WORKING PAPER

Veterans Health Administration Mental Health Program Evaluation Technical Manual

MARCELA HOROVITZ-LENNON,
KATHERINE E. WATKINS, HAROLD ALAN PINCUS,
LISA R. SHUGARMAN, BRAD SMITH,
TERYN MATTOX, THOMAS E. MANNLE, JR.

RAND HEALTH

SUMMARY

Questions are being raised nationally about access to and quality of mental health care, both within the Veterans Health Administration (VHA) and in the public and private sectors more broadly. Numerous studies have documented the discrepancies between mental health care that is known to be effective and mental health care that is actually delivered.[ii] These gaps are important because mental health conditions are a leading cause of disability and death and have serious economic, social, and personal consequences.[iii] Concurrently, U.S. policy makers and medical professionals are increasingly recognize that quality mental health care can lead to better, healthier lives for those with mental illness, and that performance measurement plays a key role in improving health care delivery and, ultimately, patient outcomes.

In 2006, the U.S. Department of Veterans Affairs (VA) funded an independent study to evaluate the quality of mental health services provided to veterans. This study is mandated by the Government Results and Performance Act of 1993, which requires federal agencies to independently evaluate high-volume, high-cost programs, and Title 38 of the U.S. Code, which regulates Veteran's Benefits. It represents one of the largest and most comprehensive evaluations of mental health services ever conducted. The evaluation focuses on five high-volume, high-cost mental health diagnoses that have the potential to greatly impair quality of life for veterans:
- Schizophrenia
- Bipolar disorder
- Post-traumatic stress disorder (PTSD)
- Major depressive disorder
- Substance use disorder.

This evaluation of the VHA mental health services is designed to present new information about how well VA is translating the promise of improved mental health care into better, healthier lives for veterans. In particular, the study team will examine whether specific gaps in services identified as targeted areas of improvement in the MHSP have been reduced by the implementation of the plan, and whether investments and/or other enhancements in VA mental health and substance use treatment services under the plan have had a positive impact on capacity, utilization, staffing, and individual users over the study period.

In order to develop and select measures that would be viewed as meaningful and useful in addition to valid, the team's work was guided by policy documents that identified the aims and characteristics of high quality health care.   The Institute of Medicine's quality of care paradigm was used explicitly to categorize all potential measures and to ensure indicators covered all six domains of effectiveness, safety, efficiency, timeliness, patient-centeredness and equity. [ivv] The VHA Mental Health Strategic Plan was modeled after the Report by the President's New Freedom Commission on Mental Health; both identified the consumer and family as the drivers of mental health care, focusing attention on the concept of recovery and on the elimination of disparities in the availability and quality of mental health services. Together these documents provided the social and political context for development and selection of the measures herein. Below we have documented the methodology employed in the development of mental health indicators.

Indicator development process

1. Conduct a Systematic Search for Previously Identified, Grade I Performance Indicators.

We conducted a systematic review of the literature including studies, technical reports, reviews, electronic databases, etc., manual review of relevant bibliographies, and outreach to

experts and industry representatives to identify an exhaustive pool of relevant performance indicators that were either in the public domain or were being prepared for near-term dissemination. All relevant measures were retrieved and the team reviewed the methodology used in their design to assess their quality. We abstracted each performance indicator, noting its data source, the disorder to which it applied the strength of the evidence for the process measured by the indicator, and IOM domain.

2. Identify recommendations with empirical support that are not covered by the existing measures, and create new performance indicators to address these gaps.

We reviewed VA and APA Clinical Practice Guidelines for the 5 disorders included in the program evaluation (the VA CPG for psychoses includes recommendations for both schizophrenia and bipolar disorder), and listed all individual recommendation statements. Multi-part recommendations were separated into individual parts and duplicative recommendations were deleted. We defined key terms, examined the recommendations for inconsistency or ambiguity, and produced a list of explicit, unambiguous measures that had empirical support for the process-outcome link. Where discrepancies existed between the APA and VA guidelines the team consulted outside experts and discussed the discrepancy until consensus was reached.

3. Select measures for further technical specification.

Because of the large number of candidate measures, we engaged in a systematic selection process. First, we identified whether the data needed to populate the indicators existed in the necessary form in either the administrative or in the medical record, and recommendations that could not be defined operationally because of lack of data were eliminated. Next, the research team reviewed the measures for meaningfulness and feasibility, and described the measures' predictive validity through an evaluation of the strength of the process-outcome link. A subset of measures was selected to be reviewed by external clinical experts who further pruned them on the basis of clinical significance. All measures were reviewed with a VA clinical advisory group, whose members were selected for their clinical expertise and familiarity with the subject matter. The advisory group evaluated recommendations for validity and feasibility, and usefulness for VA's operational management and strategic leadership. Lastly, VA and VHA leadership rated the indicators on their importance to the VHA and contribution to presenting a comprehensive quality profile. As a result of this process, we identified a core set of measures that were valid, feasible, and a VA priority. Most of them described processes that were identified with acute treatment.

4. Generate a new set of measures pertaining to the psychosocial aspects of care.

Because the process used above required measures to have an empirical basis of support, the domains of patient-centeredness and recovery were neglected. Although not evidence-based or guideline-supported, both domains are endorsed by the Institute of Medicine and the VA Mental Health Strategic Plan as critical to quality. We therefore used a collaborative process between the research team and the VA clinical advisory group to identify key constructs pertaining to patient-centeredness and recovery. Among the many possible constructs, we chose to focus on the psychosocial aspects of care such as attention to social supports, housing and employment. Indicator development involved recruiting experts and engaging them in the process of identifying a core set of cross-cutting psychosocial indicators. Because of the difficulty evaluating the predictive validity of the psychosocial aspects of care, they will be used descriptively.

5. Develop technical specifications for finalized indicators and categorize their strength of evidence

We generated detailed technical specifications for all finalized performance indicators with respect to VHA administrative data and electronic medical records, and identified data sources that efficiently provided the information necessary to populate the indicators. Each indicator contained an indicator statement and executive summary describing the source(s) of the specifications and clinical rationale for the selected indicator. We also included the indicator grade, which reflected the strength of the process-outcome link, and whether the indicator would be used as a benchmark or descriptively. We created numerators and denominators for each indicator based on the data that would be available, and defined the population to which the indicator applied. For example, if the indicator applied only to people in a new treatment episode, we defined the term 'new treatment episode'. All clinical and measurement terms were defined operationally, and we summarized anticipated data collection problems and other feasibility issues. These included any problems that we could foresee prior to starting abstraction, such as data elements that might be time-consuming to collect or which required a judgement to be made by the abstractor. For complex processes of care with multiple components of varying clinical or outcome relevance (e.g., delivery of CBT/SST or assessment of mental status), we sought expert input to select and operationalize critical components. Technical specifications were reviewed by both external clinical experts and the VA Advisory group in order to ensure that specifications were both feasible given the data available, and meaningful to this particular population.

We categorized indicators according to the strength of the process-outcome link using the grading system developed by the AHRQ's US Preventive Services Task Force.[vi] Grade I measures are those where the link between process and outcome has been established through randomized clinical trials, grade II measures are supported by well-designed, non-randomized trials, and grade III measures are supported by expert opinion. A caveat to drawing conclusions from this grading system is that sometimes the outcomes literature may not be specific enough about the ingredients of the intervention that are critical to its efficacy/effectiveness. For example, although randomized controlled trials have established the value of psychotherapy in the treatment of several disorders, not enough evidence exists to ascertain the minimum "dose" (or number of sessions) and duration required for the outcome advantage to emerge. We also note that the grading does not reflect translational validity, or the certainty that the technical specifications accurately reflect the process of care they are trying to capture.

6. Determine data abstraction elements and sequence of abstraction

Starting with the technical specifications developed above, we described the data abstraction elements and abstraction sequence for each indicator. Since many indicators required overlapping information, we removed redundancy and grouped questions for efficiency. For example, all questions about medications were placed together,, since the medications prescribed to a veteran are found in a single section of the record. We created abstraction forms for each diagnosis.

7. Pilot test indicators for translational validity and performance

Clinical nurse abstractors piloted each indicator for timing and performance using pencil and paper and modifications were made in order to keep data collection time to a minimum. We

found that some data elements were not found in the part of the medical record to which we had access, and, after review with the clinical advisory group, deleted these indicators. After the initial paper and pencil pilot test , an electronic abstraction form was created and a second pilot test was performed to make sure that the questions flowed correctly and that there were no programming errors.

Discussion

In this report we present a comprehensive set of indicators for evaluating the performance of mental health care systems with two different data sources, administrative and medical records. One of the greatest difficulties in evaluating mental health care is obtaining meaningful data to measure the key elements of the system. In order to evaluate the structure of care, we developed indicators that used a combination of both data sources available, while recognizing that both sources of data, either singly or in combination, have inherent strengths and weaknesses.

The main strength of using administrative data is their availability and comprehensive enumeration of the study population.[vii] Moreover, the databases were relatively large, enabling the study team to analyze population subgroups and specific geographic areas separately, which was particularly useful, since most problems related to access and availability are not uniform across populations or within areas. In many cases, however, items were missing or the accuracy of the information provided could not be guaranteed. This is not uncommon when data are collected and used for different purposes. Other studies also support the use of administrative data combined with chart review to assess performance.[viii]

While the structure-process-outcomes evaluation model presented herein holds promise for advancing the science of mental health care quality improvement both within and outside the VHA, a few final caveats are in order.

First, in any health care system, the progression from evidence-based practice guidelines to performance indicators to improved patient outcomes is fraught with complexity. Great care must be taken to measure precisely what is intended to be measured through effective and efficient documentation so that the burden of measurement does not outpace quality care provision. In addition, continued awareness of the complicated linkages between evidence-based practice and individual patient preferences and outcomes is essential. As recent studies amply demonstrate, even the most basic of evidence-based practice improvements can result in different outcomes for different patients and for different reasons. Attention must also be paid to ensuring that quality improvement becomes a part of the fabric of care at both the organizational and individual levels, through resource investment, staff training, etc.

Second, not all mental health care systems look or operate like the VHA mental health care system. Public and private sector mental health care functions largely as a cottage industry, with the majority of psychiatrists practicing in solo or two-physician practices; information technology is less well developed; there are few centralized administrative databases; and there is no single entity or organization responsible for implementing and monitoring quality improvement strategies. While these differences must be recognized and addressed in the context of ongoing quality improvement, the same high quality standards should nevertheless apply.

Third, to what extent this model can be adapted for use in other systems and in other contexts is not clear. It is possible that certain components of the model will be more suitable for mental health quality improvement efforts at the national or state levels or in large systems (e.g., managed care networks), while others will work well in more localized contexts (e.g., community mental health centers).

VA has undertaken the most extensive, systematic, and rigorous evaluation of the mental health care delivery ever conducted. Although this quality improvement effort is still in its early stages, and much remains to be learned, the framework, methodology, and preliminary results offer a fertile ground upon which other stakeholders in the mental health field can continue to build and expand both in the near- and longer-term.