“Usual Care” for Major Depression in the 1990s: Characteristics and Expert-Estimated Outcomes

Marcela Horvitz-Lennon, M.D., M.P.H.
Sharon-Lise T. Normand, Ph.D.
Richard G. Frank, Ph.D.
Howard H. Goldman, M.D., Ph.D.

Objective: Motivated by the public health importance of major depression and calls to study the effectiveness of treatments frequently used in routine practice, the authors characterized the usual outpatient care of acute-phase major depression in a privately insured population as well as the effectiveness of this care.

Method: Upon identifying frequent antidepressant interventions from a large administrative database spanning the years 1991–1996, and after completing an exhaustive review of published efficacy evidence, the authors undertook a two-stage elicitation process. A panel of 10 experts estimated probabilities of symptom-based outcomes for 120 combinations of treatment, provider type, and patient characteristics seen in 9,054 depressive episodes. The 10 most frequent treatments, utilized in 90% of the episodes, are reported here.

Results: A short course of psychotherapy delivered by specialists to healthy premenopausal women was the most frequent pattern of care. Only two prevalent treatments had adequate evidence. On the basis of expert-estimated rates of remission (Hamilton Depression Rating Scale score $<8$), response (score $<13$), and no change (score $\geq18$), three of the 10 most frequent treatments were highly effective, two were moderately effective, and five were minimally effective.

Conclusions: Although experts gave high marks to three commonly used treatments, the expert-estimated effectiveness for this particular system of care was quite modest. More research is needed on the outcomes associated with specific usual care practices as well as the contributors to ineffective care.

Furthering the knowledge base on the characteristics and outcomes associated with the “usual care” of psychiatric disorders is a top priority for researchers and policy makers alike (1–3). Because of its high prevalence, associated disability, and societal burden (4, 5), there is much to be gained by understanding and improving the routine care of major depressive disorder.

Over the past several decades, a large number of randomized controlled trials have substantiated an incontrovertible advantage of standardized antidepressant interventions over placebo in the management of acute-phase major depressive disorder (6, 7). Although the routine outpatient care of depressed adults has been the subject of multiple naturalistic studies (e.g., references 8–11), we only found 13 randomized controlled trials (e.g., references 12–19) that had investigated outcomes in usual care conditions. None of these randomized controlled trials had aimed at fully characterizing usual care, and only two (16, 17) had assessed outcomes associated with specific interventions. One of these trials (17) demonstrated an outcome advantage for nortriptyline when delivered in an experimental versus a usual care manner, even for the usual care subjects treated according to guidelines. Although this is, to our knowledge, the only study that has documented such an outcome differential, or “efficacy-effectiveness gap,” this finding is consistent with a common belief that such a gap exists. Evidence substantiating inadequacies in the management of major depressive disorder has continued to accumulate (20) over a time period marked by the introduction of safer and easier to administer medications and quality improvement efforts (6, 21).

Because policy and resource allocation debates place a high premium on the relationship between costs and outcomes of treatments provided in real-world conditions, we set out to develop a method to evaluate this relationship for acute-phase depression in the outpatient setting. To this end, we analyzed claims data from a large insured population, a novel methodology inasmuch as administrative data has not been extensively used to shed light on both the quality and cost of treatments for depression in a self-contained system of care. After identifying all outpatient episodes of major depression over a 6-year period, we classified the patterns of care and elicited expert opinion on the likely outcomes of the most frequent treatments. On the basis of these estimates, we calculated a measure of treatment effectiveness for this particular system of care (i.e., system effectiveness). Our finding of an improvement in the ratio of treatment costs to outcomes during the 1990s has already been reported in policy and health economics journals (22–24).

Motivated by calls to study the performance of usual mental health care, this study reports on the routine man-
agement of adults diagnosed with major depressive disorder in a privately insured population—including the extent to which such care is supported by published evidence and its effectiveness as ascertained by a group of experts.

Method

Identifying Frequent Patterns of Care

Claims data assembled by MEDSTAT for the period 1991–1996 from four large self-insured U.S. firms (representing 426,000 employees and dependents) with generous mental health benefits by industry standards were examined to identify outpatient episodes of acute-phase major depressive disorder in adults ages 18–64. The operational definition of a first or recurrent depressive episode required a course of care of up to 16 weeks prompted by an ICD-9 major depressive disorder diagnosis and preceded by a period of 8 or more weeks without depression care. Upon identifying 13,098 such episodes, we analyzed the demographic and clinical information contained in the claims and classified episodes according to the observed treatment, type of provider (primary or specialty care), and patient characteristics.

Treatments consisted of an intervention defined according to current procedural codes (Physician’s Current Procedural Terminology, 4th ed. [CPT-4]) and a time descriptor (e.g., tricyclic antidepressant for more than 30 days). Interventions included medications (antidepressants, lithium, and antipsychotics), psychotherapy, combined antidepressant drugs and psychotherapy, brief office visits (i.e., clinical management), unspecified mental health interventions (namely CPT-4 unspecified “focused” visits), and medical interventions only (e.g., diagnostic tests). Drugs were grouped according to known conventions (e.g., selective serotonin reuptake inhibitors [SSRIs]), except for trazodone, amoxapine, and bupropion, which were bundled as “other antidepressants.” Duration of drug treatments was calculated by using pharmacy claims, which provide both number of days and dosages.

Patient characteristics were demographic and clinical variables associated with either the choice of or response to treatment: gender, age (which for women was dichotomized to differentiate by menopausal status: 18–49 years versus 50–64 years), and medical and substance abuse comorbidity, either concurrent or within the previous year. Medical comorbidity was selected on the basis of its impact on treatment choice or prognostic relevance (e.g., ICD-9 disorders of the circulatory system).

After cataloguing all episodes as a combination of treatment, provider, and patient characteristics (or treatment “cell”), we excluded combinations with fewer than 30 episodes over the 6-year period unless deemed clinically relevant (e.g., lithium monotherapy). The final data set contained 9,054 episodes of depression distributed across 120 cells (Table 1). These cells corresponded to 30 treatments modified by two provider types and seven patient characteristics (e.g., treatment with an SSRI for less than 30 days by a specialty care provider for a female subject 18–49 years of age with no medical or substance abuse comorbidity).

Because we wished to characterize the effectiveness of treatment practices commonly used in 1990s usual depression care, the focus of this paper is on the 10 most frequent treatments, observed in 8,160 episodes (90%) of the larger pool (Table 1).

Characterizing Treatment Effectiveness

Expected outcomes for care frequently observed in the claims data were assessed with the two-stage “modified Delphi” technique (25). Concretely, we 1) summarized published efficacy information for prevalent treatments; 2) elicited estimates of treatment effectiveness from a panel of experts, with their ratings informed by our literature review; and 3) asked experts to rate those treatments that had substantial disagreement during a meeting in which these treatments had been discussed.

Expert Selection

From suggestions by National Institute of Mental Health staff and other highly regarded depression researchers, we created a short list of U.S.-based practitioners recognized by their peers as expert clinician-researchers in the field of depression. If they had been in clinical practice for at least 5 years and had treated depressed patients in the past year, we invited them to participate. All experts approached accepted our invitation. The expert panel (listed at the end of this article) included four psychiatrists, four clinical psychologists, and two primary care clinicians, all of whom were involved in academic research and clinically active (years practicing, mean=20, range=11–42), with a mean of 64 major depressive disorder patients treated over the previous year (range=4–200).

Literature Review

In preparation for the elicitation process, experts were provided with summaries of published outcomes for the treatments whose effectiveness would be asked to estimate. The literature review covered randomized controlled trials for adult outpatients with acute-phase major depressive disorder published between 1975 and October 1998. Studies were excluded either when the only comparator was not approved by the Food and Drug Administration for use as an antidepressant drug (except for clomipramine and fluvoxamine) or when outcome results were not reported as rates. Findings were summarized according to published rates of response.

The evidence was characterized as adequate if 10 or more studies had investigated the efficacy of the treatment or if the number of subjects was greater than 750; the evidence was characterized as inadequate if neither was the case.

The Elicitation Process

For a thorough discussion of the elicitation process, please refer to Normand et al. (23). In the first stage of the elicitation, experts independently estimated outcome, defined as treatment-related changes in score on the Hamilton Depression Rating Scale. Experts were asked to consider 100 outpatients with major depression and moderate to severe symptoms (i.e., Hamilton depression scale score=22) seeking treatment in 1998 and estimate what number would fall into each of four outcome categories after 16 weeks of usual care: remission (Hamilton score<8), significant improvement (score<13), mild improvement (score<18), and no change (score>21).

First-stage ratings were analyzed to calculate expert agreement. In the second stage of the elicitation, experts openly discussed at a face-to-face meeting treatments that had significant disagreement, and then independently rerated those treatments. At the meeting, experts agreed that effectiveness ratings should be anchored to likely outcomes given no treatment. They estimated that the no-treatment condition would have an approximate 16-week remission rate of 15%.

Computation of Effectiveness Estimates

Upon completing the elicitation, ratings were averaged across the 10 experts. Treatment response (i.e., Hamilton depression scale score<13) was not elicited but rather computed as a sum of the mean probabilities of remission and significant improvement. To illustrate the effectiveness of usual depression care, we report on expert-estimated rates of remission, response, and no change.

For clarity, we operationalized three categories of effectiveness: minimal, moderate, and high. Treatments with expert-estimated rates of ≤20% for remission, ≤45% for response, and ≥25%
for no change were classified as minimally effective. Treatments with rates of ≥30% for remission, ≥60% for response, and ≤15% for no change were classified as highly effective. Treatments with outcomes between these two poles were classified as moderately effective.

Standard errors for the weighted mean probabilities were calculated by using bootstrap methods (26). We did not evaluate the statistical significance of the difference between all 10 treatments at each outcome category because of the inevitable methodological problems associated with multiple comparisons. When we did test for differences, empirically calculated 95% confidence intervals (CIs) were used.

System effectiveness, the sum of all the effects associated with depression care in a population, was operationalized as a weighted average of expert-estimated rates of remission, response, and no change for the 10 most frequent treatments.

**Results**

**Characteristics of Usual Care**

Over the 6-year period covered by the claims data and across all episodes (N=9,054), psychotherapy alone (N=3,723) and SSRI treatment, alone or in combination (N=2,227), were the most frequently utilized interventions (Table 1). Four to nine psychotherapy visits was the most frequent treatment (14% of episodes), followed by one psychotherapy visit, one brief office visit, and unspecified...
Mental health interventions. Tricyclic antidepressants were minimally used, and only 74 episodes (0.8%) involved definitive anxiolytic use, which was always coprescribed with antidepressants. The treating clinician was a specialist in approximately three-fourths of all episodes, and the most frequent patient characteristic was women aged 18–49 without medical or substance abuse comorbidity. While 38% of episodes involved medical comorbidity, only 16 (0.2%) documented a substance abuse diagnosis.

Notable utilization trends between 1991 and 1996, based on frequencies of interventions relative to annual episodes, were a consistently high utilization of psychotherapy-only treatments (36%–45%) and a steady growth in SSRI utilization through mid-period (SSRI only: 2% to 7%–8%; combined SSRI and psychotherapy: 11% to 17%–21%).

**Expert-Estimated Effectiveness of Usual Care**

According to expert ratings of the 10 most frequent treatments, three treatments (SSRI ≥60 days with 1–3 psychotherapy visits, SSRI ≥60 days with four or more psychotherapy visits, and 10–24 psychotherapy visits) were highly effective, two treatments (4–9 psychotherapy visits and SSRI ≥60 days) were moderately effective, and the five remaining treatments were minimally effective (Table 2). The small difference in remission rates between 1) SSRI ≥60...
days with four or more psychotherapy visits and 2) 10–24 psychotherapy visits, selected for a statistical test of such difference because of their contrasting paradigms and superior effectiveness, was not statistically significant (95% CI=−0.06 to 0.05).

Weighted averages of the expert-estimated rates of remission, response, and no change for the 10 most frequent treatments were 23%, 48%, and 24%, respectively, placing the overall performance of this system (i.e., system effectiveness) at a level barely above our minimal effectiveness criteria (Table 2). Thus, while experts predicted that 23% of episodes treated with these interventions would be in remission after 16 weeks of treatment, they also predicted that 24% of episodes would be unimproved.

### Adequacy of the Evidence

Only six treatments, used in 15% of episodes, had adequate evidence (i.e., at least 10 studies or 750 subjects) (Table 3). Of the 10 most frequent treatments, only SSRI ≥60 days and 10–24 psychotherapy visits had adequate evidence. The treatments with more than one intervention (i.e., SSRI ≥60 days with 1–3 psychotherapy visits and SSRI ≥60 days with four or more psychotherapy visits) had adequate evidence for only one of the interventions (SSRI ≥60 days). The evidence base was inadequate for the remaining six treatments.

### Discussion

Common patterns of care in this insured population involved psychotherapy and/or SSRIs delivered by specialists to premenopausal women without comorbid conditions. It is not surprising that in a 1990s setting with high SSRI utilization, tricyclic antidepressants and anxiolytics were infrequently used. Utilization trends over the 6-year period included a consistently high use of psychotherapy and increasing SSRI use over time.

According to expert estimates, the most effective of high-volume treatments for moderate to severe major depressive disorder were SSRI ≥60 days with four or more psychotherapy visits, 10–24 psychotherapy visits, and SSRI ≥60 days with 1–3 psychotherapy visits. Employed in 21% of all episodes, the utilization of these treatments was supported by adequate evidence for at least one of the interventions used in the treatment (SSRI ≥60 days). Despite its inadequate research base, 4–9 psychotherapy visits was the most frequent treatment and one of two rated by experts as moderately effective. The remaining five treatments, utilized in almost half of all episodes despite inadequate evidence in support of their use, were considered minimally effective.

This system's effectiveness—a weighted average of the effects associated with the 10 most frequent treatments—was only modest, better only than outcomes predicted for

---

**Table 2. Effectiveness of the 10 Most Frequent Treatment Interventions Seen in Usual Outpatient Care of Acute-Phase Major Depression for a Privately Insured Population, 1991–1996**

<table>
<thead>
<tr>
<th>Effectiveness Level and Intervention</th>
<th>Number of Treatment Cells</th>
<th>Episodes N</th>
<th>%</th>
<th>Number of Studies</th>
<th>Mean (Weighted) Expert-Estimated Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 psychotherapy visit</td>
<td>9</td>
<td>978</td>
<td>11</td>
<td>4</td>
<td>Remission: 15, Response: 34, No Change: 35</td>
</tr>
<tr>
<td>1 brief office visit</td>
<td>11</td>
<td>901</td>
<td>10</td>
<td>0</td>
<td>Remission: 16, Response: 36, No Change: 35</td>
</tr>
<tr>
<td>Unspecified mental health intervention</td>
<td>8</td>
<td>894</td>
<td>10</td>
<td>0</td>
<td>Remission: 15, Response: 37, No Change: 31</td>
</tr>
<tr>
<td>Medical intervention only</td>
<td>9</td>
<td>816</td>
<td>9</td>
<td>0</td>
<td>Remission: 18, Response: 38, No Change: 27</td>
</tr>
<tr>
<td>2–3 psychotherapy visits</td>
<td>8</td>
<td>816</td>
<td>9</td>
<td>1</td>
<td>Remission: 19, Response: 42, No Change: 25</td>
</tr>
<tr>
<td>Moderate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4–9 psychotherapy visits</td>
<td>9</td>
<td>1,292</td>
<td>14</td>
<td>8</td>
<td>Remission: 27, Response: 57, No Change: 19</td>
</tr>
<tr>
<td>SSRI ≥60 days</td>
<td>11</td>
<td>552</td>
<td>6</td>
<td>35</td>
<td>Remission: 27, Response: 61, No Change: 16</td>
</tr>
<tr>
<td>High</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSRI ≥60 days with 1–3 psychotherapy visits</td>
<td>7</td>
<td>473</td>
<td>5</td>
<td>0</td>
<td>Remission: 30, Response: 61, No Change: 15</td>
</tr>
<tr>
<td>10–24 psychotherapy visits</td>
<td>8</td>
<td>637</td>
<td>7</td>
<td>12</td>
<td>Remission: 35, Response: 64, No Change: 15</td>
</tr>
<tr>
<td>SSRI ≥60 days with ≥4 psychotherapy visits</td>
<td>9</td>
<td>801</td>
<td>9</td>
<td>1</td>
<td>Remission: 36, Response: 65, No Change: 13</td>
</tr>
</tbody>
</table>

---

**Table 3. Treatment Interventions With Adequate Evidencea**

<table>
<thead>
<tr>
<th>Treatment Intervention</th>
<th>Percent of Episodes</th>
<th>Number of Studiesb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tricyclic antidepressant ≥60 days with optional anxiolytics</td>
<td>0.3</td>
<td>55</td>
</tr>
<tr>
<td>SSRI ≥60 days</td>
<td>6.1</td>
<td>35</td>
</tr>
<tr>
<td>SSRI &lt;30 days</td>
<td>0.7</td>
<td>32</td>
</tr>
<tr>
<td>Other antidepressant medicationc ≥60 days with optional anxiolytics</td>
<td>0.4</td>
<td>12</td>
</tr>
<tr>
<td>10–24 psychotherapy visits</td>
<td>7.0</td>
<td>12</td>
</tr>
<tr>
<td>SSRI ≥60 days with anxiolytics</td>
<td>0.2</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>14.7</td>
<td>156</td>
</tr>
</tbody>
</table>

---

a From Table 1, specific combinations of treatment, provider, and patient characteristics.

b Numbers are not additive because a given study may have offered evidence for more than one intervention type; studies are included in this count if they met review criteria.

c Remission ≤20%, response ≤45%, and no change ≥25%.

d Remission ≥30%, response ≥60%, and no change ≤15%.

e Weighted average of the expert-estimated rates of remission, response, and no change for the 10 most frequent treatments.
minimally effective interventions and for the no-treatment condition. Relative to randomized controlled trials that have included a usual care arm and have reported symptom-based outcomes at 3 or 4 months (all of them primary care-based), our system-wide 16-week response rate of 48% was higher than published results (i.e., 27%–44%) (14, 15, 19), while the 23% remission rate was consistent with published results (i.e., 21%–24%) (12, 13, 18).

Conclusions

Challenging the notion that usual care outcomes are probably worse than outcomes generated by randomized controlled trials, experts predicted high effectiveness for several treatments frequently used in this privately insured population. This finding may be explained by the better matching of patients and treatments or patients and clinicians in usual care versus experimental conditions. These encouraging results are tempered by the fact that half of high-volume treatments—utilized in half of all episodes despite their inadequate evidence—were rated by experts as minimally effective. System-wide, experts estimated that only 23% of episodes managed with the 10 most frequently utilized treatments would be in remission after 16 weeks, a modest outcome when compared with the expert-estimated 15% rate given no treatment. It may be argued that usual care patients are less severely depressed than was assumed in this study and that care judged ineffective was indeed sufficient given high odds of spontaneous remission (27, 28). In this scenario, this system’s effectiveness would have been underestimated. However, in our predominantly specialist-treated population, patients whose care was rated as ineffective experienced more recurrences than those treated with better-rated treatments (23), a fact that argues against their being a less severe subgroup.

The original study was designed to assess the system effectiveness of usual care in a private managed care environment. The study’s most important economic finding (i.e., most depression dollars were spent on the most effective interventions [22, 24]), to some degree attenuates the significance of the troubling high prevalence of minimally effective care.

Future Directions

Efficacy studies have played a fundamental role in building the evidence base for depression care, but many questions remain unanswered. Although efforts are underway to shed light on the characteristics of usual depression care and its effectiveness (27), the field still lacks firm evidence on outcomes associated with common treatments and whether (and which) factors unique to usual care lead to better outcomes vis-à-vis efficacy findings. More research needs to be conducted on the contributors to ineffective care for moderately to severely depressed patients treated in either primary or specialty care, with the aim of developing targeted quality improvement interventions. Also, more information is needed on other forms of usual care, such as that provided to low-income populations covered by public insurance.

We, like others (29), believe that effectiveness research should take advantage of large administrative databases, a design that lies “at the interface of clinical trials and effectiveness studies” (3). As we hope to have demonstrated, claims data can be used to characterize usual mental health care. Outcome associated with such care may be assessed with elicitation methods such as ours or with process or gross outcome indicators (30). More ambitious projects may entail conducting mixed-design studies, where both internal and external validity are optimized, or using such methodological advances as instrumental variables or propensity analyses to establish causality (3).

Study Limitations

The usual depression care described in this study occurred in the context of rather generous private insurance, a fact that may explain the large proportion of specialist-managed episodes. Although ready access to specialty care may not be generalizable to public insurance or more modest private insurance (31), this form of usual care is by no means uncommon in the present-day United States.

The method used to estimate treatment effectiveness relied both on administrative data—limited in the breadth and quality of clinical information available when compared with clinical trials and expert opinion—which may be considered a biased or unreliable source of outcome information.

We addressed the first set of potential problems by circumscribing this study to major depressive disorder, thus optimizing the specificity of depression diagnoses. The fact that insurance benefits for this population were quite generous may have mitigated against out-of-system and unaccounted care.

Despite justified concerns over the use of expert judgment to evaluate outcomes, we believe that the use of expert opinion in a little understood yet vital aspect of health care represents a legitimate research tool as long as the limitations of the method are clear and the findings are primarily used to motivate future research. A related precedent in this regard is the use of expert opinion in the generation of practice guidelines (e.g., APA Practice Guidelines Series [32]). Further, the modified Delphi method provided us with a technique to minimize systematic biases and maximize reliability. We do acknowledge that the manner in which the expert panel was selected—and its constitution—may have biased our effectiveness findings.

Received Sept. 4, 2001; revision received Aug. 6, 2002; accepted Nov. 4, 2002. From the Department of Psychiatry, The Cambridge Hospital, Cambridge, Mass.; the Department of Health Care Policy, Harvard Medical School, Boston; the Department of Biostatistics, Harvard School of Public Health, Boston; and the Center for Mental Health Services, Harvard School of Public Health, Boston.
USUAL CARE FOR MAJOR DEPRESSION

Health Services Research, University of Maryland, Baltimore. Address reprint requests to Dr. Horvitz-Lennon, Department of Psychiatry, The Cambridge Hospital, Harvard Medical School Department of Health Care Policy, 26 Central St., Somerville, MA 02143; horvitz@hcp.med.harvard.edu (e-mail).

Supported by a grant from the John D. & Catherine T. MacArthur Foundation (97-49667A-HE) and NIMH grant MH-62028.

The authors thank Anupa Bir and Susan Busch for their program support and the practitioners who served on the expert panel: David Adler, Christopher Callahan, Ellen Frank, Wayne Katon, Janice Krupnick, Jeanne Miranda, John Rush, H.C. Schulberg, Michael Thase, and John Williams, Jr.

References


726 http://ajp.psychiatryonline.org

Am J Psychiatry 160:4, April 2003