

RESULTS OF THE HOMELESS YOUTH DATA AUDIT

AUGUST 2002

A TECHNICAL REPORT FOR THE
OFFICE OF SCHOOL AND COMMUNITY PARTNERSHIPS

REPORT PREPARED BY:
MATT NICE
SENIOR RESEARCH AND EVALUATION ANALYST
MULTNOMAH COUNTY
PERFORMANCE MANAGEMENT GROUP
503-988-3364

<http://www.co.multnomah.or.us/budget/eru/>

EXECUTIVE SUMMARY

The Performance Management Group (PMG) initiated a data audit to assess the quality of data in the system Homeless Youth Database (HYDB), before any analysis or evaluation planning could be performed. This step was taken as a quality assurance step in light of the recent service provider changes, the county's reorganizations, and because of PMG's new role as evaluator for the Homeless Youth Continuum. The audit analyzed samples of provider data, the HYDB data, database queries, past data analysis, and the quality of the available data. This report summarizes the findings, explains the impact, and offers possible solutions to the problems identified in the HYDB. Results identified several problems with missing data forms, the database queries, data quality, and fundamental assumption of the Homeless Youth Evaluation model, leaving us to conclude that true "outcomes" cannot be evaluated under the current model. This work was performed in accordance with the American Evaluation Association's generally accepted *Guiding Principles for Evaluators*.

FINDINGS

- Based on a random sample of provider files, missing files from the HYDB ranged from 2.5% to 6% of available data. Missing data appeared to be a more recent problem (in the last two quarters), especially with older youth, and likely due to reorganizations and a lack of consistent data management.
- Based on provider services reporting, an estimated 7.5% to 17% of the minimal files for youth who received some service coordination were missing from the HYDB (these may include some or all of the 2.5% to 6% from above).
- A limited logic model found that those personnel who performed follow-ups and actually contacted a youth would have a greater likelihood to identify a successful outcome—this is a significant bias to outcomes analyses.
- Past evaluation reports and conclusions have been based on mismatched intake, exit and follow-up data, in the common situation where multiple service episodes had occurred but not all service records were available (i.e., forms were missing).
- Past evaluation reports and conclusions have been based only on most recent youth data and failed to assess past services episodes data—In many cases this was substantial (10% or 175 youths appeared to have multiple service episodes).
- Limited missing data analyses identified a significantly higher rate of missing exit data than intake or follow-up—this was specific to older youth services.
- Of all youth with self reports, less than 13% (82) of those had intake, exit and 6-month follow-up data available.

Several corrective steps have already taken place to help insure that future results will be done so with the highest level of validity and confidence. In addition, several recommendations to address other issues are discussed. Results of the audit have underscored the complexities of this evaluation and highlighted what data is available, in what quantity, and the quality of the information. Regardless, serious thought must occur to prevent future reporting errors in this complex evaluation.

BACKGROUND

The entire Homeless Youth Continuum evaluation has been based upon data from providers. A variety of providers enter their data directly into the United Database (provider database) and subsequently print out individual forms for each of the youth served. These printouts are then mailed to OSCP by the middle of the following service month. These forms are then reviewed and manually entered into the HYDB. Missing forms or those with missing data are returned to the providers. This method of multiple hands on the data increases the likelihood for missing information and errors. The model of evaluation focused upon youth strength-based outcomes; however outcomes data has been virtually non-existent due to a variety of evaluation problems.

The Homeless youth data audit was broken into three parts—external provider data (missing forms), internal data processing (processes analyses) and data quality. Because of time constraints, not all data sets could be fully audited.

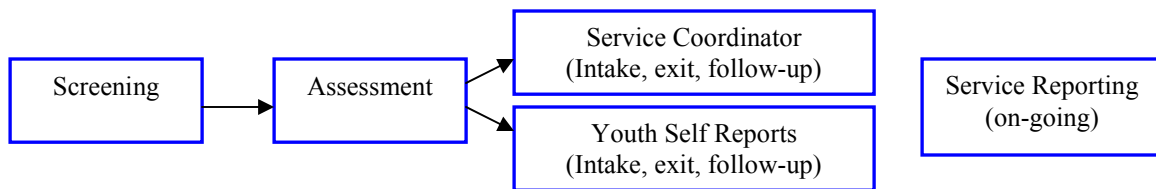


Figure 1. Reporting forms and their basic flow from providers to OSCP

There are five basic sets of forms for youth in the system (Figure 1)—initial screening and assessment forms (a first glance performed by JAAC), service coordinator forms (OI & NAFY case manager forms at intake, exit and follow-up), youth self reports (forms at intake, exit and follow-up) and service reporting forms (month services delivered). The homeless youth data system functions from the screening form—without which the youth data cannot be entered. Random provider file samples used a 95% confidence level and margins of error ranged from +/- 5.1% to 7.4%. Missing data can come in two forms: the form was never performed by the provider and/or the form was done by the provider but never entered into the HYDB.

EXTERNAL PROVIDER DATA

Screening Data. According to the HYDB there were 2,482 screenings from 1/99 to 6/02. A statistical sample of 324 cases was randomly selected from the JAAC case files.¹ Results found 21 cases (6.5%) which could not be found in the HYDB. This means that approximately 161 youth (from 37 to 285 youth) are missing their screening in the HYDB. They appeared to be missing around times of continuum change—the first being early 1999 when new service model and data collection methods were in place and again after a new service provider took over the downtown access and assessment center. These more recent problems occurred during the times of inconsistent data entry—from October

¹ Note that screenings used to take place at NAFY and OI. Those that may have been performed at NAFY or OI would not necessarily be captured in this sample. This sample had a margin of error of +/-5.1%.

2001 until January 2002 and possibly some late reporting (Figure 2). This increased the likelihood that errors went undetected.

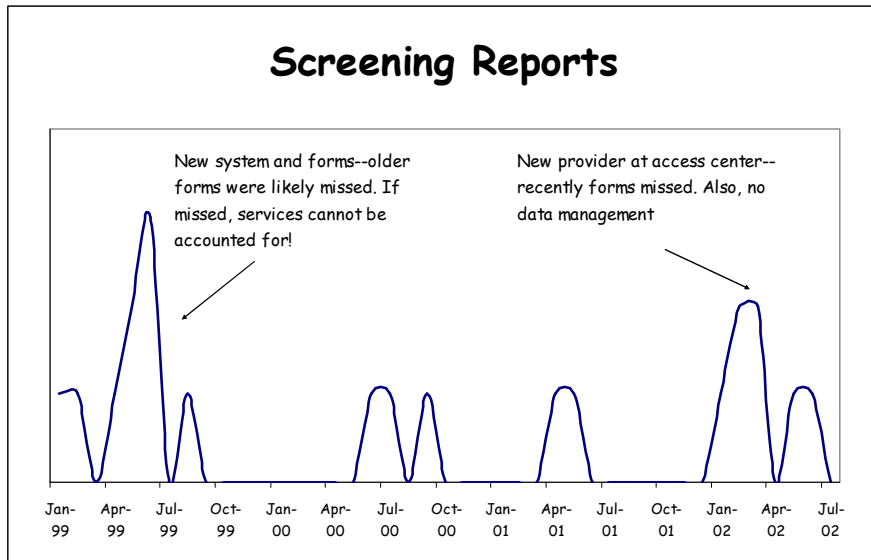


Figure 2. Missing screening frequency by month and year²

This is critical a problem because all service records, including exit and follow-up records, are tied to the screening. If a youth screening is not in the system, all the associated records/service events cannot be entered or analyzed. Initial reports of screening have shown steady declines in new screenings over time—currently the lowest levels ever (Figure 3). This may not be an accurate representation of reality as screenings since the beginning of this year have increased in likelihood of going unprocessed (Figure 2).

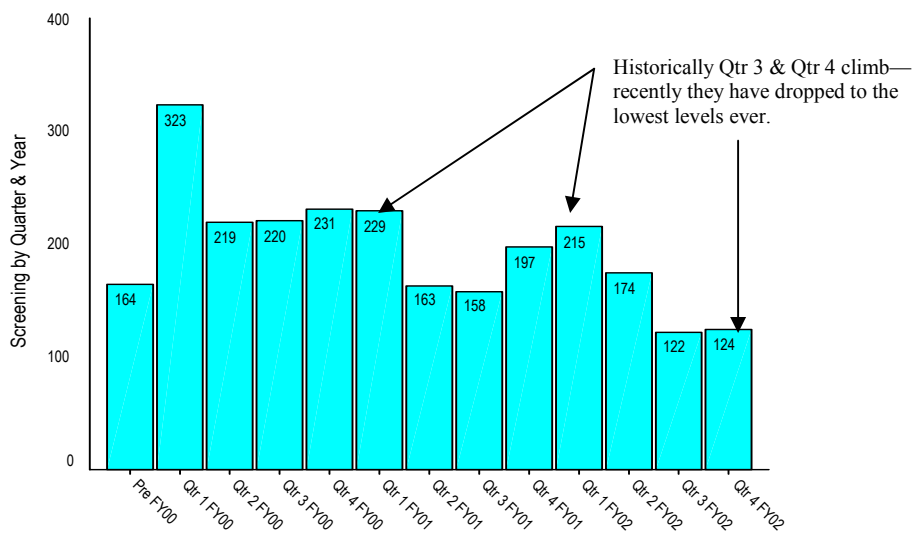


Figure 3. Screening frequency in HYDB by quarter and FY

² Graphs are best viewed not as specific counts of missing data, but when they occur and their magnitude.

Youth Assessments. A total of 1,682 youth assessments were identified in the HYDB. A sample of 161 was randomly selected (+/-7.4%). A total of four (2.5%) youth assessments could not be found in the HYDB. This is approximately 42 youth assessments which may be missing. In addition, services reporting forms identified 7.5% of cases where there were no youth initial assessments filed. The 7.5% is not independent of the 2.5% reported earlier, and level of overlap unknown.

Due to missing screening and youth assessment forms, another database (Homeless Tracker) was constructed to track the variety of missing forms and data. A significant proportion of the data entry time is devoted to tracking down missing forms or questionable responses. As of 8/22/02, there were 51 outstanding requests for missing youth forms; 71% (36) were for missing screening and/or youth assessment forms, meaning that service forms had arrived to OSCP, but were not entered due to the missing initial information.

Youth Self-report. At the time of the sample, a total of 1,288 youth self reports had been captured. A sample of 289 was drawn at random from providers (+/-5%). A total of 17 of the 289 were missing from the HYDB (6%), thus up to 77 youth self reports forms may currently be missing (Figure 4). These missing reports showed specific patterns detailed below.

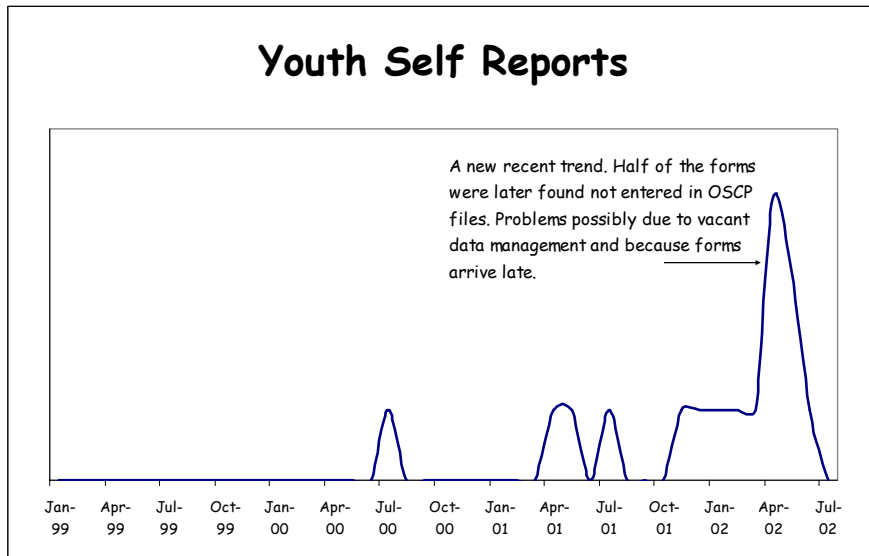


Figure 4. Missing youth self report frequency by month and year

The forms that were identified had the following characteristics:

- 53% (9) were from OI, thus older youth (none of these were for follow-up)³
- 53% (9) were later found in the OSCP paper file and had failed to be processed into the HYDB

³ Note I failed to collect NAFY form status ID, thus was unable to identify intake, exit or follow-up.

- Missing forms appear to follow a pattern based on time—the peak in April and May 2002 suggests a lack of data management at the time and likely late reporting by providers

In addition, youth services reporting data identified 17% of cases where there were no youth self reports filed (remember there should be at least an intake form, and likely an exit and possibly a follow-up). The 17% is not independent of the 6% reported earlier, thus level of overlap unknown.

Service Coordinator reports. At the time of the sample, a total of 1641 service coordinator reports had been captured. A sample of 257 was drawn at random from providers (+/- 5.6%). A total of 15 of the 257 were missing from the HYDB (6%), thus up to 98 service coordinator forms may be missing. Overtime, there have been consistent examples of missing forms, however this has spiked in the 4th quarter of FY02 (Figure 5).

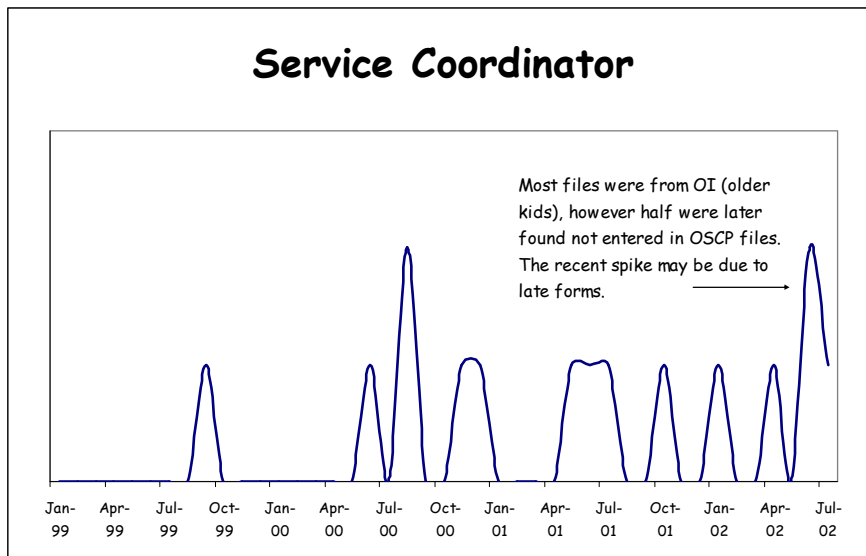


Figure 5. Missing service coordinator report frequency by month and year

The service coordinator forms that were identified had the following characteristics:

- 73% (11) were from OI, thus representing missing data from older youths
- 60% (9) were for youth exits, 27% (4) for follow-up
- 47% (7) of the forms were later found in the OSCP paper files, but not processed
- Nearly all of the OSCP paper files found were from OI (7)
- Missing forms do not follow a clear pattern of time, however a recent spike in missing forms was identified (likely late reporting by providers)

In addition, services reporting data identified 12% of cases where there were no service coordinator reports filed (remember there should be at least an intake form, and likely an exit and possibly a follow-up). The 12% is not independent 6% reported earlier, thus level of overlap unknown.

An additional “snapshot” of data taken in August 2002 (for the month of July) identified 17 cases of late arriving forms from providers. These forms should have been processed for the previous FY. Of the 73 service coordinator and youth self reports identified for June 2002, 16% (12) of forms processed for June 2002 arrived well after the reporting date. This delay reduced the volume of reports reported in FY02.

INTERNAL DATA PROCESSING

Evaluation model appears biased. The current model follows the assumed pattern of forms for intake, exit, and a 6-month follow-up based on a youth entering, exiting, and latter contact. This group of clients is highly transient, and some are very difficult if not impossible to locate after services. Currently, service providers track down the youth by telephone, mail or in person if they can be located. If the service providers are actually able to contact a youth, it would suggest increased stability (enough to locate). Those follow-up personnel who were able to contact youth would have a greater likelihood to be a success story than not. This is a significant bias to any outcomes.

Data model problematic. As noted above, the model assumes youth intake, exit, and a 6-month follow-up. While this is the theoretical model of events, the forms and data available do not often follow that pattern. It is not uncommon for a youth to have multiples of the same forms at different time periods (e.g., 30% of youth had more than one youth assessment intake form, with at least one youth having four in the system). It was also common for youth who received some service to have a combination of missing intakes, exits, or particularly follow-ups forms. This can occur for many reasons from youth refusing to fill out the form, to forms not getting to the OSCP, to forms at OSCP never getting entered (see above).

The problem occurs when analyses were attempted. Previous data retrieval methods appeared to query only the most recent data from a youth (i.e., the most recent intake, exit, and follow-up)—this occurred even if the youth had multiple service episodes over several years. The audit found several cases where service treatment sequences were out of order during data retrieval and subsequent analyses (Table 1). For example, the system matched John Doe’s (#972) youth self report intake from May of 2001 with his exit in June 2000 and a follow-up from March 2001. The analysis compared his exit performance a year *prior* to his intake. In addition, because the system took only the most recent data, John Doe’s past service episodes were ignored. This means that John Doe’s 1999 and 2000 service episodes were probably not analyzed because he had service in 2001. Several other cases were also noted below.

Table 1. Examples of mismatched intake, exit and follow-up data

Name	Intake Date	Exit Date	Follow-up Date
John Doe (#972)	5/01	6/00	3/01
Richie Rich (#273)	3/02	5/02	10/00
Sally Sue (#166)	8/00	3/01	2/02 (overrode 9/01)
Jane Doe (#470)	6/01	3/01	3/01

The query and subsequent analyses mismatched intake, exit and follow-up data in situations where multiple service episodes had occurred but not all service forms were available (i.e., forms were missing). This appears to have been a common occurrence, since at least July 2001. These problems appeared for both the service coordinator forms and the youth self assessment form, and can occur anytime multiple forms were linked. These have led to erroneous results which *may* impact prior service continuum conclusions. The magnitude of these effects cannot be assessed at this point in time. In addition, the store of additional data from prior services was likely never analyzed.

DATA QUALITY

Data missing in HYDB system. Just because forms arrive and are entered doesn't mean that the data is useful. Missing data in database is fairly common—concerns are raised when the missing data is not random, but systematic. Prior outcome assessments were based on the difference between initial and exit, and exit and follow-up scores on a large variety of measures (e.g., alcohol and drug use frequency).

An initial analysis of the youth self reports data found that of the 67 variables assessed an average of 19% were missing at exit, which was twice the rate of the intake or follow-up forms (approx 8.5%).⁴ Further investigation found that OI forms were significantly more likely to be missing data than were NAFY.⁵ Because this does not appear random, it directly effects the assessment of any outcomes. One bias introduced impacts OI which serves older youth who are anecdotally associated with greater difficulty to serve. The second impact is the importance of exit data—assessing changes in intake to exit and exit to follow-up become less reliable when unusually large segments of exit data are missing.

Further examination occurred with the HYDB youth assessment data. The HYDB identified 1,882 assessments associated with youth. However, 393 (21%) of these cases were for “place holder assessments.” Place holder assessments were entered into the system when youth screening were on file and some service delivery had occurred, but where no youth assessment could be identified. Assessments are used for intake baseline data and also to link clients who have multiple service episodes. These assessments acted as place holders for missing forms, so that other data could be entered.

Too few available outcomes. According to the HYDB 650 youth had returned at least one youth self report (intake, exit, or follow-up) since its inception. Of those 149 had at least one exit form. Of those only 82 had an intake, an exit and a follow-up. Assuming this series was in perfect order (c.f., above), it is not possible to generalize the available outcome data when less than 13% of the data is available. Taken in concert and especially with the extremely small number of outcomes available, finds that true outcomes with statistical rigor cannot occur with the available data under this model.

⁴ 8.5% may be perfectly reasonable, as some questions may not apply to all youth (e.g., sexual relations), and some questions may have been remove over time.

⁵ ANOVA. $F(1, 1079) = 55.401, p < .001$. It is not possible to assess why this was the case at the time of the report.

CONCLUSIONS

Results identified several problems with missing data forms, the database queries, data quality, and fundamental assumption of the Homeless Youth Evaluation model, leaving us to conclude that true “outcomes” cannot be evaluated under the current model. However, a significant proportion of data has been collected and appears usable. With thought, it could produce useful program information. In addition, several corrective steps have already taken place to help insure that results of data that are reported by PMG will be done so with the highest level of validity and confidence.

Screening and initial assessment data appear to be reasonably stable, although it appears to be missing more recent screening (3rd and 4th Quarter 2002). These should be remedied as time goes by, however late forms will consistently reduce the accuracy of recent data and recommend a long-term correction to the data processing model. In addition, we recommend delaying canned reporting to the outside stakeholders to assure that greater amounts of latent data are processed and accurately reported.

Service reporting data, while not examined in-depth also appears useful. With over 11,000 records detailing specific services delivered and the number of hours, this information may be quite useful if analyzed creatively.

Several corrective actions have already begun to increase the quality of HYDB data. Sampling, which revealed many cases of missing forms, have since found and subsequently been added to the HYDB, thus reducing the missing amount of data. Several lists have also been developed identifying missing forms, and if available will be tracked down and entered into the system. These steps will allow forms that are currently waiting to be entered into the system to be processed. In addition, the youth self reports have been redesigned, adding a section for provider review and signature. This increases accountability and the likelihood that forms will arrive completed, thus reducing missing data elements in the HYDB. We recommend that future data be electronically uploaded from providers to the HYDB to avoid paperwork inefficiencies and the possibility for lost forms and data.

While service coordinator and youth self report data is available, the previous evaluation model was unable to utilize them to assess outcomes (this was due to the reliance of the “follow-up form”). We recommend rethinking “outcomes” to allow for a broader, but still useful determination of program success. One possibility may be to ignore the intake, exit, and follow-up semantics of the reporting, and link the reports simply by when they were completed in time to assess program outcomes. For example, it may be possible to take the youth’s intake, exit and a subsequent intake form to assess outcomes (in all cases it is the same form). Or, instead of relying upon exit and outcome data, one may be able to assess the amount of time before a youth reenters the system for additional services and what those may be (i.e., program return). These methods introduce other limitations but mitigate the problems of missing or incomplete data and may be a better reflection of those youth after services. These outcomes taken with other models could then be used to create a mosaic of continuum performance. Regardless, the evaluation plan must undergo serious thought to prevent future reporting errors in this complex evaluation.