

# Statewide Treatment Outcome Assessment in Colorado: The Colorado Client Assessment Record (CCAR)

*Richard H. Ellis*  
*Nancy Z. Wilson*  
*F. Mark Foster*

**ABSTRACT:** Colorado has been measuring mental health treatment outcome statewide for the past four years. The data system for it is built around the Colorado Client Assessment Record (CCAR), a multi-dimensional checklist and Level-of-Functioning (LOF) measure. The CCAR is used by the Colorado Division of Mental Health for needs assessment, defining target groups, estimating bed need and doing cost-effectiveness studies. An underlying structure consisting of seven independent scales has recently been established using factor-analytic methods. A rater effect was also discovered and mathematically removed, revealing reliable and valid LOF information heretofore unknown. The improved CCAR will hasten implementation of Colorado's long-range plan for treatment outcome assessment.

## *INTRODUCTION AND BACKGROUND*

How effective are the state's mental health programs? Colorado's state lawmakers have long sought an answer to this important question. In developing its response, Colorado, like many other states, has come to rely on information obtained with a level-of-functioning (LOF) measure, using admission LOF ratings to help describe clients served by the mental health system, and changes in LOF to characterize the results of treatment in system programs. Arizona, Illinois, Michigan, South Dakota, Maryland, and possibly others are using LOF measures, as well as Colorado. The Global Assessment Scale (Spitzer, et al., 1970) and Carter-Newman LOF measures (Carter & Newman, 1975) are the instruments most often encountered. The model in which they are used generally employs admission-discharge and/or fixed interval clinician ratings with data collected on either an entire population or a carefully defined sample. Supplementary state-level outcome studies are most often done by sampling agencies, special programs or client groups. Occasionally, follow-up studies are done at a state level (Cox, et al., 1982), but as

noted in that report, good data are expensive and studies of this magnitude often go begging for support. Nevertheless, Colorado's legislators continue to pose their question in various ways and the history of their concern can be traced in Division of Mental Health (DMH) responses to it.

Colorado's budget process was becoming unwieldy in the early 1970s. Rather than accept "expert" testimony or try to do the necessary studies themselves, the legislative Joint Budget Committee began to hold the central offices of the executive branch more accountable for the information needed to make budget decisions. This led the Colorado Division of Mental Health to new or improved systems for obtaining data on staffing, budgeting and client characteristics. A minimum data set was developed as part of this effort. Admission data included a simple, four-dimension LOF measure. Client functioning on each dimension was measured on a four-point scale. Psychometric properties of the scales were questionable, however, and since no discharge information was collected with them, very little could be said about treatment outcome. Faced with this situation, a group of mental health program evaluators working under the auspices of DMH, undertook to improve the measurement of treatment outcome and, further, to develop a general scheme for obtaining, sharing and using outcome information statewide. The effort was to include study and selection of suitable measures, elaboration of the overall scheme, and development of a long-range plan for its implementation. This project is described elsewhere (Wilson, 1979) but will be summarized briefly here.

#### *THE PLAN FOR SELECTING A MEASURE—1976*

Late in 1976, members of the Evaluation Advisory Committee, a group comprised of program evaluators in the Colorado mental health system, formed the Treatment Outcome Task Force. Its mission was two-fold. First, it was to assemble a body of knowledge on treatment outcome and how it is measured. This effort was to include a survey of Colorado's legislators and mental health program managers to learn how they would use treatment outcome information. Secondly, it was to test the best instruments available in order to recommend one or two for statewide use. Later, the group was charged with articulating the knowledge it had gained into a scheme for collecting, disseminating and incorporating outcome information into system planning and management. In the first phase, over fifty instruments were studied using an approach similar to the one described in a recently published taxonomy of outcome measures (Ciarlo, 1981). Next, members of the task force participated in a panel discussion with legislative and administrative decision-makers. Topics discussed included: ways of measuring outcome, types of raters, reliability and validity issues, training, cost, and implementation. The decision-makers wanted change data, preferably derived from some sort of

“social functioning” or “total functioning” model, and from any source, client, clinician or significant other, as long as it was reliable and valid. Armed with this information, the Task Force tested several social or total functioning measures in the second phase of the project.

The main objective of the second phase was to determine the feasibility of having a statewide treatment outcome system. Practical concerns of rater cooperation, logistics and resource availability at the agencies and central office were foremost, for without rater cooperation, smooth and timely data flow, and sufficient resources, compliance would be minimal, making questions of reliability and validity moot. However, the measures selected were to provide valid assessment of client functioning and be sensitive to change.

The measures tested were: the Denver Community of Mental Health Questionnaire—DCMHQ (Ciarlo & Reihmann, 1975), the Community Adjustment Profile—CAPS (Sletten & Hedlund, 1974), the Fort Logan Evaluation Screen—FLMHC (Ellis, 1977), the Personal Role Skills Scale — PARS (Ellsworth, et al., 1967), the Global Assessment Scale—GAS and the Katz Adjustment Scale—KAS (Katz & Lyerly, 1963).

#### *FEASIBILITY OF STATEWIDE OUTCOME MEASUREMENT—1977*

After selecting the instruments, the next step was designing a test of them. The study was to include three rater types: client, clinician, and significant other; three settings: rural, urban and suburban; centers versus hospitals; and change interval: discharge or follow-up. Each instrument was to be tested at least twice. Prepost measures on a matched sample were to be used to assess change as measured by difference scores. Participating task force members coordinated data collection and analysis and the results of the three-month study were reported in 1977.

#### *Type of Rater*

Attrition was severe. It was greatest among client and significant other measures with refusal to consent accounting for about one-third of the attrition. This finding raises some questions about the applicability of these kinds of instruments for ongoing treatment outcome measurement. Attrition was least for clinician ratings and this rater type was recommended. It was also recommended that data be collected at admission and discharge on every client, as sampling would be likely to exclude special, small groups for which data needs could not be anticipated and would cause insurmountable logistics problems later.

### *Cost*

Cost became a major consideration in the final selection. Clinician ratings were the most economical. Here again, centers reporting inadequate resources had the lowest costs, but also had the lowest response rates and the poorest quality data.

### *Psychometric Properties*

Samples in the study were too small to permit serious analysis of psychometric properties. Therefore, recommendations were made on the basis of published information. There was very little published on the Fort Logan measure. However, some information was gained from this study. One measure of criterion validity is the ability to discriminate between hospital and center clients at admission. In this study, both the Fort Logan measure and the GAS were able to do this. The Global Assessment Scale did not discriminate hospital from center clients at discharge, however, while the Fort Logan Measure did. The other instruments tested were unable to do so in this study and with the exception of the Personal Distress Scale of the DCMHQ, were insensitive to change. Most of the other instruments had already been eliminated as not feasible by poor return rates for follow-ups from non-clinicians, costly interview requirements and the like, so psychometric properties were irrelevant. Given that the Fort Logan measure and the GAS performed almost equally well, the Fort Logan measure had the advantage of being multi-dimensional. For example, the differences in LOF between hospitals and centers were specific to the areas of Thinking, Personal and Interpersonal Behavior and Meeting Basic Needs.

### *Training*

Some measures, including the DCMHQ and the PARS required significant amounts of training initially and were costly to maintain. The Fort Logan measure required no training. A later study (McGuirk & Waters, 1979) showed inter-rater reliability could be improved with training and the Division undertook a "training of trainers" project which was moderately successful.

### *Implementation*

The comparative ease with which the pilot study was carried out suggested whatever logistics and implementation problems occurred could be resolved for any of the instruments tested. Some clinician resistance was encountered in the form of complaints and haphazard completion of forms, but direct in-

tervention in the form of face-to-face negotiations by the agency project coordinator was effective in reducing it.

### *Sensitivity to Change.*

Very little change was noted in client-generated measures. Poor response rates and little change were the hallmarks of significant-other measures. Change was noted on two of the six scales on the Fort Logan measure and on the GAS. A summary of instrument pros and cons is provided in Table 1.

### *The Bottom Line*

Not only did it appear statewide treatment outcome measurement was feasible, but one measure deserved to be tried state wide. The Fort Logan measure was chosen, as it was relatively quick and easy to complete, was a multi-dimensional measure, was able to differentiate between client populations, and was sensitive to change. The Personal Distress scale of the DCMHQ was selected to provide the client-generated data needed to corroborate the clinical ratings. Client satisfaction measures were federally mandated at that time at CMHCs, but were not aggregated at the state level.

## *A CLOSER LOOK AT COLORADO'S LOF MEASURE*

The original instrument, the Fort Logan Evaluation Screen, had an extensive research background, much of it unpublished. It was derived from several studies done in connection with a demonstration project at Fort Logan in the early '70s. Factor analyses of staff and patient treatment objectives selected from a menu of 125 choices were done (Potter, 1976; Ellis & Cameron, 1977). The domains that emerged were corroborated in most cases by client data and were relatively independent of one another. Two small five-point scales, one to assess overall LOF, the other to record expected LOF, accompanied the menu. Data from these scales were encouraging and they gave rise to the LOF scales that, combined with checklist items representing common problems in each domain, comprised the total instrument. That design was expanded into nine scales and more check items were added to reach the present form of the CCAR, shown in Figure 1, with one exception: the termination form no longer includes the checklist. The definition of each domain is shown in Table 2. The domains are intended to provide as broad a measure of functioning as possible, hence the concern with such things as substance use, medical problems, trouble with the law, etc. Definitions of check items and anchors for ten-point intervals of the scales were incorporated in a user's manual (Wilson, et al., 1980). A taxonomy of outcome measures (Ciarlo, 1981) did not include an "extended format" for the CCAR. One was recently completed

STATE OF COLORADO

DIVISION OF MENTAL HEALTH

CLIENT ADMISSION FORM: PART B

CARDS 1 & 2: (1-2) AGENCY NUMBER [ ] [ ] (3-11) CLIENT NUMBER [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] (12-17) DATE OF ADMISSION MO [ ] [ ] DAY [ ] [ ] YR [ ] [ ]

RECORD INFORMATION ABOUT THE CLIENT'S PRESENT PROBLEMS AND RECENT PAST THREE WEEKS FOR A REFERRAL TO FLMHC OR CSH. GIVE DETAILS AND ADDITIONAL INFORMATION ON FORM DMH - CMR-3.

INDICATE EACH PROBLEM STATUS BY PLACING 1 = YES 2 = NO 3 = DON'T KNOW 4 = N/A

<b>(CARD 1)</b>		<b>(CARD 2)</b>	
<b>I SOCIO-LEGAL</b>	<b>IV THINKING</b>	<b>VI PERSONAL BEHAVIOR</b>	<b>EMPLOYMENT (CONT.)</b>
19 UNABLE TO CARE FOR SELF, GRAVELY DISABLED	38 MEMORY	58 WITHDRAWN	19 UNABLE TO/WON'T WORK
20 DANGER TO SELF	40 INTELLECT CAPABILITY	60 INAPPROPRIATE/BIZARRE	20 FREQUENT/RECENT FIRING
21 DANGER TO OTHERS	41 REPEATED THOUGHTS	61 REPETITIVE BEHAVIOR	21 OTHER PROBLEM
22 DISREGARDS RULES/REGULATIONS	42 SUSPICIOUS/MISTRUSTFUL	62 OVERACTIVE	<b>ACADEMIC/TRAINING</b>
23 RUNAWAY BEHAVIOR	43 PARANOID/IDEAS OF REF.	63 SLEEPING PROBLEM	22 ACADEMIC ABSENTEEISM
24 DISHONEST (E.G., LYING)	44 HALLUCINATIONS	64 RESISTIVE	23 POOR ACHIEVEMENT
25 OFFENSES VS PROPERTY	45 GRANDIOSITY/DELUSIONS	65 USES/CONS OTHERS	24 BEHAVIOR PROBLEM
26 OFFENSE VS PERSONS	46 OTHER PROBLEM	66 BELLIGERENT	25 SUSPENSION/EXPULSION
27 HAS LEGAL PROBLEMS NOW	<b>V PERSONAL DISTRESS</b>	67 OTHER PROBLEM	26 OTHER PROBLEM
28 INTRA-FAMILY VIOLENCE	47 WORRY - ANXIETY	<b>VII INTERPERSONAL RELATIONS</b>	<b>MANAGEMENT PERSONAL AFFAIRS</b>
29 VICTIM OF PHYSICAL ABUSE	48 FEAR/PHOBIA	68 PRIMARY PARTNER	27 SELF-CARE MANAGEMENT
30 VICTIM OF SEXUAL ABUSE	49 DEPRESSED	69 OTHER IMMEDIATE FAMILY	28 BUDGET
31 OTHER PROBLEM	50 WORTHLESS	70 FRIENDS/OTHER	29 USE OF LEISURE TIME
<b>II SUBSTANCE USE</b>	51 NERVOUS/TENSE	71 SOCIAL SKILLS	30 OTHER PROBLEM
32 ALCOHOL	52 MARKED CHANGES IN MOOD	72 ESTABL'G REL'SHIPS	<b>IX MEETING BASIC NEEDS</b>
33 OTHER DRUG/SUBSTANCE	53 FRUSTRATED	73 MAINT'G REL'SHIPS	31 FOOD
<b>III MEDICAL/PHYSICAL ILLNESS/INJURY</b>	54 ANGRY	74 SEXUAL INTERACTION	32 CLOTHING
34 RECENT, ACUTE ILLNESS	55 GUILTY	75 OTHER PROBLEM	33 TRANSPORTATION
35 CHRONIC, RECURRENT ILLNESS/INJURY	56 LONELY	<b>VIII ROLE PERFORMANCE</b>	34 HOUSING
36 PERMANENT DISABILITY	57 BORED	<b>EMPLOYMENT</b>	35 FINANCES
37 NUTRITION/WEIGHT	58 OTHER PROBLEM	76 WORK ABSENTEEISM	36 OTHER PROBLEM
38 OTHER PROBLEM		77 POOR JOB PERFORMANCE	
		78-80 [ ] [ ] [ ] KEYPUNCH USE ONLY	

RATE THE CLIENT'S LEVEL OF FUNCTIONING IN EACH AREA. INDICATE THE RATING WITH A VERTICAL MARK ANYWHERE ON THE LINE

	ABOVE AVG. FUNCTIONING	AVERAGE FUNCTIONING	SLIGHT DYSFUNCTION	MODERATE DYSFUNCTION	SEVERE DYSFUNCTION	
	5	10	15	20	25	30
I SOCIO-LEGAL	-----	-----	-----	-----	-----	(37-38)
II SUBSTANCE USE	-----	-----	-----	-----	-----	(39-40)
III MEDICAL/PHYSICAL	-----	-----	-----	-----	-----	(41-42)
IV THINKING	-----	-----	-----	-----	-----	(43-44)
V PERSONAL DISTRESS	-----	-----	-----	-----	-----	(45-46)
VI PERSONAL BEHAVIOR	-----	-----	-----	-----	-----	(47-48)
VII INTERPERSONAL RELATIONS	-----	-----	-----	-----	-----	(49-50)
VIII ROLE PERFORMANCE	-----	-----	-----	-----	-----	(51-52)
IX MEETING BASIC NEEDS	-----	-----	-----	-----	-----	(53-54)

NAME OF REPORTER: \_\_\_\_\_ DEGREE: \_\_\_\_\_ TITLE: \_\_\_\_\_

CLIENT ADMISSION FORM PART B: PROBLEM SCREEN AND LEVEL OF FUNCTIONING RATING FIRST COPY - DMH/DATA PROCESSING (78-80) KEYPUNCH USE ONLY

PROBLEM SCREEN AND LEVEL OF FUNCTIONING RATING SECOND COPY - CHART [ ] [ ] [ ]

PROBLEM SCREEN AND LEVEL OF FUNCTIONING RATING THIRD COPY - AGENCY PROGRAM EVALUATION/LOCAL AGENCY COPY FORM DMH - CMR-2

Figure 1. The Colorado Client Assessment Record

TABLE 1

Type of Rated/Measure	Pros	Cons
<u>Client Rated</u> DCMHQ } FLMHC-P } ----- KAS-S	Measures have credibility Multi-dimensional measures provide differential results Resources required are mainly non-clinical Differentiated between hospital and CMHC clients ----- Resources required are mainly non-clinical	High attrition for both pre and post resulting in non-representative sample Attrition caused by need for consent - universal problem Attrition also caused by client's inability to understand task-agency specific problem ----- High attrition - problem as cited above Culture bound - not appropriate for many target groups and perhaps some agencies
<u>Staff Rated</u> FLMHC-S ----- GAS	Multi-dimensional measure which provides differential results Able to differentiate hospital and CMHC clients on both the pre and post measures Requires little training Low attrition on both pre and post Uses trained and skilled raters Has parallel client and community forms for validity check if desired ----- Able to differentiate between hospital and center clients on the pretest Low attrition on both pre and post measures Easy to administer	Requires clinical staff time to complete Time period varies for post - if client drops out of treatment, rating will have to be made retroactive to last visit ----- Uni-dimensional measure produces minimal information and may lead to bias Poor face validity of final level of functioning ratings
<u>Collateral Rating</u> FLMHC-C } CAPS } KAS-R }	Multi-dimensional measures Can be used as a validity check in special studies Resources required are mainly non-clinical	Monumental attrition problems Requires heavy investment of resources for data received Cannot be used for those clients without community collaterals - bias against certain agencies and target groups

by the senior author of this article and the CCAR compared favorably with the other measures on almost all dimensions.

### COLORADO'S STATEWIDE TREATMENT OUTCOME MODEL—1978

The CCAR was implemented by DMH in July of 1978. The process was monitored closely by the Task Force for another year, during which the full model and long-range plan were developed for statewide use of the data. The model has four parts:

1. a system to provide routine monitoring of treatment outcome at the state level;
2. a series of special studies to address specific issues related to treatment outcome;
3. a communications network among program evaluators and program managers, including a system of routine reports, solicitation of

Table 2

#### Definitions of CCAR scale domains

Domain	Definition
Socio-legal	Measures the extent and ease with which the person is able to maintain conduct within the limits prescribed by law, rules, and strong social mores.
Substance Use	Measures the extent to which a person's use of synthetic or natural substances is controlled and adaptive to general well-being and functioning.
Medical Physical	Measures the extent to which the person is free of or subject to illness, injury and/or disabling physical conditions.
Thinking	Measures the extent to which the person is capable of and actually uses clear, well-oriented and rational thought processes.
Personal Distress	Measures the extent to which the person's emotional life is well-moderated or out of control.
Personal Behavior	Measures the extent to which the person's daily behavior is appropriate, acceptable and understandable.
Interpersonal Relationships	Measures the adequacy with which the person establishes and maintains interpersonal relationships.
Role Performance	Measures the effectiveness with which the person manages the role most relevant to his or her contribution to society.
Meeting Basic Needs	Measures the adequacy with which the person is able to obtain such basic needs as food, shelter and transportation



managers' objectives for client change and notification of exceptions flagged by the monitoring system;

4. an ongoing effort to provide technical assistance to the agencies along with information and support for financial assistance requests there and at central office.

Routine monitoring is viewed as essential to the first three parts mentioned above. Monitoring could flag areas for more in-depth study, provide some of the information program managers needed and demonstrate the feasibility of the approach, thereby possibly attracting funding.

The long-range plan consisted of six goals:

1. begin implementing routine monitoring of target groups, e.g., adults, the elderly, the chronically mentally ill, gradually expanding until outcome information was available for all groups;
2. establish the need, if any, for other measures of outcome with selected groups, e.g. children and adolescents;
3. continue to explore the psychometric properties of the CCAR, maximizing them, where possible through error feedback, training or redesign of the forms;
4. utilize admission LOF data to help define need within the system and establish target groups;
5. integrate outcome with fiscal data in cost-effectiveness and cost-benefit studies;
6. disseminate admission severity and treatment outcome information to planners and to other interested parties.

#### *SOME PRACTICAL EXPERIENCE WITH CCAR DATA— 1979 TO PRESENT*

##### *Admission Severity*

Although it was initially intended as a system for measuring treatment outcome, it quickly became clear the CCAR contained useful information in the form of baseline admission LOF ratings. In 1979, Colorado legislators asked the Division to find a way to limit mental health dollars to those clients most in need of mental health services. CCAR data provided a means for identifying which clients were the most severely impaired at admission and this became the basis for determining eligibility for state mental health funds. A measure defined as the sum of the highest three CCAR scale ratings was selected as the measure of severity. Since each scale is a 50-point scale, scores on this variable ranged from 3 to 150. Center severity ratings were normally distributed and had a mean lower than that for hospital ratings which were also normally

distributed. Since both distributions did not completely overlap, a cutoff on the distribution for centers could be made at a point two standard deviations below the hospital mean, namely, at 78. Only admissions with scores above this cutoff qualified for reimbursement under the Division's system for performance contracting with the CMHCs. This cutoff excluded about 20 percent of the centers' admissions the first year, a percentage which has been steadily decreasing and now stands at about 10 percent.

### *Community Surveys*

Several centers in Colorado undertook needs assessment surveys, census-style randomized blocks designs that asked the resident interviewed to rate some member of the household using the scales of the CCAR (Edwards, 1979). The investigators reported means of 15 (Average Functioning) and standard deviations of 5 on most scales. Scale composites created from the CCAR checklist items associated with each LOF domain were used to increase instrument reliability for a survey of Denver's mental health needs (Shern, 1982).

### *Client Status Reports*

In 1980, the Division began a system of routine reports which tied units of service directly to outcome. Each quarter, a one percent sample of each center's admissions for one year earlier is listed using that center's own client identification numbers on a computer-generated form. The agency is asked to report the status of each case, open or terminated, units of service in each of four categories, inpatient, residential, partial and outpatient and some billing information. Admission and termination LOF on each case is then added to the file, thus making cost-effectiveness studies and studies of the relationship of outcome to units of service possible. So far, no relationship has been found between outcome and units of service, a finding also reported elsewhere (Cox, et al., 1982).

### *Open Cases Reports*

The Division also obtains data on all open cases in three categories: inpatient, residential, partial, and a ten percent sample of a fourth, outpatient, on one selected day each year. When combined with admission information, these data provide a description of all clients served. All admission form variables, thirteen checklist items and the nine LOF scales are included on the form. The most recent application of this data has been the development of a model for estimating the numbers of clients appropriately and inappropriately placed in different residential settings. With this model, it has been possible to estimate bed need in the state for various types of beds by showing which patients are now in inappropriate settings.

*Other Applications*

Checklist and LOF information is used on a day-to-day basis in many ways. A 1982 Colorado Supreme Court ruling was to have significant impact on children and adolescents in treatment in the state hospitals. The Division was able to obtain a stay of execution long enough to permit promulgation of emergency rules and regulations on the basis of CCAR data on the numbers of children affected.

Combinations of checklist items were used to assemble a definition of chronicity which allowed the Division to use its current data base to determine counts and locations of chronic patients in the system. CCAR data is also used in documenting budget issues, defining the characteristics of clients in target groups, and developing vocational programs, among others.

*SOME PROBLEMS WITH CCAR DATA AND THEIR SOLUTION—1982*

The Treatment Outcome Task Force had been aware of a problem with the Fort Logan measure, namely high interscale correlations. This problem was also evident in data from the CCAR. Despite having been developed from an empirical base designed to keep its scales independent, it appeared they were not. Two factor analyses seemed to indicate the scales could not be used independently (Winer, 1981; Stahler, 1982) and were only useful for assessing general psychopathology. No one knew if the instrument's underlying structure was the same at admission and termination. It was known that if change was to be measured by some aggregation of scales, there would be serious methodological problems. Even on the overall severity measure, the sum of the highest three scales, differences between the two state hospitals appeared for which there was no explanation. Comparison of LOFs for Colorado State Hospital, where ratings are done by the hospital staff, and the other state hospital, Fort Logan Mental Health Center, where ratings were done by CMHCs during admission prescreening, showed Fort Logan patients were consistently seven to nine points more impaired on most scales, a finding not supported by any other evidence. Unless some solution could be found, Colorado would be unable to use the multi-dimensional aspect of the CCAR LOF scales.

*RATER EFFECT*

One hypothesis was that CCAR data was possibly being distorted by a rater effect. It had been encountered once before in similar rating data (Foster, et al., 1972). Evidence for this effect included a high positive manifold in the interscale correlations, the problem with the two hospitals' data, and severity

levels for some centers that were not consistent with the type of client likely to come from their catchment areas. Controlling a possible rater effect meant identifying it first and factor analysis was the method chosen. The details of the study are reported elsewhere (Ellis & Foster, 1982), but will be explained briefly here.

### *The Sample*

A sample of clients ( $n = 27,707$ ) admitted to or discharged from the mental health system in 1980–81 was drawn from the Division's computer files. Every attempt was made to sample from all the types of clients who might be encountered by the system in one year. Duplication of clients was minimized by including only first admissions to centers and allowing only those few clients ( $n = 77$ ) who had been admitted to Colorado State Hospital's Forensic Division to represent all forensic clients. Only those clients who had been terminated were included to facilitate analysis of change in LOF from admission to termination. Only those records having no missing ratings at admission and termination were selected to facilitate all data analyses. Gross distortions, such as rating every scale the same, were also eliminated.

### *Factor Analysis*

A random sample of 5,300 records was selected from the 27,707 and the intercorrelations of their checklist items and LOF ratings were submitted to factor analysis. Ten factors were hypothesized, one for each scale, and one for the rater effect. Factor analysis in this study differed from other previously reported on CCAR data, in that the 77 checklist items were included instead of limiting the analysis to nine scales. The intent was to see if each checklist domain factored independently of the others and brought along its scale as a marker. Seven of the scales did so. No factor was found for Personal Behavior and two were found for Role Performance, one connected with job, the other with academic performance. The tenth factor was the rater effect, consisting of high positive loadings in all scales and no loadings on any of the checklist items. The factor analysis was repeated on the same clients' termination data with the same result except for minor differences in loadings of the scales on the rater effect factor. The entire analysis was successfully replicated using a new sample ( $n = 5,700$ ), drawn with replacement from the original 27,707.

### *Rater Study*

Another study was done using ratings ( $n = 863$ ) for which the raters ( $n = 11$ ) were known. Each rater was identified in a dummy variable and a correlation matrix of raters by LOF scales and checklist items was run. Some raters consistently rated high with respect to the other raters, others rated low, still

others rated certain scales high and others low in rater-specific patterns. Some raters appeared to have little or no bias in their ratings. Low bias was associated with better clinical skills or more clinical experience. A factor analysis of this matrix reproduced the seven scale factors and the rater effect. Raters either loaded on the rater effect factor or formed factors unique to the rater.

### *Community Surveys*

Data from the aforementioned community surveys incorporated the same checklist and LOF scales used on the CCAR. A fairly large sample ( $n = 2802$ ) was available and a study was needed to test whether the rater effect was idiosyncratic to a pathological population. Factor analytic studies confirmed the presence of a rater effect in community survey data, at the same time providing additional support for the same underlying structure found for the CCAR in replications on the population of mentally ill clients.

## LOF RESIDUAL SCORES

The rater effect was then mathematically removed from every rating in the state file to produce a set of residual scores.<sup>1</sup> Admission rater effect was removed from admission ratings, termination effect from termination ratings. Scale intercorrelations, which had been as high as the .60s for the scale that was subsequently dropped, fell to near zero or below in most cases as shown in Table 3.

The residuals were checked for criterion validity by selecting groups of clients by diagnosis and comparing their admission profiles. Profiles based on raw scores did not discriminate among diagnostic groups by visual inspection. Profiles based on residual scores made sense, however. The profile of scale scores for schizophrenia was marked by high disruption on Thinking, low on Personal Distress and Interpersonal Relationships, while in contrast, the profile for a depressed client was low on Thinking, high on Personal Distress and moderate on Interpersonal functioning. Organic brain syndrome was a Medical problem accompanied by high disruption in Thinking and inability to Meet Basic Needs. Diagnoses retained their characteristic profiles on termination, though shifted in the direction of less disruption, particularly on the scales that were markers for each diagnosis.

The finding of identifiable diagnostic groups in the LOF information was reliable. The profile for schizophrenics was examined for each of 23 centers and two state hospitals. Despite minor differences on single scales, the profiles were consistent and easily distinguished from other diagnostic groups. Removing the rater effect also caused the consistent difference in mean LOF between the two state hospitals to disappear in favor of a pattern more in line



exists in other instruments that measure psychopathology, but it may not be as easy to identify or control.

### *PROBLEMS REMAINING*

Despite a significant upgrading of the quality of its CCAR data, some problems with Colorado's treatment outcome model remain. Perhaps the most serious one, from a psychometric perspective, is the lack of information from clients or significant others. Whether the viewpoints of these informants and clinicians converge or diverge is not known and it is essential to have client-generated data to validate results obtained with the clinician-generated CCAR. A multi-state collaborative treatment outcome study was recently proposed (Newman, et al., 1982), in which the CCAR was to have been included. Despite its lack of success, the proposal drew together many of the most important issues in treatment outcome measurement. Hopefully, some may still be addressed by states or partnerships between states.

From a practical standpoint, the best possible method for rescaling residual scores continues to be a pressing problem. Comparing ratings for known "problem" groups with "no problem" groups, comparing normals from community surveys or validation with client data are a few of the possibilities here. Some means by which centers could correct their own data for bias, preferably manually, would also help. Not all centers have the means to factor analyze their own data or adjust ratings with large scale file manipulations on a computer. Even more to the point, some way of incorporating the CCAR approach within a more clinically relevant outcome measurement system would win clinician confidence and move the evaluator closer to what really happens in the treatment episode.

### *FUTURE DIRECTIONS*

With the residual scores from the CCAR, it will be possible to move toward outcome measurement that is closer to measuring the client's particular problems. Assessment of functioning in all domains can uncover specific areas of dysfunction; change over time can be assessed in all areas but particularly those showing dysfunction originally. This improvement in the quality of LOF information will allow further implementation of the overall model. It will be possible to:

1. Identify clusters of clients with the same profile, i.e., a client typology. It is anticipated this would converge on diagnosis, but would be somewhat broader. DSM-III is not intended for this purpose and some system of classifying mental health system clients is needed.

2. Proceed with linking data on treatment effectiveness with cost and other system variables. One way would be to correlate effectiveness measures, such as residualized gain, with variables known to drive cost, e.g., number of inpatient days.
3. Relate LOF to the social indicators approach used in Colorado's needs assessment model. This approach implies a relationship between certain characteristics of the general population or of special populations which acknowledge persons in them are at risk of acute or chronic mental illness. Community surveys done in Colorado used CCAR LOF measures and these can be matched with those of mental health center clients from the same communities to provide another way of estimating prevalence of mental illness in the community.
4. LOF can be used in both long- and short-range staffing models for state hospitals. One short-range model being tried now uses LOF to help determine nursing requirements for the daily mix of patients on an inpatient ward. Nurses are then assigned from a central pool to the units where they are most needed.
6. New dimensions can be added to the monitoring process. Some preliminary work has been done with a variable labelled 'expected LOF.' Clinicians are able to state the anticipated or desired LOF for a client as the result of the client's current functioning. Fort Logan data on this variable was used to obtain a profile of potential "chronicity" among current patients as part of justifying a hospital request for more inpatient beds.

### *SUMMARY AND CONCLUSIONS*

Colorado anticipated the need for treatment outcome information several years ago. The Treatment Outcome Task Force was a highly successful mechanism for defining the state's responsibilities, selecting an instrument and generating a plan. The plan is not fully implemented yet, but it continues to guide decisions in MIS development. The Task Force received an unusually clear message from legislators, budget analysts, and administrators: some measure of change in LOF was preferred, as it would enable them to make direct comparisons among programs. In fact, they have not used the information in this way, but it has entered the process in many other ways. The finding of a rater effect is not uncommon in scaled data. It is uncommon, however, to be able to successfully control it. Residual scores offer a unique opportunity to explore the relationships between this and other system variables. As elsewhere, cost-effectiveness studies are of particular concern in Colorado.

The ability to use multi-dimensional LOF data permits a shift in focus from measuring general psychopathology to profiles of pathological conditions.



This is analogous to the shift in education research away from interest in "general intelligence" toward specific kinds of intelligence. It is perhaps more important to know the areas in which someone may be showing the effects of being mentally ill than to know simply that they are "sicker" than someone else.

Colorado has opted for a model based on continuous monitoring at the state level with special studies as needed and strong support of agency efforts to improve their own program evaluations. When the question is asked, "Yes, but are mental health services effective?" Colorado is fortunate to be able to respond, "Yes, we have the data to show that they are."

### REFERENCE NOTE

1. Regression techniques were used to statistically remove the general rater effect from individual ratings. A technical paper describing research on the rater effect and how to remove it is in preparation. A working paper may be obtained from the Colorado Division of Mental Health.

### REFERENCES

- Carter, D.E. & Newman, F.L. *A client-oriented system of mental health service delivery and program management: A workbook and guide*. Philadelphia: Eastern Pennsylvania Psychiatric Institute, 1975.
- Ciarlo, J.A., Edwards, D.W., Kiresuk, T.J., Newman, F.L. and Brown, T.R. *The assessment of client/patient outcome techniques for use in mental health programs*. Final Report, National Institute of Mental Health, Grant No. 278-80-0005(DB), 1981.
- Ciarlo, J.A. & Reihman, J. The Denver Community Mental Health Questionnaire: Development of a multi-dimensional program evaluation instrument. In Coursey, R. Spector, G. Murrell, S., and Hunt, B. (Eds.) *Program evaluation for mental health: methods, strategies, and participants*. New York, Grune & Stratton, 1977, pp. 131-167
- Cox, G.B., Brown, T.R., Peterson, P.D. & Rowe, M.M. A report on a statewide community mental health center outcome study. *Community Mental Health Journal*, 1982, 18, 135-158.
- Edwards, M. *Community mental health needs assessment of Aurora, Colorado*. Technical Report, Aurora Community Mental Health Center, Aurora, Colorado, 1980.
- Ellis, R.H. *A measure of improvement in patients treated in an inpatient psychiatric hospital*. Technical report, Fort Logan Mental Health Center, Denver, Colorado, 1977.
- Ellis, R.H. & Cameron, Carole *Factor analysis of the Fort Logan evaluation screen*. Technical Report, Fort Logan Mental Health Center, Denver, Colorado, 1977.
- Ellis, R.H. & Foster, F.M. *Change in level of functioning as measured by the Colorado client assessment record*. Treatment Outcome Series No.2, Division of Mental Health, Department of Institutions, State of Colorado, 1982.
- Ellsworth, R.B., Foster, L., Childers, B., Arthur, G., and Kroeker, D. Hospital and community adjustment as perceived by psychiatric patients, their families and staff. *Journal of Consulting and Clinical Psychology*, 1967, 32(5) Part 2, Monograph Supplement.
- McGuirk, F. & Waters, J. *Preliminary survey for the development of treatment outcome training*. Technical report, Division of Mental Health, Department of Institutions, State of Colorado, 1979.
- Foster, F.M., Horn, J.L., and Wanberg, K.W. Dimensions of treatment outcome: A factor-analytic study of alcoholics' responses to a followup questionnaire. *Quarterly Journal of Studies of Alcoholism*, 1972, 33, 1079-1098.
- Katz, M.M. & Lyster, S.B. Methods for measuring adjustment and social behavior in the community: I. Rationale, description, discriminative validity and scale development, *Psychological Reports*, 1963, 13, 503-535, Monograph Supplement 4-VI3.

- Potter, A. *Goal attainment approach to program evaluation in mental health*, unpublished doctoral dissertation, University of Colorado, Denver, Colorado, 1976.
- Sherm, D.L. *The Denver health & hospitals mental health program community needs assessment project quality assurance study*. Paper presented at 1981 Annual Meeting of the Rocky Mountains Psychological Association, Denver, Colorado, 1981.
- Sletten, I.W. & Hedlund, J.L. The Missouri automated standard system of psychiatry; current status, special problems, and future plans, in Crawford, J.L., Morgan, D.W., and Gianturco, D.T. *Progress in mental health systems: computer applications*, New York, Ballinger Publishing Company, 1978, pp. 285-297.
- Spitzer, R.L., Endicott, T.R., Fleiss, J.L. & Cohen, J. The Psychiatric Status Schedule, *Archives of General Psychiatry*, 1970, 23, 41-45.
- Stahler, G.J. *An assessment of therapist rating bias and the Hawthorne effect in program evaluation*. Unpublished doctoral dissertation, Temple University, 1982.
- Wilson, N.W. *Treatment outcome: A feasibility pilot study*. Unpublished report, Division of Mental Health, Department of Institutions, State of Colorado, 1977.
- Wilson, N.W.(ed.) *Colorado client assessment record (CCAR): User's guide*. Division of Mental Health Department of Institutions, State of Colorado, 1980.
- Winer, D. *Assessing the relationship of patient and therapist characteristics to psychotherapeutic success as derived from the Colorado client assessment record*. Unpublished doctoral dissertation, University of Northern Colorado, 1981.