clients on antiretroviral medications, which was the only group that JSI had recorded adherence information.

The categorization of adherence in the Suffolk outcomes was not quantitative, making the direct comparison with JSI’s “adherence problem” somewhat subjective for the second best rating --- “good” or frequently adherent. Therefore, we limited the analysis to the most clear cut relationships. But within the Suffolk outcome of “excellent” routinely adherent clients, 1 in 4 was noted to have adherence problems in their medical record during the comparable year. The lack of specific dates during these 6 and 12-month intervals makes drawing firm conclusions unwise, but this is an issue for further consideration in design of outcome assessment.

Finally, the correlation between depression levels in the Suffolk outcomes and the presence of a depression diagnosis in the chart must also be considered with caution, given the lack of standardization across the projects. Nonetheless, there were a substantial number of clients reporting moderate to severe depressive symptoms in the outcome data who had no record of depression at their clinical site. These inconsistencies led to Suffolk over-estimating the extent of depression compared to the medical provider notes, which may signal a low recognition of depressive symptoms by medical providers in some clients and sites, and warrants further study.

Overall, the inconsistencies between the two data sources reveal no strong patterns or systematic biases other than the differences between medical and non-medical sites, which would be expected. Other discrepancies appear to be random events that can be largely explained by the differences in definitions and use of broad time windows for comparison. The general trend observed was that information flowing to Suffolk on laboratory and depression outcomes paints a more negative picture of client status than one would get from direct medical record information. Consequently, concerns that “human nature” might lead agency staff to report outcomes in an overly positive light as a means to confirm the impact of their services are not supported by this analysis.

Acknowledgements:

The authors thank our colleagues in the Suffolk University Center for Public Management, in particular Nicole Rivers, for sharing their data with us for this report. In addition, we are indebted to the staff of the participating agencies and HIV clinics, the Boston Public Health Commission HIV/AIDS Program, and the Massachusetts Department of Public Health HIV/AIDS Bureau, without whose assistance this work would not have been possible. Special thanks to Michael Goldrosen, Eileen Harrington and Sharon Asonganyi for their support and helpful feedback.

References:

¹ AS Stone, JS Turkkan, CA Bachrach, JB Jobe, HS Kurtzman, VS Cain Editors. *The Science of Self-Report: Implications for Research and Practice*. Lawrence Erlbaum Associates, Mahwah, New Jersey 2000.

² WE Cunningham, HM Rana, MF Shapiro and RD Hays. Reliability and validity of self-report CD4 counts in persons hospitalized with HIV disease. *J Clin Epidemiol* 50;7:829-835, 1997.