

# Image Theory's counting rule in clinical decision making: Does it describe how clinicians make patient-specific forecasts?

Paul R. Falzer\*

D. Melissa Garman†

## Abstract

The field of clinical decision making is polarized by two predominate views. One holds that treatment recommendations should conform with guidelines; the other emphasizes clinical expertise in reaching case-specific judgments. Previous work developed a test for a proposed alternative, that clinical judgment should systematically incorporate both general knowledge and patient-specific information. The test was derived from image theory's two phase-account of decision making and its "simple counting rule", which describes how possible courses of action are pre-screened for compatibility with standards and values. The current paper applies this rule to clinical forecasting, where practitioners indicate how likely a specific patient will respond favorably to a recommended treatment. Psychiatric trainees evaluated eight case vignettes that exhibited from 0 to 3 incompatible attributes. They made two forecasts, one based on a guideline recommendation, the other based on their own alternative. Both forecasts were predicted by equally- and unequally-weighted counting rules. Unequal weighting provided a better fit and exhibited a clearer rejection threshold, or point at which forecasts are not diminished by additional incompatibilities. The hypothesis that missing information is treated as an incompatibility was not confirmed. There was evidence that the rejection threshold was influenced by clinician preference. Results suggests that guidelines may have a de-biasing influence on clinical judgment. Subject to limitations pertaining to the subject sample and population, clinical paradigm, guideline, and study procedure, the data support the use of a compatibility test to describe how clinicians make patient-specific forecasts.

Keywords: decision making, clinical judgment, forecasting, evidence based medicine, treatment guidelines, patient-centered care, clinical training, naturalistic decision making, mental illness, schizophrenia.

## 1 Introduction

The naturalistic movement in human decision making was a response to what its originators believed were significant limitations of classic behavioral decision theory. Among these was a putative under-emphasis on processes and events that occur at the outset, in the course of recognizing decisional situations and identifying feasible alternatives (Beach, 1997; Klein, Orasanu, Calderwood & Zsombok, 1993). Early efforts to address this limitation included the schema-based image theory (Beach, 1998), which proposed that decision making occurs in two phases (Beach & Potter, 1992; Van Zee, Paluchowski & Beach, 1992). At the first phase, decision makers narrow the range of alternatives by applying a simple, non-compensatory, test of compatibility with their ethical standards, values, beliefs, goals, and plans (Beach

& Strom, 1989; Richmond, Bissell & Beach, 1998). This test, originally called the "simple counting rule", is conducted by tallying incompatible attributes and rejecting alternatives whose incompatibility count exceeds a threshold (Beach, Smith, Lundell & Mitchell, 1988).

The two-phase conception and the compatibility test have been applied to a variety of situations. However, most research has focused on common personal and organizational tasks, such as purchasing consumer goods, making career choices, allocating resources to product development, and selecting job candidates. The current paper brings compatibility testing to the arena of clinical decision making. It describes how clinicians make treatment decisions that incorporate distinct kinds of information, including general knowledge about treatment effectiveness that is based on clinical and epidemiological studies, and patient-specific information that is obtained through clinical examination. Clinical decision tasks have been simplified somewhat by the advent and widespread use of treatment guidelines, which identify evidence-based interventions, specify supporting evidence, and include algorithms and assessment procedures (Lomas et al., 1989; S. H. Woolf, 1990, 1993). The role and function of guidelines in clinical judgment has also fueled controversy and intensified an ongoing con-

This study was funded by a grant from the National Institute of Mental Health, 1R34 MH070871-01. The authors wish to express their appreciation to Lee Roy Beach, for his inspiration and his assistance in developing the project, and to Jonathan Baron, for his editorial help.

\*VA Connecticut Healthcare System, Clinical Epidemiology Research Center, 950 Campbell Avenue, Mailcode 151B, West Haven, CT 06516. Email: paul.falzer@yale.edu.

†State of Connecticut, Department of Mental Health and Addiction Services Southwest Connecticut Mental Health System.

flict between practitioners and service researchers (Dickenson & Vineis, 2002; Satterfield et al., 2009). The latter tend to regard guideline recommendations as standards of care, and they are inclined to gauge performance by comparing clinician decisions against guideline recommendations (Drake et al., 2001; Grimshaw & Eccles, 2004). This tack is vehemently opposed by practitioners and others, who insist that the purpose of guidelines is to assist in the expert task of “contextualizing” general recommendations by incorporating patient-specific information (Maier, 2006; Ruscio & Holohan, 2006; A. D. Woolf, 1997).

By now, the parties to the so-called “evidence debate” (McQueen, 2002) about guidelines as decision aids versus standards of care have marshaled such impressive support that discussion has reached a stalemate. On the one hand, guideline recommendations are based on evidence of treatment effectiveness; endorsing them overcomes well documented tendencies of clinicians to get lost in complications, take refuge in lore, and make decisions that are inconsistent both temporally and across geographical boundaries (Eddy, 1990; Moffic, 2006; Weisz et al., 2007). On the other hand, practitioners are mandated to treat patients, not diseases. Guidelines give categorical recommendations; they cannot be expected to factor in patient specific circumstances, account for heterogeneous conditions and treatment responses, or incorporate patient values and preferences.

Of all of the solutions to this predicament that have been proposed to date, Eddy’s (2005) is perhaps the most feasible. In his vision of “evidence-based decision making” (EBDM), a guideline serves as a point of reference. Its recommendations should be followed in most cases, but clinicians are expected to adapt them in light of patient-specific needs and idiosyncratic circumstances. The EBDM vision strikes a balance between evidence and application, but Eddy’s proposal lacks an alternative to what can be called “conformance testing,” or comparing treatment recommendations against a strict standard. In previous work, we proposed an alternative inspired by image theory’s compatibility test, and used the counting rule to examine how clinicians systematically factor patient-specific information into guideline recommendations (Falzer & Garman, 2009, 2010). Our study used a treatment guideline developed at the Yale Department of Psychiatry for patients with schizophrenia (Sernyak, Dausey, Desai & Rosenheck, 2003). This is a progressive, five-step, algorithm that recommends treatment switches for patients who have not responded adequately to full trials of antipsychotic medication. The guideline was derived from a widely disseminated set of recommendations for treating patients with schizophrenia (Lehman et al., 1998), and developed specifically to favor less expensive generic treatments over newer (second

generation or “atypical”) antipsychotic agents. The five steps represent progressive orders of unit cost (Rosenheck, Leslie & Doshi, 2008).

The guideline calls for treatment to begin with a course of first-generation antipsychotic therapy such as haloperidol or fluphenazine. At step two, patients who do not respond adequately after three months are switched to a different first-generation treatment. At step three, non-responders are switched to a different medication class, a second-generation treatment such as risperidone or olanzapine. Step four is a trial of a different second-generation treatment. Clozapine therapy is introduced at the fifth and final step. Switch recommendations are guided by two items from the Clinical Global Impression Scale (Guy, 1976): A *severity* scale measures the patient’s illness, and a *progress* scale measures the treatment response. Each scale is rated from 0 to 7, with higher scores indicating greater severity of illness or a poorer response. A switch is recommended if the illness score is 4 or higher, indicating at least moderate illness, and the progress score is 3 or higher, indicating minimal improvement at best.

Our previous work (Falzer & Garman, 2010) used case vignettes to systematically vary four factors and asked psychiatric residents to make treatment recommendations. Subjects were introduced to the five-step algorithm and switch criteria. Vignettes were constructed from a design in which each factor had either a compatible level or a discrepant level. “Discrepant” levels are technically consistent with the guideline, but they are either inconsistent with clinical practice or introduce an additional relevant factor. Discrepant information weakens the guideline recommendation and may lead clinicians to consider an alternative. The factors and their levels were:

1. For the *progress* factor: “The progress score is 3, minimally improved over the past 6 months,” versus “the progress score is 6, much worse over the past 6 months.” Minimal improvement is discrepant because it barely meets the switch criteria and indicates that the current treatment may be beneficial. “Much worse” is compatible because it clearly meets the switch criteria and strongly indicates the need for a treatment change.
2. For the *illness* factor: “The illness score is 4, moderate illness at present,” versus “The illness score is 6, severe illness at present.” Moderate illness is discrepant because it barely meets the switch criteria, whereas “severely ill” clearly meets the criteria and strongly indicates the need for a treatment change.
3. For the *adherence* factor: Subjects were presented with a 4-point adherence scale. The lowest rating was “1: never/almost never takes medications as

prescribed (0–25% of the time).” The highest rating was “4: always/almost always takes medications as prescribed (75–100% of the time).” The lowest score is discrepant because non-adherent patients are unlikely to benefit from a treatment change. The highest score is compatible because patients who take medication as prescribed are most likely to achieve its full benefit.

4. For the *likelihood* factor, subjects were presented with one of four sets of likelihoods. Three sets were discrepant because they indicated a low likelihood of a positive response to the guideline-recommended treatment. For one of the discrepant sets: “With this subset of schizophrenic patients, [following the switch recommendation] will have the following results: 10% chance of significant improvement, no longer treatment resistant; 40% chance of no significant change; 50% chance of getting significantly worse, requiring hospitalization.” (There were two other low discrepant sets: 10%-80%-10%, which suggests that a switch would probably be ineffective; and 45%-10%-45%, which suggests that a switch is risky.) The fourth set was compatible owing to a high likelihood of a positive response. The text was the same as noted above, but percentages were 50%-40%-10%, suggesting a high likelihood of significant improvement and a low likelihood of decompensation.

Note that the adherence and likelihood factors were not included in the guideline, but introduce information that clinicians would regard as relevant to making a treatment recommendation.

Each subject in the earlier study evaluated 64 vignettes. The vignettes were constructed from a 2 x 2 x 2 x 4 x 2 design. The first four variables represent the four factors described above. The last 2-level factor refers to the guideline step, 2 or 4. Overall, 42% of their recommendations concurred with the guideline. However, the endorsement rate ranged from 32% to 91%, depending on the number of discrepancies. There was a significant inverse linear relationship between the likelihood of endorsing the guideline recommendation and the number of discrepant attributes, and the discrepancy count explained 65% of the within-subject variance (Falzer & Garman, 2010).

Patients ask a variety of questions in the routine course of consultation, but two questions tend to predominate. They are: “what do I have?” and “what are my chances?” The first question requests a diagnostic classification; the second asks for a specific or tailored forecast. Most patients will not be satisfied with a general likelihood estimate that applies to a disease population, followed by a caveat to the effect that “all patients are different.” From the perspective of EBDM, forecasting requires expertise

in using a guideline to modify a general estimate in light of relevant patient-specific factors (Visweswaran et al., 2010).

The current study uses the counting rule to examine how clinicians bring a combination of general and case-specific knowledge to the task of forecasting a patient’s treatment response. The study focuses on two principal findings from image theory research: One is that probabilities are treated as attributes at the initial phase of decision making (Potter & Beach, 1994b; Van Zee, et al., 1992). The other finding is that, in some situations, certain attributes have greater weight than others (Beach, Puto, Heckler, Naylor & Marble, 1996). In practicing EBDM, we expect that, in making a patient-specific forecast, clinicians will give greater weight to general likelihoods than other clinical factors.

In the current study, subjects made two forecasts for each of 8 case vignettes. In one forecast, they projected the likelihood of a positive treatment response if the guideline recommendation is followed from step 3 forward; the other forecast projected the likelihood of a positive treatment response if the guideline recommendation is *not* followed, i.e., if the guideline recommends change and no change is made, or if the guideline recommends change to one treatment and a change is made to a different treatment. This procedure allowed us to examine how the absence of general likelihood data affects the counting rule.

So that the subjects can make these judgments, we do not provide likelihood information. General likelihoods may be absent for a variety of reasons, but most commonly because alternatives to a guideline recommendation have not been extensively investigated. Image theory studies have found that decision makers treat missing information as a violation (Potter & Beach, 1994a). In other words, a significant missing piece of information is treated as incompatible information. However, clinicians who are familiar with treatment alternatives and are accustomed to comparing guideline-recommended treatments with commonly-used alternatives may handle missing information differently, perhaps by adjusting the likelihoods of known treatments.

Asking subjects to make two forecasts also allows us to identify the higher forecast as the preferred alternative and examine how preference influences the counting rule. Forecasts may be biased in a variety of ways (Alexander, 2008; Harvey, 2007; Wolfson, Doctor & Burns, 2000). A “value induced” or preference bias is frequently mentioned in the clinical decision making literature to explain the ostensible tendency of clinicians and patients to make over-optimistic forecasts about favored alternatives (Gurmankin Levy & Hershey, 2006; also see Krizan & Windschitl, 2007; Levy & Hershey, 2008). A study by Ditto and Lopez (1992) found that preference bias is appar-

ent in decisional processes as well as summary forecasts. Specifically, judgments are reached more quickly and require less information when they are consistent with favored conclusions. This finding suggests that preference may influence what image theory's calls the "rejection threshold"—the point at which prospective alternatives are rejected because of too many discrepant attributes. A key finding of image theory research is that once the threshold is met, additional discrepancies have limited influence on whether to eliminate a prospective course of action from further consideration (Beach & Strom, 1989). The influence of the rejection threshold on likelihood forecasts can be seen by plotting mean likelihoods at each violation count. Likelihoods should decrease somewhat as the number of violations increase, then drop precipitously and flatten out. The current study examines the influence of preference on the rejection threshold by proposing that favored alternatives have a higher (that is, a more generous) threshold than non-favored alternatives.

## 2 Method

### 2.1 Guideline and task

Subjects evaluated eight case vignettes that were selected from the group of 64 that the same subjects had reviewed in performing the treatment recommendation task described above (Falzer & Garman, 2010). The vignettes were rated in the manner described below at the step three of the guideline, after the hypothetical patient had failed to respond adequately to two courses of a first-generation antipsychotic treatment. These patients comprise the roughly 15 to 25% of patients with schizophrenia who are "treatment resistant" (Brenner et al., 1990; Falzer, Garman & Moore, 2009). Ratings consisted of two forecasts of a favorable treatment response: a) if treatment followed the guideline recommendations from step three forward, and b) if treatment departed from the guideline recommendation. The forecasts were made sequentially, using a 0 to 100 scale. As with the previous study, subjects were able to consult the guideline as they performed the task and were instructed to proceed through the vignettes in the order they were presented.

As experienced psychiatric trainees, the subjects are well aware of at least three viable alternatives to following the guideline from step three forward. One alternative, especially with a partial response, is to continue the current treatment. The second is to recommend clozapine earlier than the guideline-recommended step five. There is extensive support for using clozapine for treatment-resistant patients (Falzer & Garman, 2012; Kane, 2004), but despite its effectiveness, clozapine tends to be underused (Fayek, Flowers, Signorelli & Simpson, 2003; Mistry & Osborn, 2011; Nielsen, Dahm, Lublin & Taylor,

2010). The third alternative is embodied in the American Psychiatric Association's (APA's) recommendation to introduce depot (long-acting, intra-muscular injected) medications for patients who have demonstrated poor adherence to orally administered treatments (Lehman, Lieberman, et al., 2004). Treatment with an injectable medication may begin with a first-generation formulation, even if the oral formulation was previously tried.

### 2.2 Study design

The eight vignettes were sorted into four random orders and then presented to the subjects at random. A fully balanced 2 x 2 x 2 design was created by manipulating three factors: general likelihood of a positive treatment response, course of the illness, and patient-specific adherence. Each factor had two levels. One level represented compatibility between the case and the guideline recommendation; the other level represented incompatibility and is treated as a violation. The factors and levels are as follows:

1. *General likelihood of a positive treatment response:* For the compatible level, a 50% likelihood of a positive response, a 40% likelihood of no change, and a 10% likelihood of a negative response. For the violation level, a 10% likelihood of a positive response, a 40% likelihood of no change, and a 50% likelihood of a negative response. These likelihoods may seem low, but they are consistent with current findings about the limited effectiveness of antipsychotic medication for patients with treatment-resistant schizophrenia, and the risk inherent to switching from one treatment to another.
2. *Course of the illness:* As described in the previous section, the treatment guideline uses two items from the Clinical Global Inventory (CGI) to assess patients' current condition and their progress during the current treatment. In all 8 vignettes, the progress item score was "4, no change," which calls for a switch to step 3. For the compatible level, the severity item score was "6, severely ill." For the violation level, the condition score was "4, moderately ill." To a layperson these scores may seem backwards, but for a trained psychiatrist a severe condition combined with lack of progress is attributed to the current treatment's lack of effectiveness. Consequently, a severe illness and no progress is compatible with the guideline's switch recommendation. Moderate illness combined with no progress suggests that the patient's condition is on a stable or deteriorating course. For these patients, clozapine is the treatment of choice; alternatively, the current treatment would be continued.



3. *Patient-specific adherence*: As described in the previous section, adherence was high for the compatible level (75% or greater) and low for the violation level (25% or lower). There is ample evidence that low adherence reduces the likelihood of a positive treatment response (Ascher-Svanum et al., 2006). As noted above, the APA guideline recommends a depot medication when adherence is low. What cannot be determined from the vignette, as in actual practice, is how the patient's adherence is affected by the treatment regimen, and consequently whether adherence would change with a different treatment.

### 2.3 Hypothesis testing and data analysis

The study tested three hypotheses. The first is that clinicians use an unequally-weighted counting rule in making patient-specific forecasts that follow the guideline. The hypothesis is examined by treating the first forecast—the likelihood of a positive response if the guideline is strictly followed—as the dependent variable, and two categorical “violations” variables as independent variables. One variable ranges from 0 to 3 and represents an equally-weighted violation count. It is computed by a simple sum of the violations in each vignette. The other variable ranges from 0 and 4 and represents unequally-weighted violation count. It is computed by giving the general likelihood factor twice the weight of the course and adherence factors.

The hypothesis is tested by creating two linear mixed effects models and examining each independent variable separately. For the hypothesis to be confirmed, the unequally-weighted variable must be significantly associated with the patient specific forecast. If the equally-weighted variable is also significant, the two models will be compared using Akaike's Information Criterion (AIC) index, a “lower is better” goodness-of-fit measure (Akaike, 1974). In addition, model means will be inspected for evidence of rejection threshold. Hypothesis testing uses the linear mixed model algorithm in SPSS 19 (SPSS Inc., 2011), with a diagonal covariance type. A numeric subject identifier is treated as subject-level factor; trial number (1–8) is a repeated measure.

The second hypothesis is that clinicians use an unequally-weighted counting rule in making patient-specific forecasts that do not follow the guideline. The analytic procedure is the same as for hypothesis one, with the second forecast as the dependent variable. Three mixed models are compared: The first model treats the absence of general likelihood information as a violation. The second model substitutes the violation level of guideline's general likelihood. The third model substitutes the general likelihood as a weighted violation.

The third hypothesis examines how clinicians' prefer-

ences that either favor or oppose the guideline influence their use of the counting rule. The hypothesis that preference significantly influences the counting rule is tested by creating a preference variable, then examining the preference by violations interaction. A significant preference by violations interaction for each rating confirms the hypothesis. The preference variable is created by subtracting the second forecast from the first for all eight vignettes. A positive difference indicates that for a given vignette, a subject favors the guideline recommendation. A negative difference indicates that the subject favors an alternative to the guideline recommendation. A difference of 0 indicates no preference. Violation factors that pertain to both ratings will be tested, provided that hypotheses 1 and 2 are confirmed. Tests of the first two hypotheses will also determine whether an equally- or unequally-weighted rule is used. If only one of the hypotheses is confirmed, then hypothesis 3 will be tested for that rating only. If neither hypothesis is confirmed, then hypothesis 3 will not be tested.

### 2.4 Subjects

Twenty-one volunteer psychiatric residents with experience in treating patients with schizophrenia were recruited as subjects. They were paid \$100.00 to complete a one hour session that included the task described here. The funding source and two local Human Investigation Committees required that recruitment be done passively, to minimize concerns that residents' participation could affect their status or progress in the training program. Consequently, only candidates who were interested in participating contacted the study investigator, and every candidate who contacted the investigator became a subject. The experience requirement limited the sampling frame to third and fourth year residents, and fellows with experience treating patients with schizophrenia. These criteria were verified with the candidates prior to obtaining informed consent. The residency program has no specific training in clinical decision making or in using treatment guidelines. The guideline used in this study had not been incorporated into routine clinical procedures and none of the subjects was acquainted with it.

## 3 Results

Of the 21 subjects, 11 were third year residents, 5 were fourth year residents, and 5 were fellows. The demographic characteristics correspond roughly to the population of the training program, with 14 males and 7 females and a mean age of 33.4 years ( $\pm 3.6$ ). Fourteen listed their race as Caucasian, 6 as Asian, and 1 as other. One male Caucasian resident identified himself as Hispanic. Mean

Table 1: Linear mixed model analysis of the first forecast

<b>Equal weighting</b>				
	<b>Violations</b>	<b>Vignettes</b>	<b>Observations</b>	<b>Est mean</b>
	0	1	21	56.667
	1	3	63	46.039
	2	3	63	34.912
	3	1	21	29.048
<b>Hypothesis tests</b>				
	Violations factor:		F = 12.251 (df=3/41.332), p<.001	
	Linear contrast:		t = -5.805 (df=43.705), p<.001	
	AIC:		1489.552	
	Significant pairs:		0 vs 2-3, 1 vs 2-3	
<b>Unequal weighting</b>				
	<b>Violations</b>	<b>Vignettes</b>	<b>Observations</b>	<b>Est mean</b>
	0	1	21	56.67
	1	2	42	52.81
	2	2	42	27.91
	3	2	42	29.91
	4	1	21	29.05
<b>Hypothesis tests</b>				
	Violations factor:		F = 15.046 (df=4/46.905), p<.001	
	Linear contrast:		t = -6.006 (df=51.266), p<.001	
	AIC:		1466.367	
	Significant pairs:		0 vs 2-4, 1 vs 2-4	

ratings of the two forecasts were almost identical: For the first forecast, = 39.8 ( $\pm 22.3$ ) and ranged from 3 to 90. For the second forecast, = 36.4 ( $\pm 19.7$ ) and ranged from 4 to 90. Subject age, gender, and race had no significant effect on either forecast.

Based on a comparison between the two patient-specific forecasts, the first rating was higher than the second in 66 of the 168 total vignette presentations (21 X 8), or 39.3%. An alternative was favored in 67 presentations, or 39.9%. The two ratings were identical, indicating no preference, in 35 presentations, or 20.8%. The mean difference between the ratings was 20.3 ( $\pm 11.56$ ) when the guideline was favored and -12.0 ( $\pm 9.07$ ) when an alternative was favored. Seven subjects expressed a single preference in all eight vignettes. Of these eight, one subject always favored the guideline, one always had no preference, and five always favored an alternative. Ten subjects expressed multiple preferences in at least two of the eight vignettes; three of these ten subjects expressed all three preferences. Using a multinomial GEE analysis (Hardin & Hilbe, 2003), preference was predicted by year in resi-

dency (Wald  $\chi^2 = 10.597$ , df=2, p=.005). Third year residents were less likely to favor the guideline recommendation and fellows were more likely to favor the guideline recommendation. Consequently, resident year was entered, along with the subject identifier, as a subject-level factor in the mixed model analyses.

### 3.1 First hypothesis

The first hypothesis is that subjects use an unequally-weighted counting rule when they make patient-specific forecasts that follow the guideline. Mixed model analyses of the equally- and unequally-weighted counting rules are in Table 1. Both analyses show a significant inverse linear relationship between mean estimates and the number of violations, and there is evidence of a rejection threshold at 2 violations. In the equally-weighted model, the forecast drops precipitously from a mean of 46 at 1 violation to 35 at 2. In the unequally-weighted model, the forecast drops from a mean of 53 at 1 violation to 28 at

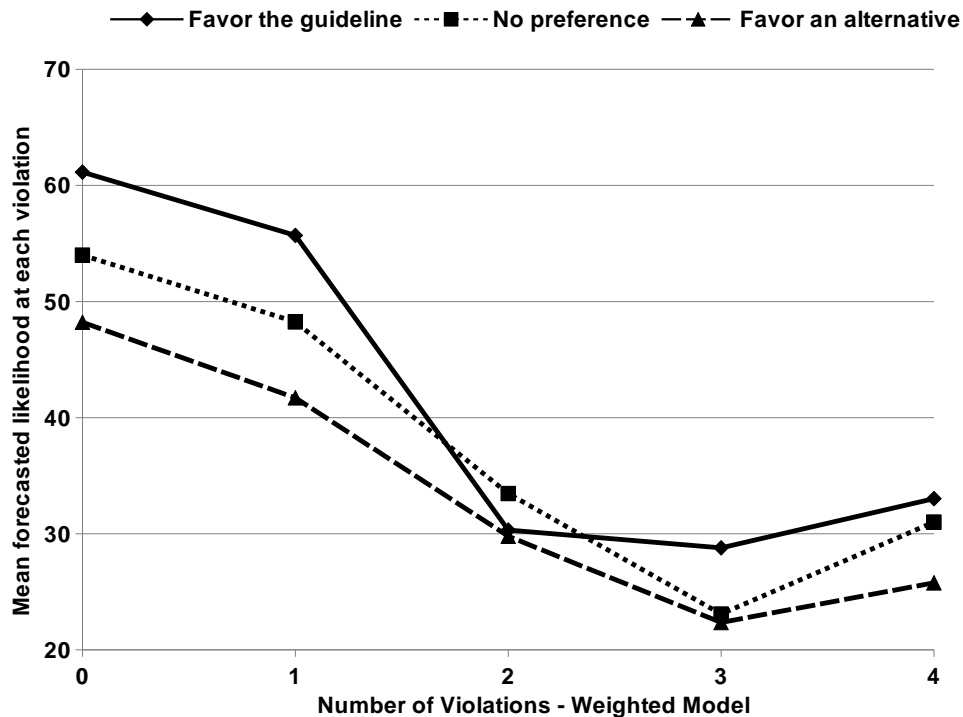
Table 2: Linear mixed model analysis of the second forecast.

<b>Missing as violation</b>				
	<b>Violations</b>	<b>Vignettes</b>	<b>Observations</b>	<b>Est mean</b>
	1	2	42	37.493
	2	4	84	35.823
	3	2	42	33.921
<b>Hypothesis tests</b>				
	Violations factor		F = .36 (df=2.94.205), p=.699	
	Linear contrast		t = -.839 (df=78.362), p=.404	
	AIC		1479.029	
	Significant pairs		None	
<b>Equal weighting</b>				
	<b>Violations</b>	<b>Vignettes</b>	<b>Observations</b>	<b>Est mean</b>
	0	1	21	49.524
	1	3	63	40.365
	2	3	63	29.339
	3	1	21	30.571
<b>Hypothesis tests</b>				
	Violations factor		F = 9.221 (df=3/43.477), p<.001	
	Linear contrast		t = -4.375 (df=44.023), p<.001	
	AIC		1454.198	
	Significant pairs		0 vs 2-3, 1 vs 2-3	
<b>Unequal weighting</b>				
	<b>Violations</b>	<b>Vignettes</b>	<b>Observations</b>	<b>Est mean</b>
	0	1	21	49.524
	1	2	42	46.211
	2	2	42	30.465
	3	2	42	25.979
	4	1	21	30.571
<b>Hypothesis tests</b>				
	Violations factor		F = 12.171 (df=4/43.594), p<.001	
	Linear contrast		t = -5.399 (df=21.279), p<.001	
	AIC		1432.359	
	Significant pairs		0 vs 2-4, 1 vs 2-4	

2. Bonferroni-corrected paired comparisons reported in Table 1 confirm that the drop between 1 violation and 2 is statistically significant for both models, and differences between 0 and 1 and between 2 and 3 were non-significant. The unequally-weighted model has the same pattern of paired-comparisons, but features a steeper drop at the rejection threshold and a slightly better goodness

of fit, as indicated by a 1.5% reduction in the AIC index. These findings indicate that study subjects employed a compatibility test in reaching patient-specific forecasts of a positive treatment response, and support the hypothesis that general likelihood is weighted more heavily than course and adherence.

Figure 1: Mean forecast ratings at each violation point for each preference category.



### 3.2 Second hypothesis

The second hypothesis is that subjects use an unequally-weighted counting rule in making patient-specific forecasts that do not follow the guideline. Tests of the second hypothesis are displayed in Table 2. Results indicate that absence of general likelihood information was not treated as a violation. The estimated means give no indication of a threshold, and neither the violations factor nor the linear contrast tests are significant. (Only the simple counting rule is tested because doubling the value of a constant does not change the results.) The alternative explanation, that subjects substitute guideline likelihoods in making forecasts, is supported by significant violations factor and linear contrast tests. These tests were significant in both the equal-weighted and unequal-weighted models. The pattern of means and rejection thresholds is similar to what was found with the first rating, and the AIC index for the unequally-weighted model is 1.5% lower. These findings indicate that study subjects employed a compatibility test in reaching patient-specific forecasts of a positive treatment response. As with the first ratings, there is limited support for the hypothesis that the general likelihood is weighted more heavily than course and adherence. Further, the fact that the two sets of findings were almost identical suggests that the two forecasts were not made independently.

### 3.3 Third hypothesis

Because an unequal-weighted violations variable provided a slightly better fit in both forecasts, it was used to examine the third hypothesis, that clinicians' preferences for or against the guideline influences their use of the counting rule. This hypothesis was tested by introducing the three-level preference factor (favoring the guideline, favoring an alternative, or indifferent) as a second independent variable and examining the violations by preference interaction for each rating. Both interactions are significant: For the first rating,  $F = 17.199$ ,  $df=14/19.104$ ,  $p<.001$ ; for the second rating,  $F=6.409$ ,  $df=14/30.595$ ,  $p<.001$ . Sub-group analyses illustrate the influence of preference on the counting rule, and specifically on the rejection threshold. Mean forecast ratings at each violation point for each preference are displayed in Figure 1. It shows a rejection threshold of 2 when the guideline recommendation is favored. When an alternative is favored, there is a sharp drop between 0 and 2 violations, as indicated by a significant pair-wise difference. However, the decrement between 1 and 2 is non-significant. Results are similar with no preference, except that the difference between 1 and 4 violations is non-significant, owing to a relatively large standard error (4.522 at 4 violations versus 4.128 at 2 and 4.069 at 3). In addition, the mean ratings at 3 violations were higher for the guideline-favored ratings



than for non-favored or no-preference ratings. Overall, the results suggest that there is a rejection threshold of 2 when the guideline is favored. Otherwise, the rejection threshold is more generous and additional violations continue to exert an influence on the forecasts. Implications of these findings are discussed in the following section.

## 4 Discussion

Current discourse in clinical decision making is daunted by a conflict between those who emphasize adherence to evidence-based practices (Chambers, 2008), and others who view clinical judgment as essential to making patient-specific treatment recommendations (Patel, Kaufman & Arocha, 2002). Questions about the value and importance of clinical judgment are routinely addressed in healthcare policy (Parks et al., 2009; Rosenheck, Leslie, Busch, Rofman & Sernyak, 2008), in discussions about quality of care (Blumenthal, 1996; Zerhouni, 2003), in medical informatics (Fiol & Haug, 2009; Lipman, 2004), and comparative effectiveness research (Basu, 2009; Helfand, 2009). Among the proposals that have been advanced to diminish the conflict and minimize its deleterious influence on healthcare education, policy, and practice, the most fully developed is Eddy's EBDM (Eddy, 2005). It requires a guideline that makes evidence-based recommendations and clinical discretion in applying the recommendations to specific cases. Although EBDM was not developed expressly with image theory in mind, its two-phase conception of decision making complements EBDM's conception of how evidence informs practice: At the first phase, clinicians decide whether to endorse the guideline's recommendation; contingent on a general endorsement, they select a specific treatment.

The current study focused on the first decisional phase and examined the applicability of image theory's simple counting rule to case-specific forecasting. It found that the counting rule describes how clinicians incorporate different kinds of knowledge. Hypotheses that clinicians weight population estimates more heavily than specific attributes were confirmed. However, the difference between the equally- and unequally-weighted counting rules was small (only 1.5%, gauged by the AIC index). The findings that all three attributes were important militates against an explanation frequently mentioned in conferences and anecdotal conversations—that clinicians adopt a “take the best” heuristic by focusing principally on a single cue (Marewski, Gaissmaier & Gigerenzer, 2010).

Asking subjects to make two forecasts of a positive response allowed us to examine how the treatment guideline in combination with clinician preference in-

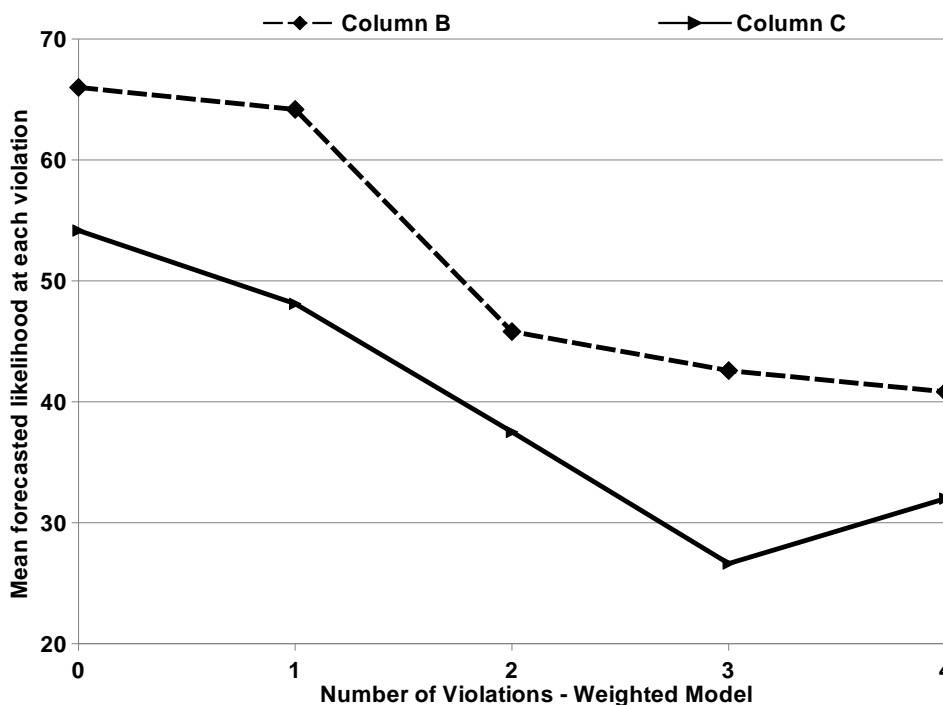
fluences their use of the counting rule. Findings by Ditto and Lopez (1992) led us to expect that favored and non-favored preferences would have different rejection thresholds. The mean forecasts in Figure 1 confirm this expectation. What Figure 1 does not show is the relationship between rating and preference. This relationship is represented in Figure 2. It reports mean forecasts at each violation point, using a weighted count, for ratings that are consistent with preference. The broken line, which displays mean forecasts of the first rating when the guideline is favored, shows a sharp drop at 2 violations. The solid line, which displays mean forecasts of the second rating when an alternative is favored, shows a gentle slope from 0 to 2 and a rejection threshold of 3. These findings raise the possibility that guidelines—specifically, the expert use of guidelines consistent with EBDM—may have a de-biasing influence on clinical judgment (Almashat, Ayotte, Edelstein & Margrett, 2008; Wolfson, et al., 2000). The stability of this finding across guidelines, illnesses, and levels of expertise, as well as its implications for education and policy, bear further investigation.

### 4.1 Limitations

The results should be qualified by the study's limitations, which pertain to the subject sample, stimulus, guideline, clinical paradigm, and the experimental procedure. The data were drawn from a small sample of psychiatric trainees at a single and fairly select facility. It cannot be assumed that similar findings would have been obtained from the same study administered at a different facility or if the subjects were experienced clinicians. Nor can the results be generalized to trainees in other disciplines, such as nursing, psychology, or social work. Vignettes are used frequently in studies of clinical decision making and clinical training (Campo et al., 2008; Peabody, Luck, Glassman, Dresselhaus & Lee, 2000). Nonetheless, their use remains controversial, particularly in comparing results with other procedures such as record reviews and standardized patients. A particular concern is whether vignette study data generalize to actual clinical practice (Fihn, 2000).

The Yale Psychiatry Sernyak guideline (YPSA) that was used in this study (Sernyak, et al., 2003) was the precursor to a “fail-first” policy (a requirement that two courses of first-generation treatment be tried before a second-generation treatment can be introduced) that was instituted briefly at the VA Connecticut Healthcare System (Rosenheck, Leslie & Doshi, 2008). The YPSA lacks the broad consensus that is enjoyed by other schizophrenia treatment guidelines, including the APA (Lehman, Lieberman, et al., 2004), the Schizophrenia PORT (Patient Outcomes Research Team: Lehman, Kreyenbuhl,

Figure 2: Mean forecast ratings at each violation point, ratings consistent with preference.



et al., 2004), and the TMAP (Texas Medication Algorithm Project: Moore et al., 2007). Two limitations of the YPSA were noted in previous sections: clozapine, which is the single most effective medication for treatment resistant schizophrenia, is postponed until four other therapies have been tried. There is no mention of injectable treatments, which are recommended for addressing adherence problems. In addition, the YPSA has no provision for so-called “adjunctive” or combination treatments that are commonly used in clinical and community practice. Switch recommendations are based solely on ratings of two items from an established assessment scale (Guy, 1976). These ratings provide very limited information about the patient’s condition and treatment response, and their use as switching criteria has not been tested independently.

The procedure called on subjects to make two sequential forecasts for each vignette. This procedure virtually invited them to use the guideline recommendation in the second rating rather than forecasting without a general likelihood estimate. This procedure did not provide an appropriate test of image theory’s hypothesis that missing information is treated as a violation. However, with a repeated measures design there is no clearly superior alternative. For instance, had subjects been asked for first forecasts of all eight vignettes, then instructed to go through the vignettes again and make second forecasts,

they could draw on memory or believe that the study was testing the consistency of their responses. Similar problems would occur if half of the ratings were made before the guideline was presented. As an alternative, subjects could be asked to rate only one or two vignettes, but given the small subject sample this procedure would have severely limited statistical power. Some of these limitations can be addressed by comparing ratings that rely on different guidelines, or by comparing guideline recommendations against specific alternatives rather than allowing subjects to pose their own.

Forecasts were made by drawing on only three clinical factors. Subjects in the guideline’s dissemination study (Sernyak, et al., 2003) identified these factors as having the greatest influence on their recommendations. However, other clinical and experiential phenomena are important, including patient perceptions of illness, stressors, coping responses, metabolic and other medical complications, and medical and psychiatric co-morbidities. The question for clinical decision makers is, at what point does additional patient-specific information over-fit the data and unduly complicate the process of forecasting? This question can be investigated by studies that vary the type, amount, and quality of information that is included in case summaries.

Perhaps the most significant procedural limitation of the current study is giving subjects information and ask-

ing them only to write their forecasts on paper. In practice, relevant information is elicited through clinical examination and forecasts are discussed with the patient in the course of treatment planning. In this study, as with many others, the communicative aspects of decision making were eliminated in order to focus on the cognitive processes of treatment providers. But in clinical and community practice, EBDM is not the sole province of providers. Active involvement of patients is both inherent and desirable, especially in treatment severe mental illness, where decisions are made progressively over a protracted period and in an evolving system of care (Nielsen, Damkier, Lublin & Taylor, 2011; Pincus et al., 2007).

## 4.2 Conclusion

Studies of medical decision making as a shared activity were occurring long before patient centered care was formally incorporated into healthcare policy (Institute of Medicine, 2001). Early studies displayed strongly differing views about how practitioners should convey expert knowledge to patients, especially in forecasting likelihoods of disease, treatment response, and outcome (see Braddock, Fihn, Levinson, Jonsen & Pearlman, 1997; Greenfield, Kaplan & Ware Jr., 1985; Strull, Lo & Charles, 1984; Vertinsky, Thompson & Uyeno, 1974). The issues have been clarified by recent work that has focused on concepts of numeracy, framing, and format (Gigerenzer & Gray, 2011; Reyna, Nelson, Han & Dieckmann, 2009; Timmermans, Ockhuysen-Vermey & Heneman, 2008). But they continue to overlook that the sole, or even principal, purpose of quantitative forecasts in clinical practice is not prediction. First and foremost, forecasts and the factors that influence them are subjects for discussion. For instance, if poor adherence is diminishing the prospect of a good treatment response, the crucial issues are why this person is not adhering to the regimen and how adherence can be improved. Persons with schizophrenia have their own criteria for gauging the effectiveness of treatment. Whether a progress score of 3 or 4 is less important than whether they can hold a job, keep an apartment, or have a relationship. Whether these aims of treatment are accomplishable, how, and over what period, are what patients want to know when they ask the question, “what are my chances?” An appropriate presentation of quantitative information makes a forecast more understandable. Incorporating this information into a treatment narrative is what makes it meaningful.

Prior to the paper that introduced the simple counting rule, Beach and associates drew a pivotal distinction between “aleatory” (calculated) and “epistemic” reasoning (Beach, Christensen-Szalanski & Barnes, 1987). This distinction became a cornerstone of image theory

and of Beach’s later work in narrative behavioral decision theory (Beach, 2010). The authors portrayed decision making as an epistemic task, that “explicitly involves knowledge about the unique characteristics of specific elements and the framework of knowledge, including the casual network and set of members, in which they are embedded” (p. 147). The counting rule can be narrowly conceived as a smart heuristic, like Gigerenzer’s “tallying rule” (Marewski et al., 2010). More appropriately, its value lies in surmounting the polemic that dominates current discussions about evidence, decision making, and clinical practice. But it has a broader and richer place in the narrative tradition of medical decision making (for instance, Cronje & Fullan, 2003; Epstein & Street, 2011; Greenhalgh, 1999; Kerstholt, van der Zwaard, Bart & Cremers, 2009; Say, Murtagh & Thomson, 2006), where the aim is neither to optimize nor satisfice, but spur an interactive and collaborative effort that determines “the best next thing for this patient at this time” (Weiner, 2004). By interpreting quantitative data in a meaningful and useful way, this process can reach an informed choice.

## References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.
- Alexander, M. (2008). Bias and asymmetric loss in expert forecasts: A study of physician prognostic behavior with respect to patient survival. *Journal of Health Economics*, *27*, 1095–1108.
- Almashat, S., Ayotte, B., Edelstein, B., & Margrett, J. (2008). Framing effect debiasing in medical decision making. *Patient Education and Counseling*, *71*, 102–107.
- Ascher-Svanum, H., Faries, D. E., Zhu, B., Ernst, F. R., Swartz, M. S., & Swanson, J. W. (2006). Medication adherence and long-term functional outcomes in the treatment of schizophrenia in usual care. *Journal of Clinical Psychiatry*, *67*, 453–460.
- Basu, A. (2009). Individualization at the heart of comparative effectiveness research: The time for i-CER has come. *Medical Decision Making*, *29*, NP9–11.
- Beach, L. R. (1997). *The psychology of decision making: People in organizations*. Thousand Oaks, CA: Sage.
- Beach, L. R. (2010). *The psychology of narrative thought: How the stories we tell ourselves shape our lives*. Xlibris.
- Beach, L. R. (Ed.). (1998). *Image theory: Theoretical and empirical foundations*. Mahwah, NJ: Lawrence Erlbaum.

- Beach, L. R., Christensen-Szalanski, J., & Barnes, V. (1987). Assessing human judgment: Has it been done, can it be done, should it be done? In G. Wright & P. Ayton (Eds.), *Judgmental forecasting* (pp. 49–62). New York: John Wiley & Sons.
- Beach, L. R., & Potter, R. E. (1992). The pre-choice screening of options. *Acta Psychologica*, *81*, 115–126.
- Beach, L. R., Puto, C. P., Heckler, S. E., Naylor, G., & Marble, T. A. (1996). Differential versus unit weighting of violations, framing, and the role of probability in image theory's compatibility test. *Organizational Behavior and Human Decision Processes*, *65*, 77–82.
- Beach, L. R., Smith, B., Lundell, J., & Mitchell, T. R. (1988). Image theory: Descriptive sufficiency of a simple rule for the compatibility test. *Journal of Behavioral Decision Making*, *1*, 17–28.
- Beach, L. R., & Strom, E. (1989). A toadstool among the mushrooms: Screening decisions and image theory's compatibility test. *Acta Psychologica*, *72*, 1–12.
- Blumenthal, D. (1996). Quality of care—what is it? *New England Journal of Medicine*, *335*, 891–894.
- Braddock, C. H. 3rd, Fihn, S. D., Levinson, W., Jonsen, A. R., & Pearlman, R. A. (1997). How doctors and patients discuss routine clinical decisions. Informed decision making in the outpatient setting. *Journal of General Internal Medicine*, *12*, 339–345.
- Brenner, H. D., Dencker, S. J., Goldstein, M. J., Hubbard, J. W., Keegan, D. L., Kruger, G., et al. (1990). Defining treatment refractoriness in schizophrenia. *Schizophrenia Bulletin*, *16*, 551–561.
- Campo, A. E., Williams, V., Williams, R. B., Segundo, M. A., Lydston, D., & Weiss, S. M. (2008). Effects of lifeskills training on medical students' performance in dealing with complex clinical cases. *Academic Psychiatry*, *32*, 188–193.
- Chambers, D. (2008). Introduction: Advancing implementation research in mental health. *Administration and Policy in Mental Health and Mental Health Services Research*, *35*, 1–2.
- Cronje, R., & Fullan, A. (2003). Evidence-based medicine: Toward a new definition of 'rational' medicine. *Health*, *7*, 353–369.
- Dickenson, D., & Vineis, P. (2002). Evidence-based medicine and quality of care. *Health Care Analysis*, *10*, 243–259.
- Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, *63*, 568–584.
- Drake, R. E., Goldman, H. H., Leff, H. S., Lehman, A. F., Dixon, L., Mueser, K. T., et al. (2001). Implementing evidence-based practices in routine mental health service settings. *Psychiatric Services*, *52*, 179–182.
- Eddy, D. M. (1990). Practice policies—what are they? *JAMA*, *263*, 877–880.
- Eddy, D. M. (2005). Evidence-based medicine: A unified approach. *Health Affairs*, *24*, 9–17.
- Epstein, R. M., & Street, R. L., Jr. (2011). Shared mind: Communication, decision making, and autonomy in serious illness. *Annals of Family Medicine*, *9*, 454–461.
- Falzer, P. R., & Garman, D. M. (2009). A conditional model of evidence based decision making. *Journal of Evaluation in Clinical Practice*, *15*, 1142–1152.
- Falzer, P. R., & Garman, D. M. (2010). Contextual decision making and the implementation of clinical guidelines: An example from mental health. *Academic Medicine*, *85*, 548–555.
- Falzer, P. R., & Garman, D. M. (2012). Optimizing clozapine through clinical decision making. *Acta Psychiatrica Scandinavica*, *In press*.
- Falzer, P. R., Garman, D. M., & Moore, B. A. (2009). Examining the influence of clinician decision making on adherence to a clinical guideline. *Psychiatric Services*, *60*, 698–701.
- Fayek, M., Flowers, C., Signorelli, D., & Simpson, G. (2003). Psychopharmacology: Underuse of evidence-based treatments in psychiatry. *Psychiatric Services*, *54*, 1453–1456.
- Fihn, S. D. (2000). The quest to quantify quality [editorial]. *JAMA*, *283*, 1740–1742.
- Fiol, G. D., & Haug, P. J. (2009). Classification models for the prediction of clinicians' information needs. *Journal of Biomedical Informatics*, *42*, 82–89.
- Gigerenzer, G., & Gray, J. A. M. (Eds.). (2011). *Better doctors, better patients, better decisions: Envisioning health care 2020*. Cambridge, MA: MIT Press.
- Greenfield, S., Kaplan, S., & Ware Jr., J. E. (1985). Expanding patient involvement in care. *Annals of Internal Medicine*, *102*, 520.
- Greenhalgh, T. (1999). Narrative based medicine: Narrative based medicine in an evidence based world. *British Medical Journal*, *318*, 323–325.
- Grimshaw, J. M., & Eccles, M. P. (2004). Is evidence-based implementation of evidence-based care possible? *Medical Journal of Australia*, *180*(6 Suppl), S50–51.
- Gurmankin Levy, A., & Hershey, J. C. (2006). Distorting the probability of treatment success to justify treatment decisions. *Organizational Behavior and Human Decision Processes*, *101*, 52–58.
- Guy, W. (1976). *ECDEU assessment manual for psychopharmacology*. Rockville, MD: U.S. Department of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, National Institute of Mental Health, Psy-



- chopharmacology Research Branch, Division of Extramural Research Programs.
- Hardin, J. W., & Hilbe, J. M. (2003). *Generalized estimating equations*. Boca Raton, FL: Chapman & Hall/CRC.
- Harvey, N. (2007). Use of heuristics: Insights from forecasting research. *Thinking and Reasoning*, 13, 5–24.
- Helfand, M. (2009). Comparative effectiveness research. *Medical Decision Making*, 29, 641.
- Institute of Medicine. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, D.C.: National Academy Press.
- Kane, J. M. (2004). PORT recommendations. *Schizophrenia Bulletin*, 30, 605–607.
- Kerstholt, J. H., van der Zwaard, F., Bart, H., & Cremers, A. (2009). Construction of health preferences: A comparison of direct value assessment and personal narratives. *Medical Decision Making*, 29, 513–520.
- Klein, G. A., Orasanu, J., Calderwood, R., & Zsombok, C. E. (Eds.). (1993). *Decision making in action: Models and methods*. Norwood, NJ: Ablex Publishing.
- Krizan, Z., & Windschitl, P. D. (2007). The influence of outcome desirability on optimism. *Psychological Bulletin*, 133, 95–121.
- Lehman, A. F., Kreyenbuhl, J., Buchanan, R. W., Dickerson, F. B., Dixon, L. B., Goldberg, R., et al. (2004). The schizophrenia patient outcomes research team (PORT): Updated treatment recommendations 2003. *Schizophrenia Bulletin*, 30, 193–217.
- Lehman, A. F., Lieberman, J. A., Dixon, L. B., McGlashan, T. H., Miller, A. L., Perkins, D. O., et al. (2004). Practice guideline for the treatment of patients with schizophrenia, 2nd ed. *American Journal of Psychiatry*, 161(2, Suppl.), 1–56.
- Lehman, A. F., Steinwachs, D. M., Dixon, L. B., Goldman, H. H., Osher, F., Postrado, L., et al. (1998). Translating research into practice: The schizophrenia patient outcomes research team (PORT) treatment recommendations. *Schizophrenia Bulletin*, 24, 1–10.
- Levy, A. G., & Hershey, J. C. (2008). Value-induced bias in medical decision making. *Medical Decision Making*, 28, 269–276.
- Lipman, T. (2004). The doctor, his patient, and the computerized evidence-based guideline. *Journal of Evaluation in Clinical Practice*, 10, 163–176.
- Lomas, J., Anderson, G. M., Domnick-Pierre, K., Vayda, E., Enkin, M. W., & Hannah, W. J. (1989). Do practice guidelines guide practice? The effect of a consensus statement on the practice of physicians. *New England Journal of Medicine*, 321, 1306–1311.
- Maier, T. (2006). Evidence-based psychiatry: Understanding the limitations of a method. *Journal of Evaluation in Clinical Practice*, 12, 325–329.
- Marewski, J. N., Gaissmaier, W., & Gigerenzer, G. (2010). Good judgments do not require complex cognition. *Cognitive Processing*, 11, 103–121.
- McQueen, D. V. (2002). The evidence debate. *Journal of Epidemiology and Community Health*, 56, 83–84.
- Mistry, H., & Osborn, D. (2011). Underuse of clozapine in treatment-resistant schizophrenia. *Advances in Psychiatric Treatment*, 18, 250–225.
- Moffic, H. S. (2006). Ethical principles for psychiatric administrators: The challenge of formularies. *Psychiatric Quarterly*, 77, 319–327.
- Moore, T. A., Buchanan, R. W., Buckley, P. F., Chiles, J. A., Conley, R. R., Crismon, M. L., et al. (2007). The Texas Medication Algorithm Project antipsychotic algorithm for schizophrenia: 2006 update. *Journal of Clinical Psychiatry*, 68, 1751–1762.
- Nielsen, J., Dahm, M., Lublin, H., & Taylor, D. (2010). Psychiatrists' attitude towards and knowledge of clozapine treatment. *Journal of Psychopharmacology*, 24, 965–971.
- Nielsen, J., Damkier, P., Lublin, H., & Taylor, D. (2011). Optimizing clozapine treatment. *Acta Psychiatrica Scandinavica*, 123, 411–422.
- Parks, J., Radke, A., Parker, G., Foti, M.-E., Eilers, R., Diamond, M., et al. (2009). Principles of antipsychotic prescribing for policy makers, circa 2008. Translating knowledge to promote individualized treatment. *Schizophrenia Bulletin*, 35, 931–936.
- Patel, V. L., Kaufman, D. R., & Arocha, J. F. (2002). Emerging paradigms of cognition in medical decision-making. *Journal of Biomedical Informatics*, 35, 52–75.
- Peabody, J. W., Luck, J., Glassman, P., Dresselhaus, T. R., & Lee, M. (2000). Comparison of vignettes, standardized patients, and chart abstraction: A prospective validation study of 3 methods for measuring quality. *Journal of the American Medical Association*, 283, 1715–1722.
- Pincus, H. A., Page, A. E. K., Druss, B., Appelbaum, P. S., Gottlieb, G., & England, M. J. (2007). Can psychiatry cross the quality chasm? Improving the quality of health care for mental and substance use conditions. *American Journal of Psychiatry*, 164, 712–719.
- Potter, R. E., & Beach, L. R. (1994a). Decision making when the acceptable options become unavailable. *Organizational Behavior and Human Decision Processes*, 57, 468–483.
- Potter, R. E., & Beach, L. R. (1994b). Imperfect information in pre-choice screening of options. *Organizational Behavior and Human Decision Processes*, 59, 313–329.
- Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehen-



- sion and medical decision making. [article.]. *Psychological Bulletin* 135, 943–973.
- Richmond, S. M., Bissell, B. L., & Beach, L. R. (1998). Image theory's compatibility test and evaluations of the status quo. *Organizational Behavior and Human Decision Processes*, 73, 39–53.
- Rosenheck, R. A., Leslie, D. L., Busch, S., Rofman, E. S., & Sernyak, M. (2008). Rethinking antipsychotic formulary policy. *Schizophrenia Bulletin*, 34, 375–380.
- Rosenheck, R. A., Leslie, D. L., & Doshi, J. A. (2008). Second-generation antipsychotics: Cost-effectiveness, policy options, and political decision making. *Psychiatric Services*, 59, 515–520.
- Ruscio, A. M., & Holohan, D. R. (2006). Applying empirically supported treatments to complex cases: Ethical, empirical, and practical considerations. *Clinical Psychology: Science and Practice*, 13, 146–162.
- Satterfield, J., Spring, B., Brownson, R. C., Mullen, E. J., Newhouse, R. P., Walker, B. B., et al. (2009). Toward a transdisciplinary model of evidence-based practice. *Milbank Quarterly*, 87, 368–390.
- Say, R., Murtagh, M., & Thomson, R. (2006). Patients' preference for involvement in medical decision making: A narrative review. *Patient Education and Counseling*, 60, 102–114.
- Sernyak, M. J., Dausey, D., Desai, R., & Rosenheck, R. A. (2003). Prescribers' nonadherence to treatment guidelines for schizophrenia when prescribing neuroleptics. *Psychiatric Services*, 54, 246–248.
- SPSS Inc. (2011). SPSS advanced 19 for windows. Chicago: SPSS Corporation.
- Strull, W. M., Lo, B., & Charles, G. (1984). Do patients want to participate in medical decision making? *Journal of the American Medical Association*, 252, 2990–2994.
- Timmermans, D. R. M., Ockhuysen-Vermey, C. F., & Henneman, L. (2008). Presenting health risk information in different formats: The effect on participants' cognitive and emotional evaluation and decisions. *Patient Education and Counseling*, 73, 443–447.
- Van Zee, E. H., Paluchowski, T. F., & Beach, L. R. (1992). The effects of screening and task partitioning upon evaluations of decision options. *Journal of Behavioral Decision Making*, 5, 1–19.
- Vertinsky, I. B., Thompson, W. A., & Uyeno, D. (1974). Measuring consumer desire for participation in clinical decision making. *Health Services Research*, 9, 121–134.
- Visweswaran, S., Angus, D. C., Hsieh, M., Weissfeld, L., Yealy, D., & Cooper, G. F. (2010). Learning patient-specific predictive models from clinical data. *Journal of Biomedical Informatics*, 43, 669–685.
- Weiner, S. J. (2004). Contextualizing medical decisions to individualize care: Lessons from the qualitative sciences. *Journal of General Internal Medicine*, 19, 281–285.
- Weisz, G., Cambrosio, A., Keating, P., Knaapen, L., Schlich, T., & Tournay, V. J. (2007). The emergence of clinical practice guidelines. *The Milbank Quarterly*, 85, 691–727.
- Wolfson, A. M., Doctor, J. N., & Burns, S. P. (2000). Clinician judgments of functional outcomes: How bias and perceived accuracy affect rating. *Archives of Physical Medicine and Rehabilitation*, 81, 1567–1574.
- Woolf, A. D. (1997). Introduction: How does evidence that is available affect decisions with an individual patient? *Bailliere's Clinical Rheumatology*, 11, 1–12.
- Woolf, S. H. (1990). Practice guidelines: A new reality in medicine: I. Recent developments. *Archives of Internal Medicine*, 150, 1811–1818.
- Woolf, S. H. (1993). Practice guidelines: A new reality in medicine. Iii. Impact on patient care. *Archives of Internal Medicine*, 153, 2646–2655.
- Zerhouni, E. (2003). Medicine: The NIH roadmap. *Science*, 302, 63–72.