CalWORKs Mental Health Outcome Monitoring: Implementation and Demonstration Study

July 2016

Daniel Chandler, Ph.D.

SUBMITTED TO THE LOS ANGELES DEPARTMENT OF MENTAL HEALTH (DMH)
BY THE CALIFORNIA INSTITUTE FOR BEHAVIORAL HEALTH SOLUTIONS
**EXECUTIVE SUMMARY**

- Outcome: Change in work
- Outcome: Change in psychiatric disability
- Four policy issues
- Methodological issues and recommendations

**CALWORKS MENTAL HEALTH OUTCOMES MONITORING: IMPLEMENTATION AND DEMONSTRATION STUDY**

- Background
- Overview of the study

**A. EMPLOYMENT OUTCOMES OF CALWORKS MENTAL HEALTH SERVICES**

- Use of Individual Placement and Support services
- Employment Pre to Post
- Employment change over four quarters
- The relationship of employment to the reason clients leave services

**B. PSYCHIATRIC OUTCOMES OF CALWORKS MENTAL HEALTH SERVICES**

- Change in measures of psychiatric disability in the group of participants as a whole
- Measuring reliable and clinically significant change

**C. POLICY ISSUES RAISED BY THE OUTCOME DATA**

- POLICY ISSUE 1: Do all clients appear to need the level of professional mental health care offered by CalWORKs mental health services?
- POLICY ISSUE 2: Can low treatment completion rates be improved?
- POLICY ISSUE 3: Should we be concerned about the variability among providers regarding effectiveness in helping clients?
- POLICY ISSUE 4: Do the prevalence and effects on treatment of co-occurring domestic abuse suggest a need for a new approach?

**D. METHODOLOGICAL ISSUES, FINDINGS, AND RECOMMENDATIONS**

- Structure of data collection when the system is fully implemented
- Performance of the instruments currently used
- Improving data collection and reducing missing data
- Burden of the study on providers
- Reporting and making data useful for DMH and providers

**APPENDICES**

- APPENDIX 1: Outcomes Monitoring Implementation Study 2014-2016: Data Security
- APPENDIX 2: Training and Implementation
- APPENDIX 3: Psychiatric Disability Rating Scales Used in the Monitoring
- APPENDIX 4: Mock-Up of Proposed Consumer Form Revision
EXECUTIVE SUMMARY

This study tests the implementation of an outcomes monitoring system in the Los Angeles CalWORKs mental health program. The study began October 1, 2014 and ended in February 2016. It included 54 providers and 659 client participants—all those admitted into a program between October 1 and December 30, 2014. To obtain information, a set of questions about each participant was filled out by clinical staff at baseline, in each of four quarters, and at discharge. Participants filled out a satisfaction form quarterly.

The report contains information about a) employment outcomes, b) change in psychiatric symptoms and functioning, c) four policy questions raised by the findings, and d) methodology and how the implementation went. We present major findings and recommendations in this executive summary.

Outcome: Change in work

- Supported employment is not being utilized sufficiently. No more than 67 out of 659 persons were in supported employment at any time or at discharge. However, those who did participate in IPS were much more likely to be employed at the end of the study. At baseline 15% of both those who received IPS and those who did not were working at least part-time. At follow-up, 26.5% of those not in IPS were working and 51% of those in IPS worked.

- Employment overall increased overall from 16% at baseline to 29% at discharge (for persons for whom we had data at both times). Having two or more children under five reduced the predicted likelihood of working.

- Persons who left services for positive reasons were far more likely to be working at discharge than those who left for negative reasons (65% of those leaving for positive reasons worked vs. 22% of those leaving for negative reasons).

- To improve the monitoring system, DMH should consider asking DPSS for follow-up employment data for all admits in each calendar year. Analysis could be part of the CIBHS contract. This would allow DMH to track employment after clients leave mental health services—which is impossible in the current monitoring system.

Outcome: Change in psychiatric disability

Change on scales measuring psychiatric disability was characterized by both statistically significant change of the group over time and by clinically significant change for individuals.

Group change

- Global Assessment of Functioning (GAF) scores of client functioning at discharge increased steadily over time with an overall change of 4.5 points (on a hundred point scale) from baseline to discharge. Those discharged after a year in the program showed a 7.9 average change. These changes were statistically significant, and the “effect size” of this change is .44, close to a medium sized effect (.50)
• The Clinical Global Impressions scale (CGI) recorded a statistically significant change from baseline to discharge in symptoms. The effect size was .59, greater than a medium effect but less than a “large” effect (.80).

• The consumer questionnaire used the K6 and Sheehan Disability scales to measure psychiatric status. However, high attrition meant that only 113 persons were present at baseline and the first two quarters. On all three scales there was a statistically significant improvement over three quarters.

• The consumer scale showed more participant satisfaction with their services among those who were there three quarters. By the end of the second quarter, 93% were at least “very satisfied” and 99% said services helped at least “some” (71% said they “helped a lot”).

**Reliable and clinically significant individual change**

• Because the reliability of the GAF and CGI, though high, is not perfect we tested adjusting outcomes to take account of scale unreliability. After doing so, the GAF showed reliable positive change for 121 persons and reliable negative change for 14 persons. The CGI showed reliable positive change for 240 and reliable negative change for 56.

• Looking at individual change on the GAF we found negative change for 12%, no change for 38%, positive change of 1-9 points for 28% and clinically significant change of 10 points or more for 23%. Only 7% moved from below the “normal” range of 70 -100 into the normal range.

• For the CGI we found that 15% had clinically significant change as judged by having moved into parts of the scale describing persons with no symptoms or unclear symptoms. When CGI results are categorized in the same way as were those of the GAF, 12% had negative change from baseline to discharge; no change was found for 38%; a one step increase in mental health was registered for 30%; and 20% had at least a two step increase. So both instruments were similar in their description of the amount and direction of individual change.

• The Clinical Global Improvement scale rates change at the time of discharge. Based on its categories 6% made “ideal improvement;” another 8% made “very considerable improvement;” and 11% made “considerable improvement.”

**Four policy issues**

*Issue 1:* Data from the monitoring study and previous outcome studies suggest up to 23% of participants have few or no symptoms at admission and so may not need services specifically from licensed mental health professionals.
Recommendations:

- It may be that this situation reflects incorrect referrals from DPSS. We would have to do another study of a different kind in order to determine whether there are large numbers of persons with need in the DPSS population who are not being referred by DPSS. If the level of need in persons referred increased, presumably the issue in the programs would disappear. DMH might suggest to DPSS a special study next year of screening and referral that focuses on improving both sensitivity (getting the right people) and specificity (not getting the wrong people).

- If some programs have a particularly high rate of persons with less severe mental health problems it may be because they are not getting enough referrals. A study to look at number of referrals, waiting lists, and percent of assessments admitted would permit getting a handle on this possible cause.

- DMH might also recognize the serious family and employment issues some clients may have even when their mental health issues are mild. In that case, these clients might be offered only case management and peer-support mental health services. Some programs already operate this way.

- Anomalies between outcome monitoring instrument scores and level of care placement might be investigated during site visits. CIBHS can supply DMH a list of these cases if desired.

Issue 2: This study confirmed a long-standing finding that only 15-20% of participants complete treatment in the sense that they achieve the goals they initially set with their therapist.

Recommendations:

- The May 2015 Outcomes Study found that those with poor attendance are much more likely to leave early than others. Given the confirmed low completion rate in this study, we suggest testing a requirement for a home visit if attendance is poor early on.

- Because the 10% of clients with PTSD diagnoses are especially likely not to complete treatment, it may be possible as a system or at least in individual providers to develop ways of engaging them.\(^1\)

- And we recommend a system-wide quality improvement effort in which each provider takes responsibility for identifying causes for poor attendance and early dropout, and then creating, testing, and evaluating solutions.\(^2\) For example,

---

\(^1\) Although there are 67 persons with PTSD, they are spread out with only 2-4 in each clinic.

\(^2\) Many models and resources are available for helping staff carry out a quality improvement project, such as the FADE and PDSA models described at [http://patientsafetyed.duhs.duke.edu/module_a/methods/pdsa.html](http://patientsafetyed.duhs.duke.edu/module_a/methods/pdsa.html). Wells has documented the success of a multi-agency QI project to improve care of depressive disorders. Wells, K. B., Sherbourne, C., Schoenbaum, M., et al. Impact of disseminating quality improvement programs for depression in managed primary care: A
a non-profit organization that helps behavioral health agencies improve engagement is NIATx. One tool they use is the “Walk Through,” in which an agency administrator notes issues as she walks through the intake process of the agency as if she were a consumer seeking services. But numerous other tools are available. There are NIATx experts available at UCLA and in Ventura Behavioral Health as well as at the national office in Madison.

**Issue 3:** Providers and Service Planning Areas show quite considerable variation. Is this a matter of concern?

The sample size in this study for most programs is too small to reliably characterize the programs in relationship to each other. However, the sample size in each of the Service Planning Areas was much larger, and findings in this study indicate large disparities in outcomes between SPAs. Whether these are of concern will depend on being able to adjust differences in population and resources so that comparisons are fair.

**Recommendations:**

Whether the high variability we found in this study is of concern cannot be resolved until the full implementation of the outcome monitoring system provides much larger sample sizes for each program than have been available to date.

However, no comparison should be instituted unless, based on advance consultation between providers and DMH, providers consider it to be fair and helpful.

**Issue 4:** Do participants experiencing domestic abuse do less well than others? If so, is the current “silo” approach to those with co-occurring mental health and domestic abuse problems adequate?

- There are substantial numbers of clients experiencing at least moderate domestic abuse (21%) and another 21% experience minor abuse.
- These rates decline significantly during treatment.
- Persons who are experiencing current domestic abuse at discharge show less positive change in their psychiatric disability than do those with no abuse.

**Recommendations:**

When CalWORKs supported services began in 1999, much joint planning and communication took place among DPSS, and mental health, substance abuse and domestic abuse agencies at the county level. That interaction has diminished very substantially, to the extent that silo services predominate.

---


• As we have proposed in other reports to the DMH, the 40% percent of persons with domestic abuse situations would suggest that each clinic have at least one clinician who is a certified domestic violence advocate, having taken the 40 hour training.

• An additional remedy would be to designate one clinic in each SPA that would have a track that integrates mental health and domestic abuse services. Persons with active safety issues would be referred to those clinics. This would involve an agreement with the county agency administering the CalWORKs domestic abuse services.

• Although clinicians appear to be good at surfacing abuse issues, a formal screening instrument might increase the detection rate. DMH is considering the HITS instrument, but there are other possibilities. One or more instrument could be tried out in the next iteration of this study.

• Because the Outcome Study of 2015 found patterns of increased abuse during treatment through the use of phone interviews with persons who had left services, DMH may wish to ask CIBHS to conduct post-discharge phone interviews with a representative sample of participants who are experiencing domestic abuse at baseline.

Methodological issues and recommendations

This implementation and demonstration study was designed to test a number of options regarding the administration of the outcome monitoring forms as well as to test the forms themselves. Below are conclusions about different aspects of the methodology.

• Sampling is not a viable option because it will produce too few cases to profile individual providers. All mental health participants should be included in the monitoring system.

• Keeping track of quarterly updates was apparently very difficult for providers. We recommend that only a baseline and discharge staff form be completed for participants.

• Consumer forms had the same problem with quarterly updates but also experienced high attrition at discharge. Because of this attrition, use of a consumer form as an outcome measure (pre and post services) is not feasible unless DMH decides to use phone follow-up interviews after clients leave services. The form can still serve its function as a measure of satisfaction and effectiveness of service design. We recommend two administrations: one a month after admission and one six months later.

---

iv Research studies with the CalWORKs population have show that approximately 80% have past abuse, so almost twice the 42% reported at baseline here.
• The three staff instruments that measure psychiatric disability all seemed useful. They should be continued, although the GAF may need to be replaced by a different instrument to measure functioning because the DMH no longer uses the GAF. Reliability of the instruments in this outcomes monitoring context might be the focus of a special study next year.

• The burden on providers appears to be reasonable and should continue to be so under the proposed full implementation.
CALWORKS MENTAL HEALTH OUTCOMES MONITORING: IMPLEMENTATION AND DEMONSTRATION STUDY

Background.

In 2003 DMH contracted with CIBHS to conduct an evaluation of the outcomes of CalWORKs mental health programs using a random sample of providers and participants. A total of four outcome studies were conducted in the next four years, some also involving substance abuse and domestic violence providers and participants. In addition, a peer reviewed journal article was published using outcome data from the 2007 study.

These studies also served to meet a DPSS requirement that DMH conduct consumer satisfaction studies. Even in years in which a full-fledged outcome study was not conducted, a random survey of consumer satisfaction and perception of their treatment was conducted. Eleven annual surveys have now been completed and show stability and change over time on a set of issues focusing on consumer satisfaction.

Both the more comprehensive outcome studies and the consumer surveys were limited in three important ways: a) they were cross-sectional rather than longitudinal; b) the sample (approximately 400 each year) was appropriate for profiling the system as a whole but not for allowing comparisons between providers; and c) the outcomes for the CalWORKs mental health system were all based on special studies rather than being part of an ongoing system in which outcome data are routinely collected and can be compared over time.

In 2010 DMH asked CIBHS to design an outcome monitoring system that would overcome these problems and be incorporated into the DMH management information system. Thus the goal became developing staff and consumer rating systems that could be administered at admission, discharge and quarterly that would reliably allow DMH to monitor how much clients improve, their satisfaction, and the reasons for leaving the program. The set of measures designed by CIMH was tested in 2011. As a consequence of the test, a number of the scales were changed and forms revised. Until this study began in 2014, however, these revisions had not themselves been tested in a longitudinal framework.

Due to DMH’s current implementation of a new information system, it has been a challenge to incorporate the CalWORKs outcome monitoring system into the information system as anticipated; however, there is ongoing communication between departments within DMH to resolve this issue.

In the interim, DMH has instituted a “Provider Report Card” that is completed monthly by each provider and which covers a variety of aggregate outcome and process measures.

5 The California Institute for Behavioral Health Solutions was known until 2014 as the California Institute for Mental Health.

6 These reports, as well as outcomes reports on the Homeless CalWORKs Families Project, are available at: http://www.cibhs.org/calworks-publications

Some of these measures will be incorporated into the outcome monitoring system either as outcomes or as important variables for cross tabulation.  

Consequently, starting October 1, 2014, DMH and CIBHS began testing the outcome monitoring system in its entirety (not just the instruments but the timing, reporting mandates, and analysis of results). The study ended February 29, 2016.

The current goal is to use study results to achieve a manageable and effective monitoring system that can be easily transferred to the new DMH information system when that system can assume the responsibility. In the interim, DMH implemented the ongoing system with the help of CIBHS and the use of a HIPAA compliant database purchased from an enterprise provider.

**Overview of the study**

All CalWORKs mental health participants who enrolled in services between September 30, 2014 and January 1, 2015 were enrolled in the study. That is, the sample consisted of a full quarter of new CalWORKs mental health clients. The actual number enrolled was 659. Clinical staff submitted information at baseline and, as people were discharged, the same information, plus reason for discharge, was entered into the study data base. For those remaining in service, additional information on the same clinical and employment-related characteristics of clients was submitted by staff. In addition to staff-completed forms, clients were asked to provide quarterly information on satisfaction with services and a self-report clinical measure, the K6. This information was also collected quarterly when possible; obtaining it at discharge was difficult as clients often dropped out rather than having a planned termination.

The monitoring implementation study uses a number of different instruments that measure psychiatric condition as well as measures of employment, education, substance abuse, and domestic violence. The symptom scales are listed below and reprinted in the appendices to this report.

- Clinical Global Impressions scale: a symptom rating scale for staff
- Clinical Improvement scale: a tool for clinicians to rate improvement at discharge
- Global Assessment of Functioning scale: includes functioning in work and family as well as symptoms as rated by staff
- Level of Care determination by clinical staff: assigns three levels of treatment intensity
- Client self-rating of symptoms (on the K6) and functioning (on the Sheehan disability scale).

---

8 Outcomes include the substance abuse and domestic abuse scales as well as hours worked per week. Cross-tab variables include basic participant characteristics such as age category and whether the client participated in supported employment.
There are three sections to this report:

- The first part examines employment outcomes and clinically significant change in psychiatric disability—the two primary goals of CalWORKs mental health treatment.
- The second part attempts to provide confirmation and more detail regarding several policy issues first identified in another CIBHS study published a year ago in May 2015.
- The third part examines what we have learned about implementing the outcome monitoring system and suggests some changes in both measures and procedures.

Appendix 1 describes measures taken to protect the safety of the data as prescribed by HIPAA. Appendix 2 describes training and other elements of study implementation, including missing data. Appendix 3 reproduces the staff and client instruments used in the study. Appendix 4 contains proposed modifications to the consumer questionnaire.

**A. EMPLOYMENT OUTCOMES OF CALWORKS MENTAL HEALTH SERVICES**

The outcome system being tested is prospective—we can follow participants from beginning to end of their mental health services. This has permitted an examination of participation in supported employment as well as changes in employment over time. It permits snapshots of individual change on a quarter by quarter basis. The basic limitation of the monitoring system is that information is not gathered in the months following participants’ termination from the program.

**Use of Individual Placement and Support services**

Since mid 2012 all contracted CalWORKs programs were required to have a supported employment program; however, some providers continued to work under their contract statement of work which did not specify that supported employment meant the Individual Placement and Support (IPS) Model. CalWORKs Administration will be issuing a Request for Proposal in 2016 which includes the provision of Supported Employment-IPS. Supported Employment-IPS is an evidence-based mental health program with very strong research support. As part of the outcome monitoring study we asked whether participants were enrolled in IPS or there was a plan to enroll them.

At the time of the baseline staff report, only 6% were enrolled in IPS. This increased to 46% for the small number who received at least four quarters of service. The column labeled “Final” is the final staff report: for 488 persons it is a discharge summary but for about 20 persons an end of study summary for persons still in treatment more than 15 months from the start of the study. However, the total column is most relevant as it indicates that over

---

9 An overview of IPS is available at [http://www.dartmouthips.org/about-ips/](http://www.dartmouthips.org/about-ips/). Although there are many randomized controlled trials of IPS, very few have focused on a TANF population of persons with psychiatric problems. Information about the implementation of the program in the Los Angeles CalWORKs Mental Health system is in an earlier publication by CIBHS: [http://www.cibhs.org/post/los-angeles-calworks-mental-health-services](http://www.cibhs.org/post/los-angeles-calworks-mental-health-services)
time only 16% of the 1,677 person/quarter slots included IPS. So we have, in summary, a small percentage of all clients who participate in IPS. However, the percentage increases the longer clients stay in treatment. As in many other aspects of CalWORKs mental health services, the high attrition rate affects overall IPS participation rates since many client leave before entering IPS. (Note. This would be a good time for readers to look at the Appendix 2 regarding attrition and missing data, especially Table 29.)

**Table 1: IPS Participation over time**

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Qtr1</th>
<th>Qtr2</th>
<th>Qtr3</th>
<th>Qtr4</th>
<th>Final</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 - Yes, in IPS</strong></td>
<td>6.1</td>
<td>16.3</td>
<td>29.0</td>
<td>30.6</td>
<td>45.7</td>
<td>18.4</td>
<td>16.0</td>
</tr>
<tr>
<td><strong>2 - No IPS</strong></td>
<td>72.0</td>
<td>66.7</td>
<td>62.5</td>
<td>64.7</td>
<td>52.2</td>
<td>75.5</td>
<td>69.6</td>
</tr>
<tr>
<td><strong>3 - IPS planned, not yet enrolled</strong></td>
<td>21.9</td>
<td>17.0</td>
<td>8.5</td>
<td>4.7</td>
<td>2.2</td>
<td>6.0</td>
<td>14.4</td>
</tr>
<tr>
<td><strong>Total [#]</strong></td>
<td>658</td>
<td>300</td>
<td>224</td>
<td>85</td>
<td>46</td>
<td>364</td>
<td>1677</td>
</tr>
</tbody>
</table>

Of interest is whether persons who did participate in IPS were more likely to be discharged for positive reasons than others. (As will be explained in detail later, 14 reasons for discharge can be collapsed into three categories: positive such as getting full time job, neutral such as moving, and negative such as going to jail. See Table 21 for the frequencies in all 14 categories.) Omitting those with neutral reasons for leaving (about 10% overall), 48% of those who participated in IPS were judged by staff to have left for positive reasons, compared to 27% of those who had not participated in IPS. This is not a controlled study, so this association doesn’t necessarily indicate causality. But it is nonetheless encouraging.

At discharge or end of study, those with IPS participation worked on average 14 hours a week compared to 7 hours for those with no IPS participation.\(^{10}\) In contrast, 18% of the non-IPS clients had at least some education units compared to 9% of the IPS clients. (DMH policy is that in order to participate in IPS clients must be interested in work not just education—which is available through GAIN programs.)

**Association of IPS with higher employment rates.** At baseline 15% of both those who received IPS and those who did not were working at least part-time. At follow-up, 26.5% of those not in IPS were working and 51% of those in IPS worked. These results are highly statistically significant. We investigated possible bias introduced by participants who did

---

\(^{10}\) Statistically significant in t-test, p=0.001
not have both a baseline and follow-up staff rating. On the basis of our analysis, we think it unlikely that having discharge data for the 171 with missing data would significantly change the results.

**Employment Pre to Post**

In theory, work can be recorded with high reliability. However, in this study, significant attrition (people who left the study but no discharge form was completed) means that there is unknown lack of reliability. Below we see the percentage working at baseline and final follow-up: there is an increase of 12% when using matched data. This change is highly statistically significant (p>.0001).

**Table 2: Basic changes in whether the client is working at baseline and discharge/end of study.**

<table>
<thead>
<tr>
<th>Measurement Time</th>
<th>Using all data whether matched or not (N=659 at baseline 513 at follow-up)</th>
<th>Matched cases baseline to final follow-up (N=488)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>15%</td>
<td>16%</td>
</tr>
<tr>
<td>Final</td>
<td>21%</td>
<td>29%</td>
</tr>
</tbody>
</table>

But we also want to know whether the change is of practical or clinical significance. One measure for simple percentages that goes beyond statistical significance (which is influenced by sample size) is the odds ratio. The odds of working (with matched data) are the odds of working at baseline (77 to 411) expressed as a ratio to the odds of working at follow-up (140 to 348), which in this case is an odds ratio of 3.3. Since an odds ratio showing no difference would be a 1, a ratio of 3.3 is quite high—equivalent to saying a person at discharge is 3.3 times more likely to be working than at admission.

Do demographic factors such as age, sex, race, diagnosis, and number of children influence whether clients take on work? As shown in Figure 1, only the number of children under five

---

11 Although we had 488 persons with matched baseline and discharge data, there were 171 persons for whom staff did not or could not complete the discharge form. We tested whether, at baseline, the group with matched forms was different from the group with no follow-up rating. Several variables were individually statistically predictive, but together they predicted a very small amount of the difference between the groups. The pseudo R-squared for the regression model was 0.04 and the area under the ROC curve was 0.63. An r-squared this low would be considered very small, and ROC values of .60 to .70 indicate “poor” discrimination.

12 Again, the issue of possible bias because of the 171 participants who did not have completed follow-up forms is relevant.

13 Note that the odds ratio cannot be calculated from Table 2 itself but only from the actual numbers as in the paragraph below the table.
had a statistically significant predictive effect, reducing the percentage working by almost half.\textsuperscript{14}

**Figure 1: Predicted effect of having children under five on follow-up percentage working**

![Effect of having kids under 5 on percent working at discharge](image)

**Employment change over four quarters**

A second question is whether the availability of data at quarterly intervals rather than just pre and post adds to our understanding of work patterns. Table 3 shows this data. The percentage working in the week before each form was completed is what was recorded. This maximized accuracy but does not tell us what was happening the other 11 weeks of the quarter. Ideally, DMH would capture continuous employment data the way DPSS does. Unfortunately there is no current way for DMH to access the DPSS data except in special studies. DMH should consider asking DPSS for employment data for all admits in each calendar year. Analysis could be part of the CIBHS contract. This would also allow DMH to track employment after clients leave mental health services—which is impossible in the current monitoring system.

---

\textsuperscript{14} Panel data model using XTLOGIT in Stata 14 with listed covariates. Only having two or more children was marginally statistically significant.
Table 3: Does quarterly data add to our understanding of changes in percentage who work?

<table>
<thead>
<tr>
<th>Timeframe</th>
<th>N</th>
<th>Proportion Working</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>658</td>
<td>0.16</td>
<td>0.36</td>
</tr>
<tr>
<td>Qtr1</td>
<td>300</td>
<td>0.23</td>
<td>0.42</td>
</tr>
<tr>
<td>Qtr2</td>
<td>224</td>
<td>0.26</td>
<td>0.44</td>
</tr>
<tr>
<td>Qtr3</td>
<td>85</td>
<td>0.27</td>
<td>0.45</td>
</tr>
<tr>
<td>Qtr4</td>
<td>46</td>
<td>0.33</td>
<td>0.47</td>
</tr>
<tr>
<td>Discharge/end of study</td>
<td>504</td>
<td>0.29</td>
<td>0.45</td>
</tr>
</tbody>
</table>

In Figure 2 we see these same data as output of a longitudinal regression model with age, race, sex, number of children, diagnosis, and IPS status controlled. The effect of time (baseline through the quarters to end of study) is highly statistically significant.\(^{15}\) Note that as the percentage working increases in each quarter the percentage of the whole study group receiving services decreases (see Table 3 for numbers). The lower numbers in later quarters means reliability is less than at other times.

**Figure 2: Increase in employment by quarter with 95% confidence intervals**

The relationship of employment to the reason clients leave services

As described above in the section on IPS, staff classified client reasons for leaving into fourteen categories (see Table 21). We further classified these reasons into positive, neutral and negative. For analysis here we leave out the neutral cases and look only at whether there is an association between positive and negative reasons for leaving and the percentage working at baseline, four quarters, and the final follow-up.

\(^{15}\) P<0.0001
At baseline, those who would eventually leave for positive reasons worked a small amount more than those who would leave for negative reasons (not a statistically significant difference). After that baseline difference, the positive-leavers outstripped the negative leavers at each measurement point. At discharge or end-of-study the positive-leavers had worked in the previous week 65% of the time vs. 22% for negative-leavers. Since 9% left because work or school was interfering with treatment and that was counted as positive there is a way in which the data reflect our definition of positive and negative. However, the magnitude of the difference shown in Figure 3 is much greater than attributable to the definitional issue.

**Figure 3: Relationship of working at baseline, quarterly, and at follow-up to positive or negative reasons for discharge**

<table>
<thead>
<tr>
<th>Timeframe</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qtr1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qtr2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qtr3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qtr4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Summary.** We have discussed supported employment and rates of employment:

- Supported employment is not being utilized sufficiently. No more than 67 out of 659 persons were in supported employment at any time or at discharge. However, those who did participate in IPS were much more likely to be employed at the end of the study. At baseline 15% of both those who received IPS and those who did not were working at least part-time. At follow-up, 26.5% of those not in IPS were working while 51% of those in IPS worked.
• The percentage in supported employment increased each quarter. But this was almost entirely due to the denominator getting smaller. In the fourth quarter 45% of participants were in IPS but that was only 21 persons—fewer than the 40 at baseline.

• At baseline 15% of both those who received IPS and those who did not were working some. At follow-up, 26.5% of those not in IPS were working and 51% of those in IPS worked. These results are highly statistically significant.

• Employment overall increased overall from 16% at baseline to 29% at discharge (for persons we had data on at both times). Having 2 or more children under 5 reduced the predicted likelihood of working.

• Persons who left services for positive reasons were far more likely to be working at discharge than those who left for negative reasons (65% vs. 22%).
B. PSYCHIATRIC OUTCOMES OF CALWORKS MENTAL HEALTH SERVICES

CalWORKs mental health participants qualify for treatment services because they have a psychiatric disability that negatively affects their capacity to be economically independent. In this outcome monitoring implementation study, several measures of psychiatric status were administered both at baseline, each quarter, and at discharge. These include the Global Assessment of Functioning (GAF), the Global Clinical Impressions Scale (CGI), the Clinical Improvement Scale, and the client self-report K6 and Sheehan Disability scales. The goal is to measure whether participants’ psychiatric disability improves with treatment and how much. This includes symptomatology (like anxiety or depression) as well as functioning in family and work situations. See Appendix 3 for copies of all of the psychiatric rating scales.

There are methodological issues we must confront in order to achieve this goal. In addition to difficulties caused by missing study forms as noted in Appendix 2, measurement of psychiatric symptoms must deal with the inevitable lack of reliability of psychiatric scales. For example, the Clinical Global Impressions scale has a Cronbach’s alpha of .88, which is quite good but definitely leaves room for chance. Additionally, since many participants are still in crisis mode when they enter services, we should expect to see reductions in symptom scores just as a matter of the passage of time. Finally, though we want to see statistically significant change, that alone is not enough; there has to be evidence of clinically significant change.

Change in measures of psychiatric disability in the group of participants as a whole

In this section we compare the group of participants at baseline (or in later quarters) with the same group at discharge. We are looking for evidence that there is a statistically significant difference between the participants when they enter the program compared to the same group as they leave.

---

16 If data are missing randomly there are procedures to take account of the missing cases. But if they are missing due to some biasing factor, no correction is possible. In this case we do not know if there was bias or not. Staff did not complete the forms. In theory completion is possible even if clients drop out. So this could have been random provider inefficiency. But staff may also have felt they had too little data—if a client dropped out fairly early or after a long period of non-attendance—so they just didn’t complete the form.

17 This is equivalent to a split halves test, but done with all possible splits.

18 This is called regression to the mean and is the primary reason why ideally we would have a randomized control group. If we do, both groups may show improvement (due to regression to the mean) but if a treatment group improves more than the other we can infer that it is due to the intervention rather than regression to the mean. With no control group inference is much less certain.

19 Statistical significance tells us, with a known degree of probability, whether differences between groups are compatible with chance variation. By convention, we say a difference between groups, or the same group at different times, is statistically significant if there is only a 5% or less chance that the measured difference could be due to chance. It is indicated as p<0.05.
GROUP CHANGE: GLOBAL ASSESSMENT OF FUNCTIONING

Pre-post data.

The mean of the GAF scale scores at baseline was 51.5 and it improved to 56.0 at discharge, a highly statistically significant change of 4.5 points overall. Statistically significant covariates were: level of care up to ten hours a week (in a regression model this decreased the mean change by 1.6 points). Three baseline diagnoses also changed the predicted GAF improvement: major depression brought the change down by 0.9, PTSD brought it down by 2.9; and adjustment disorder increased the score by 3.0 points.

Longitudinal analysis using quarterly data as well as pre-post.

As a first step, look at how discharge GAF scores changed depending on how long the clients were in service. Due to a programming glitch we don’t have exact dates for when discharge information was submitted (supposed to be within 2 weeks of discharge). But it is possible to reconstruct six different time periods in consecutive order during which discharges took place. As seen in Table 4, the longer participants stayed in treatment the higher was their discharge GAF score—up to a year of treatment.20

Table 4: Discharge GAF minus baseline GAF scores, by period of discharge

<table>
<thead>
<tr>
<th>Period Discharged</th>
<th>Number</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before Feb 2</td>
<td>55</td>
<td>1.85</td>
<td>4.28</td>
</tr>
<tr>
<td>Feb 3 - May 25</td>
<td>172</td>
<td>2.32</td>
<td>11.50</td>
</tr>
<tr>
<td>May 25 - July 1</td>
<td>32</td>
<td>5.16</td>
<td>7.44</td>
</tr>
<tr>
<td>July 2 - Aug 14</td>
<td>76</td>
<td>3.64</td>
<td>11.97</td>
</tr>
<tr>
<td>Aug 15 - Oct 1</td>
<td>35</td>
<td>7.89</td>
<td>11.90</td>
</tr>
<tr>
<td>After Oct 1</td>
<td>90</td>
<td>7.60</td>
<td>10.34</td>
</tr>
</tbody>
</table>

The second step is to use the same regression model variables used earlier but include quarterly data on GAF and other control variables. Our goal here is to see if the use of quarterly milestones shows other important variables that predict GAF scores.

First, the longitudinal analysis, as expected shows significant change from baseline for Quarter 2, Quarter 4 and at discharge or end of study as well as overall.21 So while the overall result is to show significant change due to the passage of time, change was not significant between baseline and quarter 1 and also not significant between baseline and quarter 3. That is, using quarterly data, progress is not consistent over time.

Second, with the use of quarterly data several variables are statistically significant predictors of GAF improvements that did not show up on just the pre-post analysis. These include (besides level of care and diagnosis) lower GAF at discharge if substance abuse was

\[20\text{ These are raw data, but adding period of discharge to the previous regression analysis showed the time period to be statistically significant.}\]

\[21\text{ The joint test for statistical significance with change by quarters is } p<0.0001\]
present at baseline (but not dependence), and lower GAF if domestic violence was present at baseline, and lower GAF if male as opposed to female.

**GROUP CHANGE: CLINICAL GLOBAL IMPRESSIONS SCALE**

**Pre-post data.**

The CGI scale goes from “Normal” to “Extremely ill, may need hospitalization” (see Appendix 3). In this study the full seven point range was represented at baseline. The baseline mean score was 3.96 and the score at discharge for the 483 persons with matched ratings was 3.24. Since lower scores are more normal, this is an overall improvement and one that is highly statistically significant.

**Longitudinal analysis using quarterly data as well as pre-post.**

Figure 4 shows the average scores for baseline and each of four succeeding quarters as well as the “final” score recorded at discharge. Age, sex, race, children under five, level of care, IPS status and substance abuse and domestic violence status were held constant.

**Figure 4: CGI scores over time: Lower is better**

The biggest drop was between the baseline and Quarter 1. This could be either because getting into treatment relieved the immediate crisis or it could be because those with more severe symptoms dropped out—or both. We can’t test these hypotheses exactly but we can get close. The mean score for Quarter 1 on the CGI is 3.32. The mean score on the discharge form for persons who were discharged between October 1 and February 2 was 3.80—quite

22 Lack of statistical significance for dependence is almost surely due to the small number of clients.

23 Paired t-test: p<0.0001. Note that the GAF is a 100 point scale and the CGI seven points, so they are not directly comparable. The effect size, which we will discuss below, standardizes the scales so they are capable of being compared.
a bit higher than that of Quarter 1. Thus it looks like CGI scores may decline so much in the first three months due in part to persons with greater severity dropping out.

THE K6 AND SHEEHAN DISABILITY SCALES: CLIENT SELF-REPORT

The scales we have looked at so far codify the impressions of staff clinicians. Many studies, however, have documented that client self-report can be significantly different from staff views.

As noted in Appendix 2 on attrition, it is very difficult to “capture” participants for a follow-up administration of the client self-report instruments given at baseline and quarterly. This is due to the high percentage who drop out or leave for other negative reasons. Even those leaving for neutral reasons, such as moving, may not visit the clinic a last time. Consequently follow-up data was collected for only 55 persons out of an original 454 who completed the form at baseline. Because of this pattern there is little we can learn from a discharge form.\(^{24}\) However, we can at least look at the first two quarters of repeat consumer forms. They don’t describe the clients final psychiatric status or their final impression of the clinic, but they can show the change that occurs for those who remain in treatment for three to six months.

The K6 Instrument. The client self-report instrument built-into the monitoring implementation and demonstration study is the K6.\(^ {25}\) It is a short scale focusing on symptoms that have been found to make up a common core among persons with many types of serious mental illness. It is usually supplemented by two questions on functioning from the Sheehan Disability Scale. The six items on the K6 are scored 0-4, with a total score of 24 indicating greatest severity.\(^ {26}\)

Below we show the baseline, quarter 1 and quarter 2 scores on the K6 for persons who were present all three administrations of the form, which is the only way we can judge change. Although there were 454 baseline forms, there were only 232 in quarter 1 and 197 in quarter 2. However, these were not the same people each time. Only 113 persons were present in all three quarters.

For these 113 persons, the mean score at baseline was 11.6, at quarter 1 it was 10.5, and at quarter 2 it was 10.4. Although these changes are moving towards less psychiatric distress and are statistically significant, the amount of change is relatively small.\(^ {27}\)

\(^ {24}\) This is both because of not having matched baseline and discharge cases and because with such high attrition the discharge cases cannot be considered to be representative at all.


\(^ {26}\) For most questions the percentage missing was around 2%; for some reason 5% did not answer the “restless” question.

\(^ {27}\) In the next section we introduce the concept of “effect size.” The effect size of the change from baseline to the second quarter on the K6 is .20, which is a “small” effect size.
Sheehan disability scales. The Sheehan work disruption scale ask clients to rate the disruption of work/school caused by the psychiatric symptoms reported on the K6. The Sheehan family disruption scale does the same for family problems. The scale uses the concept of “disrupted” work or family functioning. For the work scale, participants are asked to rate on a ten point scale how much during the past week “your mental health problems have disrupted your work or school or CalWORKs work-activities.” (See copies of the questions in Appendix 3.)

The amount of disruption of work/school declined from 5.32 at baseline to 4.89 in quarter 1 to 4.12 in quarter 2. For family disruption the change in scores is from 5.41 at baseline to 5.10 to 4.50. As before this is for the 113 persons present in baseline and both quarters 1 and 2. For both scales the reduction from baseline to quarter 1 was not statistically significant, but the reduction from baseline to quarter 2 was.

Two other outcome measures are based on direct questions regarding the help clients perceive they have received and their satisfaction with that help. Clients increasingly reported being helped a lot: from 50% at baseline to 67% in quarter 1 and 71% in quarter 2.

Table 5: How much services are perceived by clients to help, over time (Percent is shown. N is 113)

<table>
<thead>
<tr>
<th>Services Helped?</th>
<th>Baseline</th>
<th>Qtr1</th>
<th>Qtr2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - helped a lot</td>
<td>49.54</td>
<td>66.96</td>
<td>70.80</td>
<td>62.57</td>
</tr>
<tr>
<td>2 - helped some</td>
<td>37.61</td>
<td>25.89</td>
<td>28.32</td>
<td>30.54</td>
</tr>
<tr>
<td>3 - helped a little</td>
<td>11.01</td>
<td>6.25</td>
<td>0.88</td>
<td>5.99</td>
</tr>
<tr>
<td>4 - did not help</td>
<td>0.92</td>
<td>0.89</td>
<td>0.00</td>
<td>0.60</td>
</tr>
<tr>
<td>5 - made things worse</td>
<td>0.92</td>
<td>0.00</td>
<td>0.00</td>
<td>0.30</td>
</tr>
<tr>
<td>Total</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Satisfaction also increases over time. “Extremely satisfied” goes from 35% to 42% to 54% in the course of six months. The percentage of either “very satisfied” or “extremely satisfied” increases from 80% to 93%. Very small percentages (less than 2%) express dissatisfaction.28

---

28 The changes for “helped” are statistically significant at p<0.01 and satisfaction at p<0.10.
Table 6: How satisfied with services clients say they are, over time (Percent is shown. N is 113)

<table>
<thead>
<tr>
<th>How satisfied?</th>
<th>Baseline</th>
<th>Qtr1</th>
<th>Qtr2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extremely satisfied</td>
<td>35.14</td>
<td>41.59</td>
<td>53.98</td>
<td>43.62</td>
</tr>
<tr>
<td>Very satisfied</td>
<td>45.05</td>
<td>42.48</td>
<td>38.94</td>
<td>42.14</td>
</tr>
<tr>
<td>Moderately satisfied</td>
<td>14.41</td>
<td>14.16</td>
<td>4.42</td>
<td>10.98</td>
</tr>
<tr>
<td>Somewhat satisfied</td>
<td>3.60</td>
<td>0.88</td>
<td>1.77</td>
<td>2.08</td>
</tr>
<tr>
<td>A little dissatisfied</td>
<td>0.90</td>
<td>0.88</td>
<td>0.88</td>
<td>0.89</td>
</tr>
<tr>
<td>Dissatisfied</td>
<td>0.90</td>
<td>0.00</td>
<td>0.00</td>
<td>0.30</td>
</tr>
<tr>
<td>Total</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Measuring reliable and clinically significant change

We present data from each of the three staff reported scales using conceptions appropriate to each scale of what constitutes clinical change.

Reliable change. Before getting into measures of clinical change, though, we looked at reliable change in two staff-rated instruments. Some analysts say that only change that is reliable, that is not due to the unreliability of measurement, should be considered as candidates for clinical change. We computed the Reliable Change Index for the GAF. This index attempts to correct for the fact that due to unreliability of rating instruments, a proportion would change just by chance. The RCI excludes from the outcome measurement the persons whose change score falls within the bounds which could be attained by chance alone (using a 95% confidence interval). Based on the standard deviation of the baseline GAF and a positive reliability of .88, it turns out that the change for 340 of the 475 persons with pre and post scores had a difference pre to post that could be due to chance. There were 14 out of the 475 usable results where there was reliable negative change (an average of -10.6 points). There were 121 persons with reliable positive change (an average of 18.1 points).

The second scale we use, the Clinical Global Impressions scale, did not achieve reliability for 175 persons (change was less than might have occurred by chance). On this scale, change of over .92 was reliable. So 56 persons had reliable negative change and 240 had reliable improvement.

Since these data demonstrate strong positive change even adjusting for unreliability of measurement, the discussion of clinical change will use all of the data.

Clinically significant change. This is a difficult concept, both in how to define it and how to measure it. However, it is important because statistical significance alone may not indicate that a patient has improved to a meaningful extent in the view of either patient or clinician. One approach is to use the widely recognized concept of effect size, which standardizes the

29 There are multiple versions of this index. We used the basic one presented by Chris Evans at: http://www.psyctc.org/stats/rcsc.htm
amount of change by taking account of the variability at baseline,\(^{30}\) in this regard being similar to the Reliable Change Index. Like statistically significant change, this is a group measure not one that is applied to individuals.

In some scales a definition of clinical change is built-in. The GAF is described in 10 point intervals, so it would seem to make sense to call a change of 10 points “clinically significant.” In addition, the GAF rating categories are described in such a way that anything over 70 could describe normal reactions to stressors. So moving from below 70 to over 70 would be a test of whether “normal” functioning was restored.

**CLINICAL CHANGE: GLOBAL ASSESSMENT OF FUNCTIONING (GAF)**

Here are three measures of clinically significant change in GAF scores from baseline to discharge. The second focus on specific achievements of individuals rather than group means.

- The effect size using Cohen’s \( d \) comes out to .44, or close to a medium effect. One way of thinking of it is that given an effect size of .44, if the pre-score was a 50, then the post-score would be a 67. A second interpretation is that about 25% of the post-scores do not overlap with the pre-scores because they are higher.\(^{31}\)

- The actual amount of individual change is used to characterize clinical significance in the table below.

**Table 7: “Clinically significant” change in GAF “difference” scores from baseline to discharge**

<table>
<thead>
<tr>
<th>GAF Difference Score</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative change</td>
<td>55</td>
<td>11.58</td>
<td>11.58</td>
</tr>
<tr>
<td>No change</td>
<td>179</td>
<td>37.68</td>
<td>49.26</td>
</tr>
<tr>
<td>1-9 increase</td>
<td>132</td>
<td>27.79</td>
<td>77.05</td>
</tr>
<tr>
<td>Clinically significant: 10+ increase</td>
<td>109</td>
<td>22.95</td>
<td>100</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>475</strong></td>
<td><strong>100</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

As shown in Table 7, a decline in GAF functioning occurred for 11.6% and another 37.7% had no change at all, for a total of 49% without positive change. A positive

---

\(^{30}\) This conception was developed by Jacob Cohen and is used widely in the social sciences. Cohen, J. (1988) Statistical power analysis for the behavioral sciences, 2nd ed. Hillsdale, NJ: Lawrence Earlbaum Associates. The effect size is the difference pre to post divided by the standard deviation of the pre score, so it is essentially a statement of the difference in standard deviation units. There are modifications of the approach in more recent literature. By convention:

\( d = 0.2 \) is a small effect size

\( d = 0.5 \) is a medium effect size

\( d = 0.8 \) is a large effect size

\(^{31}\) Using the reliable change index, we also found 25% that had reliable positive change on the GAF.
change, but less than ten points, occurred for 28%; and 23% improved their GAF scores by 10 points or more.

We used ordinal logistic regression to model these categories and looked for client or service variables that predict membership in the different categories. We found only two predictive variables. Participants with two or more children under five were more likely to have negative or no change and less likely to have either level of positive change. The other associated variable was being in alcohol/drug recovery at baseline. Persons in recovery were more likely to have no change and less likely to have positive change.

- The final measure is the percentage who moved from below “normal” (as defined above) to the normal range of 70 to 100. At baseline only 3 persons had valid GAF scores over 70. At follow up 31, or 6.6% had moved into the 71 or over category of “normal” GAF scores. From this analysis, at least it does not appear that CalWORKs mental health treatment will achieve the program goal of “removing barriers to employment” for most clients if that is defined as moving clients into a range of functioning that could be considered normal (albeit with manageable stresses).

In considering all three measures we should be quite cautious. Each assumes that the GAF was measured reliably at two points in time. While GAF reliability is quite good (up to .90) in research studies staff were not trained specifically on the GAF, so reliability in this sample may be lower. However, even reliability of .90 is considered somewhat low for measuring differences of single individuals as opposed to comparing groups.

In summary, we see highly statistically significant improvement with close to a medium effect size overall. Looking at individuals, about half make some positive change and a quarter show positive change that is clearly clinically significant but few moved into a normal range.

**CLINICAL CHANGE: THE CLINICAL GLOBAL IMPRESSIONS SCALE (CGI)**

We also use three different measures of clinical significance for the CGI.

- The first clinically significant change measure is effect size. The difference between pre and post scores leads to an effect size of .59, which is classified as somewhat more than a medium sized effect. If the pre-scores were at 50 on a hundred point scale the post-scores would be at 77. And the second interpretation is that there is no overlap between 38% of the pre and post scores.

- The highest two categories of the CGI are “normal” and “borderline normal.” In the latter case there is suspicion that there is a psychiatric disorder but it is not conclusive. The third category is “mild” disability and it requires that symptoms be clear. So one

---

32 DMH staff report that clinics often have understandings about what GAF scores are appropriate, e.g. that scores of over 70 would not be given. Clearly a few were, but it seems likely that the CGI is more valid at the top end of the scale as no clinician norms exist for it.

33 Reliability: intraclass r is .91 among patients. Test-retest r ranged from .62 to .82. Spitzer, Gibbon, Williams, & Endicott (1996) GAF scale. In Sederer & Dickey (Ed.) Outcomes Assessment in Clinical Practice, p.76-78.
measure of clinically significant change is to see what percentage start in categories 3 to 7 (mild symptoms to extremely ill) but during the course of treatment make progress and end in the top two categories.

By discharge 9.6% of participants who started with a psychiatric disorder were classed by clinicians as “normal” and another 5.6% were classed as borderline with no clear symptoms. Thus 15% could be considered to have clinically significant change from pre to post measurements using the scale definition.

**Table 8: Distribution of Discharge CGI scores if participants were in categories 3 – 6 at baseline (no participants were in category 7)**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - Normal, no psychiatric disorder</td>
<td>43</td>
<td>9.6</td>
<td>9.6</td>
</tr>
<tr>
<td>2 - Borderline psychiatric disorder</td>
<td>25</td>
<td>5.58</td>
<td>15.18</td>
</tr>
<tr>
<td>3 - Mildly ill, clear symptoms, minimal dysfunction</td>
<td>160</td>
<td>35.71</td>
<td>50.89</td>
</tr>
<tr>
<td>4 - Moderately ill, overt symptoms</td>
<td>153</td>
<td>34.15</td>
<td>85.04</td>
</tr>
<tr>
<td>5 - Markedly ill, intrusive symptoms</td>
<td>61</td>
<td>13.62</td>
<td>98.66</td>
</tr>
<tr>
<td>6 - Severely ill, disruptive pathology</td>
<td>6</td>
<td>1.34</td>
<td>100</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>448</strong></td>
<td><strong>100</strong></td>
<td></td>
</tr>
</tbody>
</table>

- Of course, one could argue that it is clinically as important to move from more severe to milder symptoms; that is not a “cure” but it is clinically significant in the same way moving 10 points on the GAF scale is.

So we use a similar categorization for the CGI. As shown in Table 9, 12% had negative change from baseline to discharge (or end of study for a few); no change is found for 38%; a one step increase in mental health was registered for 30% and 20% had at least a two step increase. (These percentages are very similar to those for the GAF in Table : 12%, 38%, 28% and 23.).
Table 9: “Clinically significant” change in CGI “difference” scores from baseline to discharge

<table>
<thead>
<tr>
<th>Change on the CGI</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative change</td>
<td>56</td>
<td>11.64</td>
<td>11.64</td>
</tr>
<tr>
<td>No change</td>
<td>185</td>
<td>38.46</td>
<td>50.1</td>
</tr>
<tr>
<td>1 step increase</td>
<td>142</td>
<td>29.52</td>
<td>79.63</td>
</tr>
<tr>
<td>2-5 step increase</td>
<td>98</td>
<td>20.37</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>481</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

CLINICAL CHANGE: THE CLINICAL GLOBAL IMPROVEMENT SCALE

At discharge we asked clinicians themselves to judge and calibrate “clinically significant” change using the Clinical Global Improvement scale. This scale asks clinicians directly how much clients improved. Because there was no baseline measure, there is no measured change other than the clinician’s judgment of the client’s progress over time. It is interesting that unlike the GAF and the CGI the improvement scale shows almost no one who was more ill at discharge than at baseline. This perhaps indicates some unreliability and a bias among those using the scale to see more positive change than when pre-post measurements are used.

Table 10: Clinicians’ judgment of improvement from baseline to discharge

<table>
<thead>
<tr>
<th>Improvement</th>
<th>N</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - Ideal improvement</td>
<td>31</td>
<td>6.08</td>
<td>6.08</td>
</tr>
<tr>
<td>2 - Very considerable improvement</td>
<td>43</td>
<td>8.43</td>
<td>14.51</td>
</tr>
<tr>
<td>3 - Considerable improvement</td>
<td>55</td>
<td>10.78</td>
<td>25.29</td>
</tr>
<tr>
<td>4 - Moderate improvement</td>
<td>90</td>
<td>17.65</td>
<td>42.94</td>
</tr>
<tr>
<td>5 - Slight improvement</td>
<td>88</td>
<td>17.25</td>
<td>60.2</td>
</tr>
<tr>
<td>6 - Very slight improvement</td>
<td>42</td>
<td>8.24</td>
<td>68.43</td>
</tr>
<tr>
<td>7 - Status unchanged.</td>
<td>159</td>
<td>31.18</td>
<td>99.61</td>
</tr>
<tr>
<td>8 - Minimally worse</td>
<td>2</td>
<td>0.39</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>510</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

The main variable affecting scores is the point at which discharge occurred, which translates into the time duration of treatment during which improvement might have occurred. In Figure 5 below remember that higher scores mean less improvement as shown in Table 10. We see here that in roughly the first six months—in the period from October 1 to February 2 and then to a lesser extent in the period from February 3 to May 25—improvement is quite limited. The average rating is “slight improvement.” Subsequently the ratings go down to around “moderate improvement” but don’t change that much from period to period. Contrast this with the amount of change on the GAF scale reported in Table 4, where there is significant change over each time period.

34 The GAF indicates more health as one gets a higher score; the CGI does the opposite. So the CGI scale was reversed for this analysis.
Summary: Change on scales measuring psychiatric disability was characterized by both statistically significant change of the group over time and by clinically significant change for individuals.

Group change

- GAF scores at discharge increased steadily over time with an overall increase of 4.5 points (on a hundred point scale) from baseline to discharge. Those discharged after a year in the program showed a 7.9 average change. These changes were statistically significant, and the “effect size” of this change is .44, close to a medium sized effect (.50).
- The CGI also recorded a statistically significant change from baseline to discharge; the effect size was .59, so greater than a medium effect but less than a “large” effect (.80).
- The consumer form used the K6 and Sheehan Disability scales to measure psychiatric status. However, the high attrition meant that only 113 persons were present at baseline and the first two quarters. On all three scales there was a statistically significant change over three three quarters.
- The consumer scale also showed more participant satisfaction with their services among those who were there all three quarters. By the end of the second quarter, 93% were at least “very satisfied” and 99% said services helped at least “some” (71% said they “helped a lot”).

Reliable and clinically significant individual change

- Because the reliability of the GAF and CGI, though high, is not perfect we tested adjusting outcomes to take account of scale unreliability. After doing so, the GAF showed reliable positive change for 121 persons and reliable negative change for
14 persons. The CGI showed reliable positive change for 240 and reliable negative change for 56.

- Looking at GAF change (no adjustment for reliability) we found negative change for 12%, no change for 38%, positive change of 11-9 points for 28% and clinically significant change of 10 points or more for 23%. Only 7% moved from below the “normal” range of 70 -100 into the normal range.

- For the CGI we used slightly different measures and found that 15% had clinically significant change as judged by having moved into parts of the scale with no symptoms or unclear symptoms. When measured like the GAF, 12% had negative change from baseline to discharge (or end of study for a few); no change is found for 38%; a one step increase in mental health was registered for 30% and 20% had at least a two step increase. So both instruments were similar in this regard.

- The Clinical Global Improvement scale rates change at the time of discharge. Based on its categories 6% made “ideal improvement;” another 8% made “very considerable improvement;” and 11% made “considerable improvement.”
C. POLICY ISSUES RAISED BY THE OUTCOME DATA

The policy questions we focus on here were originally raised in Outcomes of Los Angeles County CalWORKs Mental Health Services published in May 2015 (available in PDF from DMH administrators and referred to hereafter as the Outcome Study of May 2015). That report used data originally collected for the first phase of a study on IPS supported employment. The data were very detailed and were multi-sourced, so offered a wealth of information. However, the data were not designed to be a representative sample of the entire system. In contrast, the Outcome Monitoring study includes all participants entering services in a three month period in the fall of 2014. Thus, if conclusions from the May report are confirmed by data from the Monitoring implementation study, conclusions that might warrant policy interventions will be based on representative and reliable information.

The policy issues addressed are:

• Does monitoring data suggests that CalWORKs MH might look at whether the 23% of persons with very mild or non-existent disorders need care from licensed mental health professionals?
• Can we confirm the 15% to 18% “completion of treatment” figure that has been steady in our studies for 15 years? What does what we find indicate about the possibility of using quality-improvement efforts focused in each provider (but with an overall design) to improve attendance and completion?
• How much provider variability is there with regard to the issues above and in success in finding employment? From the standpoint of administrators wishing to improve system outcomes this is a key issue.
• Can we confirm that 20% to 28% of clients have domestic abuse issues and that these issues appear to negatively affect their success in CalWORKs mental health treatment?

POLICY ISSUE 1: Do all clients appear to need the level of professional mental health care offered by CalWORKs mental health services?

Current CalWORKs mental health screening and admission criteria may not be assuring that those with greatest psychiatric needs are being served. The May 2015 Outcomes Study found that clinicians using the CGI classified 9.5% of participants at baseline as normal and another 14% as having only subtle, suspected or mild symptoms. This may indicate that the screening processes, in place for many years now, may not be referring those most in need of clinical mental health services. Since about 6% of the welfare-to-work population is served in CalWORKs mental health and since most epidemiological studies show 3 to 5 times as many as this need services, a significant pool of potential referrals with severe symptoms appears to exist. Aside from whether the most impaired are offered treatment, DMH administrators need to know whether programs are accepting clients who do not have a level of mental health
disorder that justifies, or is likely to respond to, care from licensed professionals. If so, alternative services such as peer self-help groups should be made available along with case management or supported employment.

We can use the monitoring implementation study to verify the finding from the May 2015 Outcomes Study that there are persons accepted for treatment who are at the very low end of need. We would have to do another study of a different kind in order to determine whether there are large numbers of persons with need in the DPSS population who are not being referred by DPSS. If the level of need in referrals increased, presumably the issue in the programs would disappear.

We will be using the same instruments we looked at in part one. There we focused on change from baseline to discharge. Here the focus is on the question of whether all participants need care from licensed mental health professionals.

GLOBAL ASSESSMENT OF FUNCTIONING SCALE

The Global Assessment of Functioning Scale (GAF) measures a composite of symptoms and how they affect functioning in work, school and family. It is rated by clinicians at intake and at discharge. The GAF is known to be a reliable instrument when staff are trained in applying it the same way across programs, but there are conceptual difficulties with it—such as whether staff put more emphasis on functioning or symptoms.

Table 11 shows that on the GAF there are essentially no participants in the normal range of 70 or over. As we will see this contrasts with the CGI, in which about 23% are classed as essentially normal.

Table 11: Admit GAF scores showing range: 70 and above is within the normal range

<table>
<thead>
<tr>
<th>GAF Scores</th>
<th>N</th>
<th>Percentage</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 20</td>
<td>4</td>
<td>0.62</td>
<td>0.62</td>
</tr>
<tr>
<td>21-30</td>
<td>4</td>
<td>0.62</td>
<td>1.23</td>
</tr>
<tr>
<td>31-40</td>
<td>42</td>
<td>6.48</td>
<td>7.72</td>
</tr>
<tr>
<td>41-50</td>
<td>256</td>
<td>39.51</td>
<td>47.22</td>
</tr>
<tr>
<td>51-60</td>
<td>270</td>
<td>41.67</td>
<td>88.89</td>
</tr>
<tr>
<td>61-70</td>
<td>69</td>
<td>10.65</td>
<td>99.54</td>
</tr>
<tr>
<td>71-80</td>
<td>2</td>
<td>0.31</td>
<td>99.85</td>
</tr>
<tr>
<td>81-100</td>
<td>1</td>
<td>0.15</td>
<td>100.00</td>
</tr>
<tr>
<td>Total</td>
<td>648</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>

---

35 DMH administrators point out this may reflect clinic informal norms which are designed to ensure clients do not look too “well.”
CLINICAL GLOBAL IMPRESSIONS SCALE

As discussed above, the Clinical Global Impressions Scale is a widely used measure of symptomatology which goes from “Normal, no psychiatric disorder” to “Extremely ill, may need hospitalization.” Note that it does not measure specific symptoms, such as anxiety or depression, but focuses on the sum total of symptoms. Earlier we looked at change over time, here we focus on severity at baseline.

Table 12: Baseline symptom severity on the CGI

<table>
<thead>
<tr>
<th>Description</th>
<th>Number</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - Normal – no psychiatric disorder</td>
<td>18</td>
<td>2.73</td>
<td>2.73</td>
</tr>
<tr>
<td>2 - Borderline psychiatric disorder, subtle or suspected pathology</td>
<td>18</td>
<td>2.73</td>
<td>5.46</td>
</tr>
<tr>
<td>3 - Mildly ill, clear symptoms, minimal distress or difficulty in social/occupational functioning</td>
<td>117</td>
<td>17.75</td>
<td>23.22</td>
</tr>
<tr>
<td>4 - Moderately ill, overt symptoms, noticeable but modest functional impairment</td>
<td>319</td>
<td>48.41</td>
<td>71.62</td>
</tr>
<tr>
<td>5 - Markedly ill, intrusive symptoms, distinctly impaired social/occupational function</td>
<td>166</td>
<td>25.19</td>
<td>96.81</td>
</tr>
<tr>
<td>6 - Severely ill, disruptive pathology, behavior and function compromised, requires assistance from others</td>
<td>20</td>
<td>3.03</td>
<td>99.85</td>
</tr>
<tr>
<td>7 - Extremely ill, drastic interference in function, may need hospitalization</td>
<td>1</td>
<td>0.15</td>
<td>100.00</td>
</tr>
<tr>
<td>Total</td>
<td>659</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>

The fact that in the monitoring implementation study 2.7% are rated as “normal” and another 2.7% as having only subtle or suspected pathology raises questions about whether these 5.4% are appropriate for CalWORKs mental health. Another 18% have “minimal” distress or difficulty in social or occupational functioning” (that is, they have stressors but are coping with them), for a total of 23% who by this measure seem marginally suited for professional treatment by licensed mental health staff.

CGI differences by SPA

From a policy point of view, this issue is very different if the 23% occur fairly uniformly at each provider than if some providers have far more “mild” cases and others far more
“severe” cases. In the latter case, remediation would focus on the assessment practices of particular providers rather than on changing the screening process.\textsuperscript{36}

In Table 13 below we show the CGI mean and standard deviation for each SPA. Recall that the CGI is scored so that lower scores mean less psychiatric disability. Only SPA 3 and SPA 7 are statistically different from the overall mean, though overall the differences between the SPA means is highly statistically significant.

Table 13: Overall severity of caseload based on CGI, by Service Planning Area (lower scores mean less severe psychiatric symptom profiles)

<table>
<thead>
<tr>
<th>Region</th>
<th>N</th>
<th>Baseline Mean GAF</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPA A</td>
<td>73</td>
<td>4.15</td>
<td>0.76</td>
</tr>
<tr>
<td>SPA B</td>
<td>99</td>
<td>4.02</td>
<td>1.08</td>
</tr>
<tr>
<td>SPA C</td>
<td>69</td>
<td>3.78</td>
<td>1.11</td>
</tr>
<tr>
<td>SPA D</td>
<td>58</td>
<td>4.09</td>
<td>0.76</td>
</tr>
<tr>
<td>SPA D</td>
<td>16</td>
<td>4.19</td>
<td>0.54</td>
</tr>
<tr>
<td>SPA F</td>
<td>120</td>
<td>3.89</td>
<td>1.11</td>
</tr>
<tr>
<td>SPA G</td>
<td>136</td>
<td>4.03</td>
<td>0.93</td>
</tr>
<tr>
<td>SPA H</td>
<td>87</td>
<td>4.07</td>
<td>0.83</td>
</tr>
<tr>
<td>Total</td>
<td>658</td>
<td>4.00</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Below we show the percentage of cases served in each SPA in which the baseline clinical rating was “normal,” “borderline” or “mild.” These differences by SPA are statistically significant.

Table 14: Percentage of admissions with “mild” or no symptoms based on CGI, by SPA

<table>
<thead>
<tr>
<th>Provider</th>
<th>Cases</th>
<th>% Mild CGI</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPA A</td>
<td>73</td>
<td>14%</td>
</tr>
<tr>
<td>SPA B</td>
<td>110</td>
<td>19%</td>
</tr>
<tr>
<td>SPA C</td>
<td>73</td>
<td>29%</td>
</tr>
<tr>
<td>SPA D</td>
<td>61</td>
<td>18%</td>
</tr>
<tr>
<td>SPA D</td>
<td>16</td>
<td>6%</td>
</tr>
<tr>
<td>SPA F</td>
<td>123</td>
<td>33%</td>
</tr>
<tr>
<td>SPA G</td>
<td>142</td>
<td>22%</td>
</tr>
<tr>
<td>SPA H</td>
<td>88</td>
<td>19%</td>
</tr>
<tr>
<td>Total</td>
<td>686</td>
<td>22%</td>
</tr>
</tbody>
</table>

In order to begin to see how the GAF and CGI results relate to each other, it is useful to look at GAF averages among those that the CGI scale classifies as normal, borderline, or mild (the top 3 categories). Table 15 shows the baseline GAF scores for the top three CGI categories. Only 21% of those rated in the top three CGI categories were at 50 or

\textsuperscript{36} Though it could focus on both if the pattern appears mixed: some universal but some provider specific effects.
lower and about 30% of this group was rated over 60 on the GAF. Another way to look at it is that the overall average score on the GAF is 51.9. For those in the least severe CGI categories, 75% are above this average.

These comparisons do show a strong correlation between the measures, but they also indicate that relying on the CGI scale alone for determining whether some clients may not need this level of services could introduce error in perhaps 25% of cases. In practice, of course a provider would never rely on one instrument except for screening. Instead a clinical assessment makes the determination of need for treatment. But the variability shown here so far, raises questions about whether services by licensed professionals are needed in all cases.

Table 15: GAF scores for participants in the four bottom CGI categories versus those in the three top CGI categories, percentage in each GAF category

<table>
<thead>
<tr>
<th>GAF Score</th>
<th>Moderate to severe CGI N 497</th>
<th>3 Least severe CGI categories N 152</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 50</td>
<td>54.93%</td>
<td>21.71%</td>
</tr>
<tr>
<td>51-60</td>
<td>39.44%</td>
<td>49.34%</td>
</tr>
<tr>
<td>61-70</td>
<td>5.63%</td>
<td>26.97%</td>
</tr>
<tr>
<td>71-80</td>
<td>0.00%</td>
<td>1.32%</td>
</tr>
<tr>
<td>81-90</td>
<td>0.00%</td>
<td>0.66%</td>
</tr>
</tbody>
</table>

LEVEL OF CARE RATINGS

There is another measure staff provide at baseline that should correlate with both GAF and CGI severity and is directly related to the issue of whether some participants need care from licensed professionals. Programs are required to classify clients by level of care, which is a measure of intensity of services.

In the table below we show the number of clients in each level of service with their average GAF score and average CGI score. The most intensive level, Level 1, allows treatment of from 1 to 10 hours per week and includes 29% of the sample. Level 2 allows up to 7 hours a week and includes 52% of the sample. Level 3 only allows up to 4 hours a week and includes 18% of the sample. Recalling that less severe is the lower numbers for CGI and the higher numbers for GAF, we see that there is some correlation between the level of care clients are assigned and their GAF and CGI scores. Differences are not great, however, as the GAF for Level 2 and Level 3 are almost identical.

\[37\] The 10 cases in which a “99” was scored as a placeholder were dropped.

\[38\] Again, this may reflect clinic norms about how GAF scores are assigned.
Table 16: Level of Care in Relationship to Measures of Severity

<table>
<thead>
<tr>
<th>Level of Care</th>
<th>Mean GAF at intake (lower is more severe)</th>
<th>Mean CGI score at intake (lower is less severe)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – Level of care</td>
<td>50.2</td>
<td>4.4</td>
</tr>
<tr>
<td>2 – Level of care</td>
<td>52.6</td>
<td>4.0</td>
</tr>
<tr>
<td>3 – Level of care</td>
<td>52.7</td>
<td>3.5</td>
</tr>
</tbody>
</table>

These scores are averages. Are there a significant number of outliers in the three levels of care? That is, are there persons whose CGI scores would suggest a level of care different from their assignment? The graph below is a dotplot. Each client is a round dot (many dots at the same score look like a bar with rounded ends). We see some very severe clients, based on CGI, in all three LOC categories. The CGI scores raise the question on both ends of the distribution: why are persons with CGI scores of 3 or less (mild or non-existent mental illness) present in those getting up to 7-10 hours of treatment a week; and why are persons assigned a 5 (markedly ill) assigned to 1-4 hours a week? Undoubtedly there are clinical reasons for most such decisions but the data make it clear that there are some anomalies in the classification process. Identifying some of these in the DMH site review process and inquiring about the clinical justification might encourage LOC classification accuracy in providers.

Figure 6: Distribution of CGI scores in the three LOC categories (note mild symptoms on the Severity axis are rated 1-3, moderate are a 4, and severe are 5-7)

The K6 Client Self-Rating Scale

The scales we have looked at so far codify the impressions of staff clinicians. Many studies, however, have documented that client self-report can be significantly different from staff views.
The Mental Health Surveillance Study sponsored by SAMHSA summarized the percent of the population expected, on the basis of two previous calibration studies, to have particular K6 scores and the percentage at each score expected to have some sort of serious mental distress. These are shown in Table 1.

Table 17: Outcome Monitoring K6 scores compared to US averages

<table>
<thead>
<tr>
<th></th>
<th>Baseline Percent</th>
<th>US Population Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=447</td>
<td></td>
</tr>
<tr>
<td>0-3</td>
<td>7.2</td>
<td>48.0</td>
</tr>
<tr>
<td>4-5</td>
<td>6.9</td>
<td>14.0</td>
</tr>
<tr>
<td>6-7</td>
<td>5.4</td>
<td>11.2</td>
</tr>
<tr>
<td>8-9</td>
<td>11.4</td>
<td>7.0</td>
</tr>
<tr>
<td>10-11</td>
<td>10.3</td>
<td>5.5</td>
</tr>
<tr>
<td>12-15</td>
<td>29.3</td>
<td>8.0</td>
</tr>
<tr>
<td>16 or higher</td>
<td>29.5</td>
<td>6.3</td>
</tr>
<tr>
<td>Total</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

The K6 scores contrast greatly with scores in the general population survey. While 48% of the general population has scores of 0 to 3, only 5-7% of the CalWORKs mental health clients do. At the other end of the scale 6.3% of the general population have scores of 16 or higher (which indicates a 67% likelihood of serious psychological distress) compared to 30-37% of the CalWORKs mental health samples. So, as we would expect, the K6 is showing far more CalWORKs mental health clients to have core symptoms of serious mental illness than a cross-section of Americans.

However, the K6 results do not show that a high percentage of clients exceed the cut-off for “possible serious mental illness.” Less than one third have scores that correspond to that cut-off of 16 or more.

---

39 Serious mental illness is defined by federal regulations as “Persons aged 18 and over, who currently or at any time during the past year, have had diagnosable mental, behavioral, or emotional disorder of sufficient duration to meet diagnostic criteria specified within DSM that has resulted in functional impairment, which substantially interferes with or limits one or more major life activities. All of these disorders have episodic, recurrent, or persistent features; however, they vary in terms of severity or disabling effects. Functional impairment is defined as difficulties that substantially interfere with or limit role functioning in one or more major life activities including basic daily living skills (e.g. eating, bathing, dressing; instrumental living skills (e.g. maintaining a household, managing money, getting around the community, taking prescribed medication; and functioning in social, family, and vocational/educational contexts.” CalWORKs clients would be expected to fit in this general category because they must have a DSM diagnosis which is interfering, at least, with their financial independence and often with multiple functional domains. See the CIBHS recent report on CalWORKs mental health outcomes that includes functional challenges: Outcomes of Los Angeles County CalWORKs Mental Health Services, May 2015.
**K6 scores in relationship to staff rated scales**

Since we know the arbitrary study ID that links staff and client forms, we can look at whether client self-assessment of symptoms corresponds to that of staff on the CGI scale. In the *Outcomes* Study of May 2015 we did not have ratings that corresponded precisely in terms of time but these are very close.

Because there are small numbers of participants in the first and last two categories of the CGI, we look at the average K6 ratings in terms of “mild” (top 3 CGI categories, “moderate” (middle category, and “severe” (the last three categories. These show that the K6 consumer measure corresponds in trend to the CGI categories:

**Table 18: Relationship of staff CGI to client K6 scores**

<table>
<thead>
<tr>
<th>Broad CGI Categories</th>
<th>N</th>
<th>Mean K6 Symptom Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – Mild severity</td>
<td>92</td>
<td>10.3</td>
</tr>
<tr>
<td>2 – Moderate</td>
<td>237</td>
<td>12.3</td>
</tr>
<tr>
<td>3 – Severe</td>
<td>116</td>
<td>13.7</td>
</tr>
<tr>
<td>Total</td>
<td>445</td>
<td>12.2</td>
</tr>
</tbody>
</table>

However, a similar comparison (not shown) using GAF scores did not reveal that the scores are related in any predictive way. This may be because GAF combines symptoms and functioning, while both CGI and K6 are symptom scales.

While the K6 scores of persons admitted to CalWORKs mental health in this study indicate that as a group they are far from “normal” they also provide further support for the possibility that some participants have symptom profiles below the threshold of what is needed to justify licensed professional care.

**THE SHEEHAN DISABILITY SCALE**

Because the K6 measures self-reported symptoms but not functional impairment, it is usually administered along with the Sheehan Disability Scale.

In the baseline monitoring survey 69% of participants reported at least moderate disruption of work, but 33% reported no disruption or mild disruption.

**Table 19: Sheehan Disability Sub-Scales: Disrupted Work**

<table>
<thead>
<tr>
<th>DISRUPTION of work</th>
<th>Baseline of Monitoring Study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=447 Percent</td>
</tr>
<tr>
<td>Not at all</td>
<td>12.30</td>
</tr>
<tr>
<td>Mildly</td>
<td>20.81</td>
</tr>
<tr>
<td>Moderately</td>
<td>36.24</td>
</tr>
<tr>
<td>Markedly</td>
<td>21.70</td>
</tr>
<tr>
<td>Extremely</td>
<td>8.95</td>
</tr>
<tr>
<td>Total</td>
<td>100.00</td>
</tr>
</tbody>
</table>
The disrupted family question reads: “Rate how much during the past week your mental health problems have disrupted your family life and home responsibilities.”

In the baseline monitoring survey 65% reported at least moderate disruption of family but again 33% reported none at all or only mild disruption.

**Table 20: Sheehan Disability Sub-Scales: Disrupted Family Life**

<table>
<thead>
<tr>
<th>DISRUPTION of family</th>
<th>Baseline of Monitoring Study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=447</td>
</tr>
<tr>
<td></td>
<td>Percent</td>
</tr>
<tr>
<td>Not at all</td>
<td>9.62</td>
</tr>
<tr>
<td>Mildly</td>
<td>22.82</td>
</tr>
<tr>
<td>Moderately</td>
<td>32.89</td>
</tr>
<tr>
<td>Markedly</td>
<td>28.41</td>
</tr>
<tr>
<td>Extremely</td>
<td>6.26</td>
</tr>
<tr>
<td>Total</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

In the baseline of the monitoring study at least moderate disruption for both work and family at the same time was reported by 61% and Extreme disability for work or family occurred in 9%. This does leave a third whose disability would be categorized as mild by Sheehan.

In summary, there is a large degree of consistency in how CalWORKs mental health clients rate their own work or family disability: about two thirds believed they had at least moderate functional impairments due to mental health problems. The other third provide client self-report confirmation that some participants may not need the level of clinical services provided by CalWORKs mental health.

**Summary and potential remedies to possible mismatch of services and severity**

The findings reported here---of substantial percentages of CalWORKs mental health clients without severe symptoms and with some SPAs having much higher percentages than others—point in two possible directions. First, it may be that DPSS screening is not sending enough persons with moderate and severe symptoms to be assessed—despite there being many more in the CalWORKs population than can be served with existing resources. Secondly, it may be that internal clinic processes of assessing clients and planning treatment need to be evaluated and/or monitored in at least some clinics.

- DPSS CalWORKs mental health screening and referral has not been investigated for some time. There are better tools available now than there were in 1999. So DMH might suggest to DPSS a special study next year of screening and referral that focuses on improving both sensitivity (getting the right people) and
specificity (not getting the wrong people). The K6 and Sheehan would be candidates for screening tools.

- If some programs have a particularly high rate of persons with less severe mental health problems it may be because they are not getting enough referrals. A study to look at number of referrals, waiting lists, and percent of assessments admitted would permit getting a handle on this possible cause.

- DMH might also recognize the serious family and employment issues some clients may have even when their mental health issues are mild. In that case, these clients might be offered only case management and peer-support mental health services. We know that some programs already operate this way.

- Anomalies between outcome monitoring instrument scores and LOC placement might be investigated during site visits. CIBHS can supply DMH a list of these cases if desired.

**POLICY ISSUE 2: Can low treatment completion rates be improved?**

The CalWORKs mental health system problem most in need of improvement is the seeming difficulty providers have in helping clients resolve the many issues that cause disengagement. Completing treatment is a good predictor of positive employment and mental health outcomes. In the *Outcomes Study* published in May 2015, as in our previous Los Angeles studies over the past 10 years, only around 15% of participants complete their treatment goals. Results of the *Outcomes Study* indicate that addressing attendance problems early might prevent some negative discharges. That study also found that levels of the admission GAF score were predictive of completing treatment in multiple regression analysis. Finally, it suggested that there is very wide variability by provider. We use outcome monitoring demonstration study data to explore these issues farther.

Table 21 shows the percentage of persons discharged for each of 13 reasons. For analysis we have reclassified these reasons into those that are positive (the 15% who met goals and the 10% who were in school or work so much it interfered with treatment); those that are neutral (client choosing to leave 4% but not dissatisfied, moving 3%, and becoming eligible for other income or insurance 1% including SSI), and those that are negative (non-compliance or drop-out 40%, practical difficulties that CalWORKs or clinic did not help with 2%, losing CalWORKs eligibility 8%, or dissatisfaction with services 1%). We did not have enough information to classify the 6% with “other” reasons. In summary, when we leave out the unclassifiable 5%, positive reasons were 26%, neutral were 10%, and negative were 64%.
### Table 21: Rating by staff at discharge of reasons clients had left treatment

<table>
<thead>
<tr>
<th>Reason for Leaving Treatment</th>
<th>N</th>
<th>Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met goals</td>
<td>73</td>
<td>14.93</td>
<td>14.93</td>
</tr>
<tr>
<td>Client choice—not dissatisfaction</td>
<td>22</td>
<td>4.5</td>
<td>19.43</td>
</tr>
<tr>
<td>Work or school interfered with treatment</td>
<td>48</td>
<td>9.82</td>
<td>29.24</td>
</tr>
<tr>
<td>Dissatisfaction with services</td>
<td>5</td>
<td>1.02</td>
<td>30.27</td>
</tr>
<tr>
<td>Noncompliance with treatment requirements</td>
<td>55</td>
<td>11.25</td>
<td>41.51</td>
</tr>
<tr>
<td>Client is transferring to SSI, SSDI</td>
<td>3</td>
<td>0.61</td>
<td>42.13</td>
</tr>
<tr>
<td>Loss of CalWORKs eligibility (negative)</td>
<td>38</td>
<td>7.77</td>
<td>49.9</td>
</tr>
<tr>
<td>Becoming eligible for other insurance</td>
<td>6</td>
<td>1.23</td>
<td>51.12</td>
</tr>
<tr>
<td>Practical difficulties in getting treatment (such as transportation)</td>
<td>12</td>
<td>2.45</td>
<td>53.58</td>
</tr>
<tr>
<td>Provider terminated for non-compliance (other than no-shows)</td>
<td>8</td>
<td>1.64</td>
<td>55.21</td>
</tr>
<tr>
<td>Early drop out: Only attended one visit</td>
<td>24</td>
<td>4.91</td>
<td>60.12</td>
</tr>
<tr>
<td>Dropped by provider for failing to show for more than 30 days</td>
<td>151</td>
<td>30.88</td>
<td>91</td>
</tr>
<tr>
<td>Client moved out of program service area</td>
<td>13</td>
<td>2.66</td>
<td>93.66</td>
</tr>
<tr>
<td>A different primary reason for leaving</td>
<td>31</td>
<td>6.34</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>489</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

Presumably some of the negative reasons having to do with loss of eligibility for CalWORKs cannot be affected by a mental health program, but since many of the reasons for loss of eligibility have to do with compliance with CalWORKs program requirements it does not seem to be unreasonable to term those negative reasons.\(^\text{40}\)

When we drop out the neutral reasons for analysis, we get 29% positive and 71% negative.

For the outcome monitoring analysis presented here, we used all of the variables available (baseline demographics, baseline clinical measures, discharge clinical measures) to create a regression model which might predict which factors are associated with ending treatment for positive reasons. Without taking into account variability by provider, the statistically significant predictors are shown below. Each

\(^{40}\) There are neutral reasons for loss of CalWORKs eligibility, too, such as the child in question becoming too old for CalWORKs. However, if we drop the persons with possibly neutral reasons, the percentage with positive reasons only increases by 5%.
one is the predicted percentage with a positive reason for discharge with all the other variables shown held constant.

- A diagnosis of PTSD decreases the likelihood of a positive termination (15% vs. at least 25% for other diagnoses).41
- Having a lower admit GAF score increases the likelihood of a positive termination (for example, 48% would have a positive exit at GAF=40 vs. 25% at GAF=60).
- Being enrolled in an educational program at admit increases the likelihood of a positive termination: 45% vs. 29% if not in an education program.
- Being in level of care 3 at discharge (1-4 hours a week) greatly increases the likelihood of a positive termination compared to those in other levels (45% for level 3 vs 12% for level 1, and 28% for level 2).
- Being in IPS at discharge (43% positive discharge if in IPS, 29% if not in IPS)
- Having a higher GAF score at discharge increases the likelihood of a positive termination (for example, 6% would have a positive exit at GAF=40 vs. 37% at GAF=60).

That a severe entry GAF and a low severity discharge GAF are both associated with treatment success seems paradoxical, but was also found in the May 2015 Outcomes Study. The amount of variability accounted for by these predictors taken together is 36%, quite high for sociological studies. But unfortunately they are not predictors that programs can manipulate in order to increase positive discharges, except for focusing more on persons with PTSD

We then added in the variability that is attributable to different SPAs. For each SPA we calculated the percent positive discharges (neutral were not included). All of the variables discussed above as significant predictors remained significant when we reran the model with SPA differences included.

Below in Table 22 we show the percentage with positive reason for discharge for the eight regions. The variability is quite high, going from 45% to 22% (we disregard the SPA with only six cases).

41 Note that 64% of those with PTSD did not report domestic abuse to their clinician, so for most persons the PTSD probably relates to other kinds of trauma.
Table 22: Percentage of Positive Reasons for Discharge, Other Variables Held Constant, by SPA (Ordered by percent positive discharges)

<table>
<thead>
<tr>
<th>Region</th>
<th>N</th>
<th>Percent Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPA A</td>
<td>44</td>
<td>27%</td>
</tr>
<tr>
<td>SPA B</td>
<td>66</td>
<td>23%</td>
</tr>
<tr>
<td>SPA C</td>
<td>46</td>
<td>22%</td>
</tr>
<tr>
<td>SPA D</td>
<td>31</td>
<td>45%</td>
</tr>
<tr>
<td>SPA D</td>
<td>6</td>
<td>0%</td>
</tr>
<tr>
<td>SPA F</td>
<td>79</td>
<td>22%</td>
</tr>
<tr>
<td>SPA G</td>
<td>101</td>
<td>42%</td>
</tr>
<tr>
<td>SPA H</td>
<td>41</td>
<td>27%</td>
</tr>
<tr>
<td>Total</td>
<td>414</td>
<td>29%</td>
</tr>
</tbody>
</table>

On the graph in Figure 7, the diagonal line shows how much would be explained by predicting individual cases randomly. The thicker line shows the actual predictive power using the regression model we used (including SPA). If the model accurately predicted 100% of those with positive reasons for leaving, this thick line would be up in the left hand corner at 1.0. Since the area under the curve is instead .91, it indicates the model is capable of very good, if not perfect, prediction of which participants will leave for positive reasons.

**Figure 7: The regression model that includes provider variability is a good predictor of positive/negative reasons for leaving treatment**
Potential remedies to increase positive and decrease negative reasons for leaving services

1. The May 2015 Outcomes Study found that those with poor attendance are much more likely to leave early than others. Based on that study (since this study did not include attendance), we suggest testing a requirement for a home visit if attendance is poor early on.

2. Because the 10% of clients with PTSD diagnoses are especially likely not to complete treatment, it may be possible as a system or at least in individual providers to develop ways of engaging them.42

3. And we recommend a system-wide quality improvement effort in which each provider takes responsibility for identifying causes for poor attendance and early dropout, and creating, testing, and evaluating solutions.43 For example, a non-profit organization that helps behavioral health agencies improve engagement is NIATx. One tool they use is the “Walk Through,” in which an agency administrator notes issues as she walks through the intake process of the agency as if she were a consumer seeking services. But numerous other tools are available.44 There are NIATx experts available at UCLA and in Ventura Behavioral Health as well as at the national office in Madison.45

POLICY ISSUE 3: Should we be concerned about the variability among providers regarding effectiveness in helping clients?

In Policy Issue 1 policy we looked at whether some SPAs are serving too great a percentage of clients with mild disorders. As seen in Table 14, the average percentage in the normal-subtle-mild group 23%, but there are SPAs with rates as high as 33%. In Policy Issue 2 we looked at percentages with positive reasons for termination by SPA and found a range of 22% to 45%.

---

42 Although there are 67 persons with PTSD, they are spread out with only 2-4 in each clinic.

43 Many models and resources are available for helping staff carry out a quality improvement project, such as the FADE and PDSA models described at http://patientsafety.duhs.duke.edu/module_a/methods/pdsa.html. Wells has documented the success of a multi-agency QI project to improve care of depressive disorders. Wells, K. B., Sherbourne, C., Schoenbaum, M., et al. Impact of disseminating quality improvement programs for depression in managed primary care: A randomized controlled trial. (2000. Journal of the American Medical Association, 283(2) 212–220. doi:10.1001/jama.283.2.212


45 NIATx itself has been used in many mental health programs as well as substance abuse programs since the issues are basically the same regarding engagement and retention. Ventura, for example, is using NIATx in both Alcohol/Drug and Mental Health divisions.
In this section we continue to look at this variability by looking at employment rates. As shown in Table 23, the rates of participants’ increases in employment range from 8% to 21% (disregarding SPA 5 because of a low N).

<table>
<thead>
<tr>
<th>SPA</th>
<th>Cases</th>
<th>Working Baseline</th>
<th>Working Discharge</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPA A</td>
<td>60</td>
<td>10%</td>
<td>22%</td>
<td>12%</td>
</tr>
<tr>
<td>SPA B</td>
<td>77</td>
<td>16%</td>
<td>31%</td>
<td>16%</td>
</tr>
<tr>
<td>SPA C</td>
<td>52</td>
<td>17%</td>
<td>25%</td>
<td>8%</td>
</tr>
<tr>
<td>SPA D</td>
<td>33</td>
<td>24%</td>
<td>33%</td>
<td>9%</td>
</tr>
<tr>
<td>SPA D</td>
<td>6</td>
<td>17%</td>
<td>0%</td>
<td>-17%</td>
</tr>
<tr>
<td>SPA F</td>
<td>89</td>
<td>18%</td>
<td>27%</td>
<td>9%</td>
</tr>
<tr>
<td>SPA G</td>
<td>113</td>
<td>16%</td>
<td>37%</td>
<td>21%</td>
</tr>
<tr>
<td>SPA H</td>
<td>58</td>
<td>12%</td>
<td>22%</td>
<td>10%</td>
</tr>
<tr>
<td>Total</td>
<td>488</td>
<td>16%</td>
<td>29%</td>
<td>13%</td>
</tr>
</tbody>
</table>

As noted above, so far we have not presented the actual variation by provider because the relatively small sample sizes in many providers mean rates may not be stable over time. However, ultimately DMH will want to look at individual providers.

As an example, below we show variation by provider in the percentage of clients who improved their work status by the final rating. We don’t show the provider names but want to make the point that even controlling for employment at baseline, variability by provider is quite high. It is not possible for DMH to take corrective action, though, until it can show that excessive variability is consistently associated with some providers and not others.

The change in the percent working from baseline (16%) to follow-up (29%) is statistically significant and substantial though far less than ideal. However, there were large differences between providers in the amount of change, as shown in Table 24. We have omitted providers who had ten or fewer discharges. The left column shows the arbitrary provider identifier used in the study, the next column is the percentage working at baseline, the following column is percentage working at discharge or end of study, and the right hand column is the difference between those two percentages. There are providers whose clients are less likely to be working at follow-up and other providers whose clients were working at rates as high as 38% more than at baseline. However, the small sample size in most programs means that a large part of the variation we see is simply chance.

For comparisons between providers to be fair—and to be considered so by the providers themselves—we have to have more than large sample sizes. Just as when the federal government compares hospitals, we have to ensure that the differences in the patient populations served is taken account of, a process known as risk adjustment. For example, one provider might serve an area that has more drug and alcohol abuse so that its clients are likely to have mental health and substance abuse problems. It would be unfair to expect the same completion of treatment rate in that program as for one that has very few
participants with drug and alcohol co-morbidity. Likewise, comparison of employment rates in different programs has to adjust for the unemployment rate in the area the clinic serves. In this report we are just exploring the issue of variability. Figuring out how to compare programs or even SPAs fairly and then how to use those comparisons in a way providers find helpful will be a large challenge.

Table 24: Average percentage increase or decrease in percentage of clients working: Baseline to follow-up

<table>
<thead>
<tr>
<th>Arbitrary Provider ID</th>
<th>Number of Cases</th>
<th>Working at Baseline</th>
<th>Working at Follow-up</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1101</td>
<td>15</td>
<td>31%</td>
<td>44%</td>
<td>13%</td>
</tr>
<tr>
<td>1103</td>
<td>18</td>
<td>22%</td>
<td>33%</td>
<td>11%</td>
</tr>
<tr>
<td>1104</td>
<td>16</td>
<td>0%</td>
<td>17%</td>
<td>17%</td>
</tr>
<tr>
<td>1107</td>
<td>24</td>
<td>13%</td>
<td>25%</td>
<td>13%</td>
</tr>
<tr>
<td>1115</td>
<td>16</td>
<td>19%</td>
<td>6%</td>
<td>-13%</td>
</tr>
<tr>
<td>1126</td>
<td>13</td>
<td>0%</td>
<td>38%</td>
<td>38%</td>
</tr>
<tr>
<td>1129</td>
<td>19</td>
<td>12%</td>
<td>24%</td>
<td>12%</td>
</tr>
<tr>
<td>1133</td>
<td>14</td>
<td>22%</td>
<td>50%</td>
<td>28%</td>
</tr>
<tr>
<td>1135</td>
<td>12</td>
<td>25%</td>
<td>25%</td>
<td>0%</td>
</tr>
<tr>
<td>1136</td>
<td>21</td>
<td>19%</td>
<td>14%</td>
<td>-5%</td>
</tr>
<tr>
<td>1137</td>
<td>17</td>
<td>10%</td>
<td>33%</td>
<td>24%</td>
</tr>
<tr>
<td>1140</td>
<td>14</td>
<td>8%</td>
<td>31%</td>
<td>23%</td>
</tr>
<tr>
<td>1141</td>
<td>28</td>
<td>16%</td>
<td>40%</td>
<td>24%</td>
</tr>
<tr>
<td>1142</td>
<td>17</td>
<td>12%</td>
<td>24%</td>
<td>12%</td>
</tr>
<tr>
<td>1143</td>
<td>14</td>
<td>17%</td>
<td>17%</td>
<td>0%</td>
</tr>
<tr>
<td>1144</td>
<td>16</td>
<td>22%</td>
<td>56%</td>
<td>33%</td>
</tr>
<tr>
<td>1146</td>
<td>18</td>
<td>15%</td>
<td>30%</td>
<td>15%</td>
</tr>
<tr>
<td>1154</td>
<td>13</td>
<td>8%</td>
<td>42%</td>
<td>33%</td>
</tr>
</tbody>
</table>

Potential remedies for the issues surrounding excessive provider variation on important outcomes

DMH monitors providers closely with both monthly reports and on-site reviews. It may be that some of the issues raised in this report can be investigated as part of the on-site reviews. It is one thing for there to be information that shows a problem, but finding a solution that works for the program requires involvement by program staff and detailed knowledge of the processes behind the data reported here.

It might be worthwhile for DMH to convene a task force of contract and county providers to discuss possible solutions to the policy issues raised by the May 2015 Outcomes Study and this follow-up report. Although waiting until a large sample has assured all stakeholders that the data are stable would be prudent if discussion will focus on provider differences.
In any case, for DMH to use provider variation as a way to improve system performance will require methods, as yet undeveloped, that providers regard as fair and helpful.

**POLICY ISSUE 4: Do the prevalence and effects on treatment of co-occurring domestic abuse suggest a need for a new approach?**

In the May 2015 *Outcome Study*, 20 of 101 women interviewed said they felt unsafe at baseline, but 28 reported feeling unsafe at follow-up. Whether this indicates actual negative change or change in clients’ willingness to report abuse is unclear. That these domestic abuse situations affect treatment is shown by the May *Outcome Study* finding (page 46 of the report) that persons with domestic abuse at baseline score lower than those without on the self-rated Recovery Assessment Scale at baseline; but that they score even lower *after* treatment. That is, if there is no domestic abuse then RAS improves; if there is domestic abuse, then RAS declines during treatment.

The monitoring implementation and demonstration study does not include domestic abuse measures from clients, only staff ratings. So the validity of domestic abuse prevalence and change data may be lower in this study. At baseline 56% of the clients admitted to treatment were judged by staff to have neither past nor present domestic abuse. This is probably a low judgment as studies of CalWORKs participants have shown rates of lifetime abuse approximating 80% and current abuse warranting services in about 25%. Of note is the fact that in these studies current abuse declined markedly over three years: only 18% had current service needs after two years and only 8% had current service needs in all three study years. In a number of TANF studies those suffering domestic abuse have been much less likely to work than others.

Based on this distribution, at least 21% might need some assistance with domestic abuse. It would not be surprising if a more intensive screening program yielded addition cases.

---

46 Meisel J., Chandler D. and Rienzi B. Domestic violence prevalence and effects on employment in two California TANF populations. *Violence Against Women* 9(10:1191-213 (2003. It is possible that in subsequent years the prevalence of abuse in CalWORKs mental health clients has declined, but there is no reason on the face of it to assume so.
Table 25: Prevalence and severity of domestic abuse at baseline in monitoring study

<table>
<thead>
<tr>
<th>Domestic abuse at baseline</th>
<th>Percent</th>
<th>N=654</th>
</tr>
</thead>
<tbody>
<tr>
<td>No domestic abuse from current or past partners</td>
<td>58%</td>
<td></td>
</tr>
<tr>
<td>Minor abuse issues from current or past partners</td>
<td>21%</td>
<td></td>
</tr>
<tr>
<td>Moderate domestic abuse interfering with or making it difficult for client to get</td>
<td>12%</td>
<td></td>
</tr>
<tr>
<td>mental health treatment; pushing, slapping, threatening with a fist; following client;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>limiting access to income.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serious domestic abuse such as serious physical abuse or major threats</td>
<td>8%</td>
<td></td>
</tr>
<tr>
<td>Very serious past domestic abuse: client lives in a domestic violence shelter or lives</td>
<td>1%</td>
<td></td>
</tr>
<tr>
<td>in a secret location.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the table below, we show that domestic abuse rates as judged by staff declined significantly by discharge or end of study. In order to be sure the change reflects actual changes in the same persons rather than sampling differences, only those present at baseline and in the subsequent rating are shown. This fact does not confirm the finding of the May 2015 Outcomes Study that current abuse does not lessen. Again, however, it is important to recognize that the this study is using staff rather than consumer judgments.

While minor abuse changed little over the course of treatment, moderate abuse declined substantially as did serious abuse. By the second rating shown here the potential need for services had declined to 10%. Whether the situations resolved themselves or were addressed in the mental health services is not known, but the change appears positive in either case.

Table 26: Comparison of Pre and Post domestic abuse rates

<table>
<thead>
<tr>
<th>Domestic abuse at Baseline and Discharge or Most Recent Quarter</th>
<th>Baseline Percent</th>
<th>Most Recent Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=488</td>
<td>N=488</td>
</tr>
<tr>
<td>No domestic abuse from current or past</td>
<td>59.22</td>
<td>73.00</td>
</tr>
<tr>
<td>Minor abuse issues from current or past</td>
<td>19.88</td>
<td>16.63</td>
</tr>
<tr>
<td>Moderate domestic abuse</td>
<td>12.50</td>
<td>6.65</td>
</tr>
<tr>
<td>Serious domestic abuse</td>
<td>7.58</td>
<td>2.91</td>
</tr>
<tr>
<td>Very serious past domestic abuse</td>
<td>0.82</td>
<td>0.83</td>
</tr>
<tr>
<td>Total</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

In the May 2015 Outcomes Study, those with domestic abuse not only did less well in finding employment but showed declines on the self-rated Recovery Assessment Scale. We do not have the same instruments in this study, but can assess the same issues.
**Does the presence of domestic abuse affect employment?** In research studies domestic abuse has had paradoxical effects: for some women it spurs them to find employment in order to become independent, while for others it lessens the changes for becoming employed. In this study, neither domestic abuse at baseline or at discharge was predictive of whether clients were working the week before the final staff rating.

**Does the presence of domestic abuse affect the reason for discharge?** When we looked at whether the reasons for discharge were positive or negative, only the experience of minor abuse at discharge had a significant negative effect on the likelihood of a positive discharge. (The effect of serious abuse also had a negative effect but the numbers were too small to reach statistical significance. Paradoxically, though, moderate domestic abuse did not have a negative effect.)

**Does the presence of domestic abuse impede recovery from mental health problems?** The most direct measure of improvement on mental health status is a staff rating of the amount of improvement. The differences between those having domestic abuse at follow-up and those with no domestic abuse is close to statistically significant (p<0.09).

**Table 27: Follow-up domestic abuse by CGI symptom categories**

<table>
<thead>
<tr>
<th>Clinician judged improvement at discharge</th>
<th>Abuse at follow-up</th>
<th>No Abuse</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=138</td>
<td>N=368</td>
</tr>
<tr>
<td>1 - Ideal improvement</td>
<td>3.62</td>
<td>7.07</td>
</tr>
<tr>
<td>2 - Very considerable improvement</td>
<td>5.80</td>
<td>9.51</td>
</tr>
<tr>
<td>3 - Considerable improvement</td>
<td>6.52</td>
<td>12.50</td>
</tr>
<tr>
<td>4 - Moderate improvement</td>
<td>23.19</td>
<td>15.49</td>
</tr>
<tr>
<td>5 - Slight improvement</td>
<td>20.29</td>
<td>16.03</td>
</tr>
<tr>
<td>6 - Very slight improvement</td>
<td>7.97</td>
<td>8.42</td>
</tr>
<tr>
<td>7 - Status unchanged.</td>
<td>32.61</td>
<td>30.43</td>
</tr>
<tr>
<td>8 - Slightly worse</td>
<td>0.00</td>
<td>0.54</td>
</tr>
<tr>
<td>Total</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Another measure of clinical improvement is whether the GAF score of those with domestic abuse goes up or down at follow-up (with baseline GAF and other client characteristics controlled). Here we see the paradoxical effect discussed earlier. Those who had experienced minor or serious abuse at baseline (left graph in Figure 8) made improvements in predicted GAF by the time of the last rating. (Again, perhaps because they were highly motivated.) But those who had experienced any level of abuse at the follow-up rating had lower predicted GAF scores at follow-up. (Again, these are called predicted

---

changes because they “hold constant” a number of other predictors.) In short, domestic abuse, if current, does reduce recovery from mental health problems in this study.

Figure 8: Paradoxically persons with domestic abuse at admission have positive GAF change while those with domestic abuse at follow-up show negative change

Potential remedies for the issues surrounding domestic abuse

Since findings about domestic abuse differ somewhat in the monitoring study and the May 2015 Outcome Study, it is probably wise to mark this issue as one for follow up when the introduction of on-going outcome monitoring has produced a much higher number of cases.

When CalWORKs supported services began in 1999, much joint planning and communication took place among DPSS, mental health, substance abuse and domestic abuse agencies at the county level. That interaction has diminished very substantially, to the extent that silo services predominate.

1. As we have suggested before, the percentage of persons with domestic abuse situations would indicate that each clinic have at least one clinician who is a certified domestic violence advocate (having taken the 40 hour training).

2. An additional remedy would be to designate one clinic in each SPA that would have a track that integrates mental health and domestic abuse services. Persons with active safety issues would be referred to those clinics. This would involve an agreement with the county agency administering the CalWORKs domestic abuse services.

3. Although clinicians appear to be good at surfacing abuse issues, a formal screening instrument might increase the detection rate. DMH is considering the HITS instrument, but there are other possibilities. One or more instrument could be tried out in the next iteration of this study.

---

48 Research studies with the CalWORKs population have show that approximately 80% have past abuse, so almost twice the 42% reported at baseline here.
4. Because the Outcome Study of 2015 found different patterns through the use of phone interviews with persons who had left services, DMH may wish to ask CIBHS to conduct post-discharge phone interviews with a sample of participants who are experiencing domestic abuse at baseline.

D. METHODOLOGICAL ISSUES, FINDINGS, AND RECOMMENDATIONS

Structure of data collection when the system is fully implemented

Sampling as an option

The monitoring implementation study reported on here is sample-based: all participants in one quarter were sampled. Selecting a block of time as a sample is easier for providers than having to keep track of a random sample of participants.

However, any sample-based system is likely not to generate enough cases so that providers and SPAs can be reliably measured and compared. Which is to say that if DMH would be satisfied with an annual characterization of outcomes in the CalWORKs mental health system as a whole then sampling one quarter a year would be an approach that minimizes the burden on providers. One potential problem, as was evident in the monitoring implementation study reported here is that because the reporting is not routinized for every participant, it leads to tracking errors.

Recommendation: Include all CalWORKs mental health participants in the outcome measurement system.

Quarterly measurement

In this outcome monitoring trial we measured the same outcomes on each of five successive quarters (if clients were still in treatment): the baseline quarter and the next four quarters. We were interested in what could be gained by multiple measurements on the same persons as well as whether quarterly measures imposed too great a burden on providers.

We found several advantages and several disadvantages to quarterly measurement.

Advantages

• Statistical power is maximized by having several measurements on the same person. This was shown in the analysis above by the fact that the longitudinal analysis using baseline, quarters, and discharge revealed more statistically significant predictors of outcomes than just baseline and discharge alone.

• The amount of time required to do the quarterly updates was not large because of the steep drop off of the number being served. Table 28 below shows hours spent by clerical and clinical staff for each administration. Only about one quarter of the time was for quarterly forms.
Disadvantages

- Although we sent an email reminder for every client every quarter, it appears that tracking clients quarterly was very difficult for many providers—hence the large amounts of missing data at each quarter.

- Participants naturally leave the program over time. That means the number being served at each quarter is smaller. By the fourth quarter is was less than one hundred of the 659 we started with. This has two consequences. One is statistical significance is a function of effect size and of sample size. With small samples it is hard to obtain significant results in the separate quarters. Graphs have large error bars, for example (like Figure 2). The last couple of quarters in a year will be way too small to use in looking at provider variability.

Recommendation: Ultimately it seems to CIBHS that the improved statistical power provided by quarterly data is not worth the burden it imposes on providers.

Measurement cycle for the consumer form.

As noted in Appendix 2, there was very large attrition for the consumer form. At every stage the number of consumers who were persuaded to complete the form were far less than the number receiving service and quite a bit less than the number that staff rated. Getting a consumer evaluation filled out as participants leave the agency services is particularly problematic. All the problems of tracking clients in order to complete forms were apparently multiplied when it came to the coordination required for identifying a client in a timely way and asking for participation.

We did not capture refusals. In other studies using the same form, refusal has been quite low. But it might have been higher in this case because clients knew that they were identifiable so that staff could potentially see their responses (even though they were told they wouldn’t see them). However, the amount of attrition from baseline to quarter 1 and thereafter seems pretty much parallel for staff and consumer forms.

We did not capture the time required for the consumer forms, but the number of forms was quite small—much smaller than the staff forms—so the time would also have been considerably less.49

The most important point is that this study shows it is not feasible to collect information at discharge from most participants. This in turn means the consumer form is not capable of measuring outcomes.

Recommendations:

- Conduct the consumer survey one month after admission and six months later. This will provide a balance of evaluations when clients are new to the program and when they have significant experience with it.

49 In this study the time spent on the consumer forms was quite similar to that required in past years in order to meet the DPSS consumer satisfaction requirement.
Because such a high percentage of participants drop out or leave for negative reasons, it would be very useful to have follow-up data on client satisfaction and well-being after discharge. DMH might consider phone follow-up on a sample of clients perhaps three months after discharge. This could be done once a year and not necessarily tied to the timeframe of the surveys administered when clients were still in service.

Performance of the instruments currently used

All three staff scales, were sensitive to change over time but varied in other important traits.

Let us look first at how well the four scales used in the monitoring study predict positive or negative reasons for discharge and how well they predict whether participants work at discharge, as these are the two most important CalWORKs mental health variables.

Above in Figure 7 we showed a Receiver Operating Curve for positive vs. negative reasons for leaving service. We tried this same analysis—seeing what proportion of the area in the graph is “under the curve”—with the only predictor being respectively the four scales we used. The CGI by itself predicted an area under the curve of .79, the GAF predicted .75, the K6 participant scale .80 and the global improvement scale (GIS) predicted .87 (remember zerp predictive power is .50).

We tried the same exercise in prediction from baseline scores but this time using whether any work was recorded in the week before discharge. The GIS result was .72, the GAF was .62, the CGI .67 and the K6 was .68. So in general, psychiatric disability was less predictive of work than of leaving for positive or negative reasons. But the global improvement scale (GIS) was the best of the four on both tasks.

Looking at reliability, the improvement scale is quite different in that it does not have a baseline with which to compare it. In general, administering the same scale two or more times is likely to be more reliable than a single administration of an improvement scale.50 One reason for this is that in many cases the person doing the improvement scale rating at discharge has not followed the client from admission. This is a different and more difficult issue than that raised by pre post scales being rated by different persons—which is acceptable if inter-rater reliability is high.

However, neither the GAF or the Clinical Global Impressions scale has been subjected to extensive psychometric testing—which is very surprising consider the GAF is contained in DSM IV and the CGI is very widely used in clinical drug trials.

50 There are special scales called Rasch scales such that the difference between two ratings represents real change (without reliability uncertainty), but to our knowledge there are none that are useful for pre-post evaluation of clinical change.
The GAF has a broader purpose than the CGI. While including a rating on symptoms which parallels the CGI, it also measures functioning in family and work or school so is perhaps more suitable for CalWORKs participants. But it would be ideal to have separate symptom and function ratings (the GAF combines them).

The GAF and CGI each ended up designating 20% to 23% who made clear clinical improvement. However, the persons described as changing are in many cases not the same. This is apparent from the correlation between the GAF change score and the CGI change score: only 33%.

The correlation between the discharge rating of the GAF and the CGI is -47%; the correlation between the improvement scale and the GAF is 41% and between the improvement scale and the CGI is .48. While a correlation of 40% to 50% is fairly large it by no means indicates informational redundancy. The amount of variance explained by correlations of this magnitude is roughly 16% to 25%.

We have noted that the GAF mixes functioning and symptoms. DSM V drops the GAF and uses the World Health Organization DAS 2.0 to measure functioning alone. However, DMH has officially dropped the GAF but at this point has not adopted anything in its place.

**Recommendations:**

- Based solely on the performance in this study we would recommend keeping the CGI scale, and the CIS scale. We also recommend that for CalWORKs mental health, the GAF be included in the monitoring system or another scale that measures functioning not just symptoms. The three instruments do not correlate that highly with each other and are not consistent in predicting working or positive discharge, so there is little ground for eliminating one over another.

**Measuring participant status and satisfaction:** The current consumer form has several questions measuring satisfaction with services, including the amount of help participants feel they have received. It also includes the K6 symptom scale and the Sheehan Disability scales. As a result of using these forms in the outcome monitoring study and also as a consequence of changes in other DMH forms we reach the following recommendations:

**Recommendations:**

- If, as seems to be advisable, we limit the participant form to one month and then six months later, the value of the K6 and disability items comes into question. It would only make sense to keep them if we can collect a post version via a phone survey as was suggested earlier. So at this point we suggest dropping the K6 and disability scales. This decision could be revisited if DMH decides on a follow-up phone survey.

- The satisfaction items remain useful but should be added to. (See Appendix 4 for a mock-up of the suggested changes.)
Improving data collection and reducing missing data

Recommendations:

- We have recommended dropping the quarterly updates in favor of baseline and discharge (or six months for clients) administrations. We suggest having the baseline and discharge forms be in the admissions and discharge packets and perhaps even printed back to back with another required form.

- The only “tickler” reminders via email that would be needed—or possible—would be a reminder to get the first client form one month after admission and the six months client form six months after that (if no discharge form has been entered in the meantime).

- Questions on work and education hours in the staff forms need to be redone with fixed response categories so as to avoid missing or misconstrued responses.

Burden of the study on providers

Table 28 shows the total hours of clerical and clinical staff time used for the study as well as the average time in minutes used for each form. Note that the staff time would be about 25% greater if forms had been completed for all persons. The average number of minutes per case went down as clerical staff and clinicians got experience with the forms. According to Salaries.com the median wage for licensed clinical social workers is just about $70,000 a year or roughly $35 an hour. A secretary I makes $39,000 a year or roughly $19 an hour. Very approximately, then, the study cost $2,052 in clerical time and $7,980 in clinical time.

If our recommendation to employ staff forms for all participants at admit and discharge is adopted, the total time and costs are likely to go up.

Recommendation:

In our judgment the burden on providers does not outweigh the gain from implementing the system pre-post for all participants on an ongoing basis.
Table 28: Time spend on staff forms by clerical and clinical staff, over time

<table>
<thead>
<tr>
<th>Timeframe</th>
<th>Type of Time</th>
<th>N</th>
<th>Total Hours</th>
<th>Average Minutes per case</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong></td>
<td>Clerical</td>
<td>658</td>
<td>45.4</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>Clinical</td>
<td>658</td>
<td>102.7</td>
<td>9.4</td>
</tr>
<tr>
<td><strong>Quarter 1</strong></td>
<td>Clerical</td>
<td>300</td>
<td>16.5</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>Clinical</td>
<td>300</td>
<td>35.0</td>
<td>7.0</td>
</tr>
<tr>
<td><strong>Quarter 2</strong></td>
<td>Clerical</td>
<td>224</td>
<td>10.9</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>Clinical</td>
<td>224</td>
<td>22.5</td>
<td>6.0</td>
</tr>
<tr>
<td><strong>Quarter 3</strong></td>
<td>Clerical</td>
<td>85</td>
<td>4.5</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>Clinical</td>
<td>85</td>
<td>8.1</td>
<td>5.7</td>
</tr>
<tr>
<td><strong>Quarter 4</strong></td>
<td>Clerical</td>
<td>46</td>
<td>2.1</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>Clinical</td>
<td>46</td>
<td>4.5</td>
<td>5.9</td>
</tr>
<tr>
<td><strong>Discharge/End of Study</strong></td>
<td>Clerical</td>
<td>429</td>
<td>28.6</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>Clinical</td>
<td>500</td>
<td>55.3</td>
<td>6.6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>Clerical</td>
<td>1742</td>
<td>108.0</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>Clinical</td>
<td>1813</td>
<td>228.1</td>
<td>7.5</td>
</tr>
</tbody>
</table>

Reporting and making data useful for DMH and providers

- In this study we had to wait almost a year from study inception to have enough data to analyze. With a system that has continuous data collection, after the first year analyses can be completed on whatever schedule DMH finds useful.

- The data collected in the monitoring study and to be collected in the future offer considerable leeway for investigating issues of interest. This is illustrated in the discussion of policy questions in this report (an early version of which was given to providers in November of 2015). The information collected also allows feedback, so that if DMH initiates a policy change the effects of it can be measured down the road.

- We believe the system is flexible enough that there can be some change in what is collected over time as different needs surface.

- The most important factor in making data useful to providers is to have a sufficient sample size that they (and DMH) have confidence in the reliability of the trends shown in the data.

Recommendation:

DMH and CIBHS should work out a report-back schedule for the initial year and then a more frequent schedule for subsequent years.
APPENDICES
APPENDIX 1: Outcomes Monitoring Implementation Study 2014-2016: Data Security

DMH has a business associate agreement with CIBHS which includes CIBHS responsibility for data security. In this study, security derives from the fact that no Personal Health Information is stored on the external data base used in the study.

- CIBHS contracted with the data base provider Caspio to host the webforms and the data. Caspio restricts access by giving each "user" a unique password. Each provider is one "user" and must control the small number of persons authorized to enter data (or obtain reports). The "user" at each provider is able to enter data for their site, not view data from other providers.
- No identifying data or other Protected Health Information is stored on the database. Data are identified by an arbitrary provider number (not the one used by DMH) and an arbitrary client study identifier.
- Sites themselves link the arbitrary data base identifier with the DMH MIS identifier in order to keep the different forms together and be sure they are completed for each client admitted and at quarterly intervals. The forms that contain this link will be destroyed when the all data have been entered for a client. No one outside of the provider can link information on the data base to client PHI.
- From CIBHS, only Daniel Chandler, had access to the (anonymized) data for the purposes of analysis.

These procedures and the Caspio organization were reviewed and approved by the HIPAA compliance unit at DMH.

Note that since the study began Caspio has developed a HIPAA compliant data base and procedures which will be used in the full system implementation.
APPENDIX 2: Training and Implementation

Training

Training was conducted at the September 2014 Quarterly Providers’ meeting using a PowerPoint presentation by Dan Chandler, Ph.D. A second presentation at the December Quarterly Providers’ meeting reviewed problems that had come up and offered solutions.

In addition to live training, providers received a set of procedural guidelines with step-by-step instructions for how to generate the data required, and how to enter that data on the Caspio website.

DMH staff and Dr. Chandler performed trouble shooting for individual providers encountering difficulties.

Sampling frame

Several estimates of how many new admissions there would be for each provider in a three-month period were generated by using a “representative” base period made up of October 2013, January 2014 and April 2014. Although ultimately admissions and discharges must balance, providers vary considerably in the number of admissions per month and the number of discharges. Estimates varied from 1099, which was the maximum for each provider of admits/discharges extrapolated to three months, to 928, which was the mean of the three months multiplied by three, to 852, which was the median for the three months multiplied by three. In actual fact, the number of admissions during the study period of October 1–December 31, 2014 was 659. We do not believe any significant group of clients admitted to a program were not included in the study; however, there were 28 clients for whom staff completed discharge forms who had no baseline form.

Time frame

Although a small fraction of CalWORKs mental health participants stays in treatment for up to 18 months, for practical reasons the time frame was set for 15 months. That is, quarterly data from both staff and clients would be collected for five quarters after admission or until the client left services, whichever occurred first.

When implemented as an ongoing system of monitoring, this could be extended to capture all episodes and to try to determine how those staying a year or longer differ from others.

Data collection

Each provider was asked to designate one staff member to data-enter both client and staff forms using only an arbitrary provider ID (not known to other providers) and arbitrary client IDs. Data are entered on web forms requiring the email address of the data enterer (which must match what is on records) and a password set by the data enterer and unknown even to the study coordinator, Dr. Chandler.
Each quarter, starting 90 days from the date of admission, providers receive an email reminder that new quarterly forms (or a discharge form) are due. And at the end of the quarter a list of missing forms is generated and sent to providers and to their DMH program monitors. Tracking of the quarterly forms clearly poses difficulties for providers, despite the individual reminder emails for each client on the anniversary of the enrollment date. After quarter 2, for example, there were 229 quarterly forms that should have been entered and had not been. This necessitated follow up by DMH coordinators with most of the providers. Some of the problems appear to have been with the email reminder system; the problems were fixed in the spring of 2015. A number of providers expressed gratitude for the email reminders.

Keeping track of consumer forms causes an even greater problem for providers judging by the high number of consumer forms that were duplicates. At baseline, for example, there were 29 cases with one duplicate, 12 with three duplicates and 11 with four or five duplicates.

Data completion and attrition

A. Staff data

In theory there should be no missing data that is due to staff failing to enter online forms. Even if clients leave precipitously staff should be able to make the judgments required to complete the online forms. In fact, though, there were many missing forms.

a. There were 659 cases at baseline. For unknown reasons staff submitted discharge forms for 25 persons who never had a baseline form. However, they failed to submit a discharge or end of study form for 170 persons (26% of those with a baseline form). This is an unacceptable level of missing data.

b. Quarterly totals compared to Baseline, other quarters and End of Study:

- In each quarter there were a few persons for whom a form was completed but who had not had forms done in previous quarters. For example. In quarter 2, there were 14 persons with a form who did not have one in quarter 1. In sum there were 42 persons who got added to the study somehow after the enrollment period. (Or whose baseline forms were never submitted.)

Clearly keeping track of study participants caused major difficulties for many providers.

The difficulties in dealing with the quarterly updates must count against having quarterly measurements in addition to pre and post measurements when the system is implemented for routine administration. On the other hand, when the data collection shifts to using the DMH identifiers and all clients are participants it seems

---

51 Actually staff usually complete each form on a paper version and then it is entered on the online data entry site.
likely staff will have less trouble keeping track of who needs a particular form: all will, so keeping track of who is in the study is simplified.

B. Consumer data

At baseline, there should have been as many consumer forms as staff forms reporting an admission. In fact, there were forms completed by only 456 of 659 admissions. In quarter I (January through March), there were only 232 completed.

We do not have a way of tracking the reason for non-completion. One possibility is that a client could have refused. A second, noted above, is that providers found the logistics of getting the consumer satisfaction forms completed approximately 3-4 weeks after the client was admitted difficult to manage.

When consumers remain in treatment, then the quarterly consumer satisfaction forms should be collectible by staff with relatively little missing data. However, when participants leave a treatment program, they often do not give notice in advance. In fact, it is common for termination to arise from participant failure to attend services. Staff are requested to obtain a termination consumer form if possible from clients who have given notice. However, the data in Table 29 make it clear that most persons leave in the first quarter and do so without completing a consumer form at discharge.

Other problems with the consumer form included duplicate submissions and very late submissions (up to 5 months past the end of the quarter). The duplicates might have been due to finding forms in the chart and re-entering or entering data from later quarters in an earlier quarter.

And as with the staff forms, there were participants entered in quarter 1 and subsequent quarters who had not had a form entered in the baseline. We do not know if clients had refused in the baseline or it just failed to get done. Since there were over 200 consumers for whom a baseline form was not entered, there were clearly many who could have been “captured” in later quarters.
Table 29: Completed forms

<table>
<thead>
<tr>
<th>TIMEFRAME</th>
<th>STAFF</th>
<th>CONSUMER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>661</td>
<td>454</td>
</tr>
<tr>
<td>Quarter 1</td>
<td>300</td>
<td>232</td>
</tr>
<tr>
<td>Quarter 2</td>
<td>224</td>
<td>197</td>
</tr>
<tr>
<td>Quarter 3</td>
<td>85</td>
<td>106</td>
</tr>
<tr>
<td>Quarter 4</td>
<td>46</td>
<td>35</td>
</tr>
<tr>
<td>Discharge</td>
<td>498</td>
<td>55</td>
</tr>
<tr>
<td>End of Study</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>Discharge and End of Study combined</td>
<td>514</td>
<td>67</td>
</tr>
<tr>
<td>Matched Baseline and Discharge/End of Study(^{52})</td>
<td>488</td>
<td>NA</td>
</tr>
</tbody>
</table>

\(^{52}\) 28 of the unmatched cases were in the discharge cases but not in the baseline database and 173 were in the baseline but missing at Discharge or End of Study.
APPENDIX 3: Psychiatric Disability Rating Scales Used in the Monitoring

Discharge form (contains all staff rated scales)

Specify main reason for termination of CalWORKs mental health services at this point (Circle one number)

1. Client and clinician agreed client has met goals relevant to removing mental health barriers to employment

2. Client chooses to terminate even though all goals not attained; choice is not based on dissatisfaction with services

3. Client is working regularly or going to school full-time and time or place of work or school interfered with attendance at mental health services

4. Client states the reason for leaving is dissatisfaction with services or with progress being made (can include any reason for dissatisfaction including hours, location, therapist, treatment modality, or linguistic incompatibility)

5. Provider has decided to terminate services due to client non-compliance with treatment requirements or client not having a mental disorder that is a barrier to work

6. Client is transferring to SSI, SSDI or other income source (and/or became Target Population client)

7. Loss of CalWORKs eligibility due to not meeting program requirements, no longer having eligible children, having timed off welfare, being sanctioned, or declared exempt.

8. Other loss of CalWORKs or Medi-Cal eligibility including becoming eligible for other insurance through a job or marriage

9. Discontinued treatment due to practical difficulties including such things as transportation, child care, child health, illness, becoming homeless, several young children, or assuming responsibility for more children. (Select this response if practical difficulties were cited by client at termination visit or on a follow-up call or visit after client dropped out and other categories do not apply.)

10. Discontinued treatment due to behavioral health difficulties (include here substance abuse, domestic violence, jail time)

11. Only attended assessment or one visit before not returning

12. No shows for a period of over 30 days; failed to respond to calls/letters

13. Client moved out of program service area

14. A different primary reason for being terminated

Current Severity of Psychiatric Disability (Circle one number; Rate as of last visit with a licensed clinician) CLINICAL GLOBAL IMPRESSIONS SCALE

1. Normal – no psychiatric disorder

2. Borderline psychiatric disorder, subtle or suspected pathology

3. Mildly ill, clear symptoms, minimal distress or difficulty in social/occupational functioning

4. Moderately ill, overt symptoms, noticeable but modest functional impairment

5. Markedly ill, intrusive symptoms, distinctly impaired social/occupational function

6. Severely ill, disruptive pathology, behavior and function compromised, requires assistance from others
7 - Extremely ill, drastic interference in function, may need hospitalization
Current GAF (Write in current GAF score between 1 and 99) GLOBAL ASSESSMENT OF FUNCTION SCALE

91 - 100 No symptoms. Superior functioning in a wide range of activities, life's problems never seem to get out of hand, is sought out by others because of his or her many positive qualities.

81 - 90 Absent or minimal symptoms (e.g., mild anxiety before an exam), good functioning in all areas, interested and involved in a wide range of activities, socially effective, generally satisfied with life, no more than everyday problems or concerns.

71 - 80 If symptoms are present, they are transient and expectable reactions to psychosocial stressors (e.g., difficulty concentrating after family argument); no more than slight impairment in social, occupational, or school functioning (e.g., temporarily falling behind in schoolwork).

61 - 70 Some mild symptoms (e.g., depressed mood and mild insomnia) or some difficulty in social, occupational, or school functioning (e.g., occasional truancy, or theft within the household), but generally functioning pretty well, has some meaningful interpersonal relationships.

51 - 60 Moderate symptoms (e.g., flat affect and circumlocutory speech, occasional panic attacks) or moderate difficulty in social, occupational, or school functioning (e.g., few friends, conflicts with peers or co-workers).

41 - 50 Serious symptoms (e.g., suicidal ideation, severe obsessional rituals, frequent shoplifting) or any serious impairment in social, occupational, or school functioning (e.g., no friends, unable to keep a job, cannot work).

31 - 40 Some impairment in reality testing or communication (e.g., speech is at times illogical, obscure, or irrelevant) or major impairment in several areas, such as work or school, family relations, judgment, thinking, or mood (e.g., depressed adult avoids friends, neglects family, and is unable to work; child frequently beats up younger children, is defiant at home, and is failing at school).

21 - 30 Behavior is considerably influenced by delusions or hallucinations or serious impairment, in communication or judgment (e.g., sometimes incoherent, acts grossly inappropriately, suicidal preoccupation) or inability to function in almost all areas (e.g., stays in bed all day, no job, home, or friends)

11 - 20 Some danger of hurting self or others (e.g., suicide attempts without clear expectation of death; frequently violent; manic excitement) or occasionally fails to maintain minimal personal hygiene (e.g., smears feces) or gross impairment in communication (e.g., largely incoherent or mute).
1 - 10 Persistent danger of severely hurting self or others (e.g., recurrent violence) or persistent inability to maintain minimal personal hygiene or serious suicidal act with clear expectation of death.
0 Inadequate information

**Domestic Violence During Past Month (Circle one number)**
1 - No domestic abuse from current or past partners
2 - Minor abuse issues from current or past partners
3 - Moderate domestic abuse interfering with or making it difficult for client to get mental health treatment; pushing, slapping, threatening with a fist; following client; limiting access to income.
4 - Serious domestic abuse such as serious physical abuse or major threats
5 - Very serious past domestic abuse: client lives in a domestic violence shelter or lives in a secret location.

**Substance Abuse During Past Month (Circle one number)**
1 - Abstinent or Use without impairment: Client does not use alcohol or other drugs or use does not cause recurrent social, occupational, or other challenge
2 - In recovery, working with a substance abuse counselor or going to a self-help group regarding past abuse
3 - Abuse (DSM IV abuse diagnosis): Client uses alcohol or other drugs during this time interval and there is evidence of persistent or recurrent social, occupational, or other challenges related to use or evidence of recurrent dangerous use
4 - Dependence (DSM IV dependence diagnosis): Exhibits at least three of the following: greater amounts or intervals of use than intended, much of time used obtaining or using substance, frequent intoxication or withdrawal interferes with other activities, important activities given up because of drug use, continued use despite knowledge of substance-related problems, marked tolerance, characteristic withdrawal symptoms, drugs taken to relieve or avoid withdrawal symptoms

**Extent of clinical and functional change since Admission (CLINICAL GLOBAL IMPROVEMENT SCALE, MEASURED ONLY AT DISCHARGE)**
1 - Ideal improvement
2 - Very considerable improvement
3 - Considerable improvement
4 - Moderate improvement
5 - Slight improvement
6 - Very slight improvement
7 - Status unchanged.
8 - Minimally worse
9 - Much worse
10 - Very much worse

**Current Level of Care (Circle one number)**
1 - Level of care 1 (treatment 1-10 hours a week)
2 - Level of care 2 (treatment 1-7 hours a week)
3 - Level of care 3 (treatment 1-4 hours a week)
Paid Work Hours in Last Complete Week Before This Report (Enter a Number): ____

Current Education Units (Enter a Number): _____

IPS Status:
  1 – YES, was in IPS
  2 – NO, was not in IPS
  3 – IPS was planned, but had not enrolled

Enter Clinical Time In Minutes Used to Complete this Form: ____

Enter Clerical Time in Minutes Used to Complete and Data Enter this Form: ____

Describe Problems Completing Form, if any:

__________________________________________________________
Client form (contains the two client self-report scales, questions 1-10)

How are you doing?

During the past 30 days, how much of the time did you feel...

(PLEASE CIRCLE ONE RESPONSE)

ON THE RIGHT FOR EACH FEELING)

<table>
<thead>
<tr>
<th>Feeling</th>
<th>All of the time</th>
<th>Most of the time</th>
<th>Some of the time</th>
<th>A little of the time</th>
<th>None of the time</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. ...nervous?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4. ...hopeless.................</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>5. ...............restless or fidgety</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6. ....so depressed that nothing could cheer you up? ........</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>7. ....that everything was an effort?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>8. ....worthless?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

9. Please mark one circle to describe how much during the past week your mental health problems have disrupted your work or school or CalWORKs work-activities.

10. Please mark one circle to describe how much during the past week your mental health problems have disrupted your family life and home responsibilities.

How are we doing?
Please tell us how well our services are meeting your needs.

11. Overall, how satisfied are you with the services you received at this agency? *(Please circle one number.)*
   1. EXTREMELY SATISFIED
   2. VERY SATISFIED
   3. MODERATELY SATISFIED
   4. SOMEWHAT SATISFIED
   5. A LITTLE SATISFIED
   6. DISSATISFIED

12. Have you been treated with respect by all the staff at this program?
   1. I HAVE BEEN TREATED WITH RESPECT BY ALL STAFF FROM RECEPTIONISTS TO COUNSELORS
   2. I HAVE BEEN TREATED WITH RESPECT BY ALMOST ALL STAFF
   3. I HAVE BEEN TREATED WITH RESPECT BY MOST STAFF
   4. I HAVE BEEN TREATED WITH RESPECT BY ONLY SOME STAFF
   5. I DO NOT FEEL RESPECTED AT THIS PROGRAM

13. How likely are you to recommend this program to a friend with the same kinds of problems?
   1. EXTREMELY LIKELY
   2. VERY LIKELY
   3. QUITE LIKELY
   4. SOMEWHAT LIKELY
   5. A LITTLE LIKELY
   6. I WOULD NOT RECOMMEND THAT MY FRIEND COME TO THIS PROGRAM

14. Overall, how much have the services you have received from the agency helped you to improve your situation or deal with your problems? *(Please circle the number in front of the best answer. Only circle one number.)*
   1. HELPED A LOT
   2. HELPED SOME
   3. HELPED A LITTLE
   4. DID NOT HELP AT ALL
   5. MADE THINGS WORSE
Thank you very much! Please feel free to discuss your answers with your counselor if you wish.
APPENDIX 4: Mock-Up of Proposed Consumer Form Revision

A Chance to Help!

Please participate in this study of mental health programs in Los Angeles County.

Please answer these questions then return the form to the person who gave it to you.

Your participation is voluntary, but if you participate it will help us make services better.

What you write will also help your counselor or therapist, but if there is anything you do not want to share with her or him just leave that question blank.

Your answers will not affect the benefits or services you receive in any way.

In six months we will ask these questions again to see how your situation has changed.

Thank you very much!

1. Overall, how satisfied are you with the services you received at this agency? (Please circle one number.)
   1. VERY SATISFIED
   2. SOMEWHAT SATISFIED
   3. SOMEWHAT DISSATISFIED
   4. VERY DISSATISFIED
2. Have you been treated with respect by everyone at this program? (Please circle one number.)
   1. YES BY EVERYONE
   2. BY MOST PEOPLE
   3. BY SOME PEOPLE
   4. I HAVE NOT BEEN RESPECTED

3. Would you recommend this program to a friend with the same kinds of problems? (Please circle one number.)
   1. YES
   2. NO
   3. NOT SURE

4. Do you trust the person you have worked with most? (Please circle one number.)
   1. YES
   2. NO
   3. SOMEWHAT

5. Overall, did the services you received from the agency help you to improve your situation or deal with your problems? (Please circle the number in front of the best answer. Only circle one number.)
   1. HELPED A LOT
   2. HELPED SOME
   3. HELPED A LITTLE
   4. DID NOT HELP AT ALL
   5. MADE THINGS WORSE

6. Is this the first time you have received counseling or medications for an emotional or mental health problem or for your nerves? (Circle one.)
   1. NO
   2. YES
7. Please circle YES or NO for each question about your services.

YES  NO  Did you have to wait too long to be admitted to this program?
YES  NO  Have people you know pressured you to stop coming to this program?
YES  NO  Are services available in your primary language?
YES  NO  Are services available at times that are convenient for you?
YES  NO  Does your work schedule make it difficult to attend?
YES  NO  Do you currently have trouble with transportation to this program?
YES  NO  Do you currently have problems with child care while you are here?

8. Are you working for pay right now?
1. NO
2. YES

9. Have you done any of these things in the last three months? (Circle YES or NO for each.)

YES  NO  Written a resume
YES  NO  Applied for a job
YES  NO  Gone to job interview(s)
YES  NO  Had a GAIN employment/vocational assessment
YES  NO  Received work skills training such as learning to use a computer or answering the phone
YES  NO  Done volunteer work
YES  NO  Attended a school, college or GED program
YES  NO  Attended a vocational training program or trade school
10. How helpful to you have the people working at the *CalWORKs* program been? (Not this mental health program but the staff at DPSS.)
   1. THEY HAVE HELPED ME A LOT
   2. THEY HAVE HELPED SOME
   3. THEY HAVE HELPED A LITTLE
   4. THEY HAVE NOT HELPED AT ALL
   5. THEY HAVE MADE THINGS WORSE FOR ME

11. Have you ever been homeless on the street or in a shelter? (Please circle one number.)
   1. YES, I am homeless now
   2. YES, within the last six months
   3. YES, more than six months ago
   4. NO, I have never been homeless on the street or in a shelter

12. Is there someone in your life who is currently making you feel unsafe?
   1. YES
   2. NO

Please use the space below to tell us anything else you want about the services you are receiving here or what you think about *CalWORKs*.

Thank you very much for your help!