Law and Human Behavior, Vol. 29, No. 2, April 2005 (© 2005)
DOI: 10.1007/s10979-005-3401-z

Gender and Risk Assessment Accuracy: Underestimating Women's Violence Potential

Jennifer Skeem,[1,7] **Carol Schubert,**[2] **Stephanie Stowman,**[3] **Stacy Beeson,**[3] **Edward Mulvey,**[4] **William Gardner,**[5] **and Charles Lidz**[6]

Understanding factors that contribute to mental health professionals' (MHPs') accuracy in assessing patients risk of violence can inform efforts to improve accuracy and to integrate risk assessment technology with practice. Based on a sample of 147 clinicians who assessed 680 patients in a psychiatric emergency room, this study investigates the influence of patient gender, MHP gender, and their potential interaction on MHPS' risk assessment accuracy. The results indicate that MHPs of both genders are particularly limited in their ability to assess female patients' risk of future violence. This finding was not limited to a particular professional group and was not attributable to gender-related differences in violence. Implications for future research on the judgment processes that may underlie MHPs' limited accuracy with women and for training programs in violence risk assessment are discussed.

KEY WORDS: mental illness; gender; violence; risk assessment.

Violence risk assessment has become an essential aspect of professional practice for mental health professionals (MHPs). These assessments are required in a variety of criminal and civil contexts, ranging from decisions about sentencing to admission to mental health facilities. Currently, however, MHPs are most likely to assess risk to fulfill their responsibilities to protect third parties while the MHPs are working with potentially violent patients in the community (Monahan, 1996; Skeem & Mulvey, 2002).

Errors in such risk assessments have serious implications for the patient and the public. When MHPs incorrectly deem patients to be at risk for violence (that is, they

[1]University of California, Irvine.
[2]University of Pittsburgh Medical Center, Pennsylvania.
[3]University of Nevada, Las Vegas, Nevada.
[4]University of Pittsburgh Medical Center, Pennsylvania.
[5]Ohio State University, Ohio.
[6]University of Massachusetts Medical Center, Massachusetts.
[7]To whom correspondence should be addressed at Psychology & Social Behavior, 3311 Social Ecology II, University of California, Irvine, Irvine, California 92697-7085; e-mail: skeem@uci.edu.

make false positive errors), the consequences for the patient may include involuntary commitment, mandated outpatient treatment, forced use of psychotropic medication, the loss of important civil rights (Monahan et al., 2001; Slovenko, 1992), and damage to the therapeutic relationship between the patient and the MHP (see Bloch & Chodoff, 1991). Conversely, when MHPs fail to identify patients at risk (that is, they make false negative errors), patients may harm themselves, their loved ones, or members of the general public (Council of State Governments, 2002). Given the costs of these errors, researchers have made great efforts to improve MHPs' accuracy in forecasting patients' violence.

Historically, MHPs have not been viewed as being particularly skilled in this task (Borum, 1996; Dershowitz, 1969). In fact, early studies suggested that MHPs were accurate for only one of three cases at best (Monahan, 1981). More recent and methodologically sound studies have shown that MHPs possess at least modest rates of accuracy (Lidz, Mulvey, & Gardner, 1993; McNiel, Sandberg, & Binder, 1998; Otto, 1992, 1994; Borum, 1996). Nevertheless, the field's attention has largely shifted from studying MHPs' accuracy to developing violence risk assessment tools to improve it. This shift reflects a growing consensus that *unaided* clinical judgment often performs more poorly than clinical judgment that is structured by, or even replaced with, risk assessment tools (Dawes, Faust, & Meehl, 1989; Douglas & Skeem, in press; Gardner, Lidz, Mulvey, & Shaw, 1996; Grove & Meehl, 1996; Meehl, 1954; McNiel, Binder, & Greenfield, 1988; compare Litwack, 2001).

The field's current dialogue about how narrow a role clinical judgment should play in violence risk assessment assumes, perhaps prematurely, that the door is closed on unaided clinical judgment (Litwack, 2001). There are three reasons to continue to use unaided clinical judgments. First, the performance of actuarial assessments may not be as far superior to clinical assessments as is typically assumed (see Grove, Zald, Lebos, Snitz, & Nelson, 2000). Second, unstructured assessments probably are the most common grounds for making decisions about violence risk today and are likely to remain so, unless or until barriers to transferring risk assessment technology to real world settings are addressed (Borum, 1996; Elbogen, 2002). Third, understanding clinical judgment will remain important even after technology transfer takes place. This is true both for risk assessment tools that (a) structure (rather than replace) that judgment (Monahan, 1997; Webster, 1994), and (b) replace judgment, but necessarily rely upon MHPs to evaluate dynamic predictors (Douglas & Skeem, 2004; Monahan & Steadman, 1994; Ward & Eccleston, 2000), and to communicate risk information to others involved in the patient's care (Heilbrun, O'Neill, Strohman, Bowman, & Philipson, 2000). Improving understanding of clinical judgment is therefore not only reasonable, but is actually a necessary step toward effectively integrating risk assessment technology with contemporary practice.

This article enriches the picture of the factors that contribute to MHPs' accuracy in violence risk assessment. Only a few investigators have gone beyond assessing MHPs' basic accuracy to identify factors associated with accurate and inaccurate risk assessments (for example, McNeil, & Binder, 1991, 1995). One key predictor of accuracy that emerges consistently, however, is patient gender (see Spector, 2001; Blashfield & Herkov, 1996). Although MHPs' overall risk assessment accuracy is

only modest, their accuracy with women is worse, chiefly because they underestimate the likelihood that female patients will become involved in violence (Coontz, Lidz, & Mulvey, 1994; Elbogen, Williams, Kim, Tomkins, & Scaolora, 2001; Lidz et al., 1993; McNeil & Binder, 1995). For example, Nicholls, Ledwidge, and Ogloff (2003) examined psychiatrists' accuracy in assessing 63 psychiatric patients' risk of physical violence (rated as low, moderate, or high) within 1 year of assessment. During that period, 23% of male and 38% of female inpatients engaged in violence. The probability that a psychiatrist had rated a man involved in violence as being a moderate to high risk was low (.43), nevertheless it was nearly twice the probability that this rating had been given to a woman involved in violence (.25).

Despite relatively consistent findings that patient gender affects accuracy, the effects of MHPs' own gender on accuracy in assessing risk is unclear. Coontz and her colleagues (Coontz et al., 1994) examined transcripts of MHP interactions with 62 individuals who were referred to a psychiatric emergency room within 72 hr of involvement in violence. The authors found that MHPs' gender did not predict the amount of time MHPs discussed violence potential with patients, although MHPs were more likely to discuss the issue of violence with men than women. In contrast, Elbogen et al. (2001) found that MHPs' ($N = 81$) gender interacted with patient gender in predicting the accuracy of violence risk assessments. MHPs had been asked to select eight of their recent patients and rate their likelihood of future violence. Although MHPs of both genders rated women as similarly likely to be violent, female MHPs rated men as more dangerous than did male MHPs. Female MHPs also perceived a larger "gender gap" in violence than male MHPs, in that, their ratings of dangerousness for men were significantly higher than those for women.

The influence of patient gender, MHP gender, and their potential interaction on MHPs' risk assessments has important implications. To the extent that MHPs inaccurately perceive female patients as disproportionately safe, female patients are at risk for inadequate supervision and intervention (Binder & McNeil, 1990). Given findings that male and female patients are at equal risk for future violence, patient- and/or MHP- gender judgment biases should be addressed through training (Borum, 1996; Elbogen et al., 2001).

The present study was designed to increase understanding of the extent to which MHPs' accuracy in assessing violence risk varies with patient and MHP gender. Based on data from one of the few large-scale, methodologically sound assessments of MHPs' accuracy in risk assessment (Lidz et al., 1993), we assess the relations between accuracy and patient gender, MHP gender, and the patient by MHP gender interaction.

METHOD

Lidz et al. (1993) interviewed patients who presented at a psychiatric emergency room, as well as the MHPs who assessed them. After these patients were discharged, the investigators attempted to locate them in the community and interview both patients and collateral informants (i.e., a family member, significant other, or close friend who knew them well) three times during a 6-month period.

These interviews were supplemented by reviews of official records. Baseline interviews focused on MHPs' assessments of patients' violence potential, whereas follow-up interviews focused on patients' involvement in violence. To ensure that the data accurately reflected whether at least one violent incident had occurred during the follow-up, only cases with at least two patient follow-up interviews or three follow-up interviews, including at least one with the patient, were included for analysis.

The emergency room had a team-based sequence for processing patients, in which (a) a staff clinician interviewed the patient in detail, (b) the attending psychiatrist briefly "re-interviewed" the patient, and (c) the clinician and attending came to a case disposition. After a disposition was determined, investigators asked the clinician and attending to independently rate their degree of concern that the patient might be violent toward others in the next 6 months (on a six-point scale, ranging from 0 = *no concern* to 5 = *great concern*). Lidz et al. (1993) summed these ratings to generate a total concern score (ranging from 0 to 10) and used this score to create two groups of patients: "predicted violent" (total concern ≥ 3) and "predicted nonviolent" (total concern = 0). (The group of patients with total concern scores of 1 and 2 were not included in analyses.) Each "predicted violent" patient was matched with a "predicted nonviolent" patient on age, sex, race, and admission status (admitted/not admitted) to prevent MHPs from obtaining high rates of predictive accuracy based on demographic differences. The original sample consisted of 357 matched pairs ($N = 714$).

To address the present study's aim, the original data set was reconfigured. Specifically, clinicians' and attendings' independent ratings of concern about future violence were extracted to permit an assessment of the relation between predictive accuracy and gender (of MHP and patient). The use of MHPs' individual ratings destroyed the original matching system, given that patients had been matched using summed, team-based scores.[8]

Participants

Patients

Clinicians provided ratings of concern for 680 cases and attendings provided ratings of concern for 667 cases. These are fewer cases than those described in the study presented by Lidz et al. (1993, $N = 714$).[9] Based on hospital procedure, clinicians and attendings typically saw the same patients (approximately 90% of the

[8] Although the use of individual ratings destroyed the original matching system, cases classified as "predicted violent" and "predicted nonviolent" based on individual ratings did not differ on the original study's matching variables (gender, race, age, and admission status). This was true at both the univariate (i.e., *t*- and chi-square tests) and multivariate (i.e., discriminant function analysis [DFA]) level, for both clinicians' and attendings' predictions.

[9] This differences reflects differences in the units of study: individual ratings of concern (in the present study) or team-based ratings of concern (in Lidz et al., 1993). For "predicted nonviolent" cases, the number of team-based ratings is equal to or greater than the number of individual ratings. (The numbers are equal if both MHPs for the team-based ratings had "zero" levels of concern, but team-based ratings exceed individual ratings when a single MHP was counted as a "team" in the original study.) For "predicted violent" cases, the number of team-based ratings is greater than the number of individual ratings. (For team-based ratings, only one MHP had to have significant concern to qualify the case

time). Thus, to simplify the presentation of results, only clinicians' data are reported here (attendings' results are reported when they differ substantively from those reported for clinicians). Based on clinicians' 680 cases, the current study's sample consisted of relatively young ($M = 28.3$ years, $SD = 11$), chiefly high-school educated ($M = 12.2$ years, $SD = 3.2$), male (60.3%) and female (39.7%) participants who were equally likely to be White (52.2%) or African-American (47.8%). Research participants' most common diagnosis was substance abuse disorder (33.1%), although diagnoses of affective disorder (19.6%), personality (17.0%), schizophrenia or schizoaffective (14.2%), and other (16.2%) disorders were also common. Relatively few were diagnosed *only* with substance abuse (2.6%) or personality (7.0%) disorders. Instead, 30.5% of participants were diagnosed with a comorbid substance abuse and Axis I mental disorder, and 10.0% were diagnosed with a comorbid personality and Axis I mental disorder. According to hospital records, over one-quarter (26.2%) of participants had a history of at least one arrest.

Mental Health Professionals

Of the 147 clinicians who assessed risk, 37% were junior resident psychiatrists, 24% were master's-level clinicians, 20% were nurse-clinicians, 12% were doctoral psychology interns, and 7% were licensed social workers. Sixty percent were women and 40% were men. The distribution of ratings across clinicians was examined to see if the results might be affected by a small number of clinicians performing a large proportion of the assessments. Although cases were distributed relatively evenly among clinicians, four outlying clinicians assessed 27–32 cases each (nearly 5% of total assessments).

Of the 62 attendings who assessed risk in this study, 65% were attending psychiatrists and 35% were senior resident psychiatrists. Nevertheless, attending psychiatrists completed the vast majority (95%) of assessments. Most attendings (87%) were men. Three outlying attending psychiatrists assessed 50–79 cases each. Analyses were conducted that omitted these outlying psychiatrists and clinicians to determine whether they unduly affected the results (see below).

Measures

Predictions

Each MHPs' rating of their degree of concern about violence within the next 6 months was left in its original scaled form (0 = *no concern* to 5 = *great concern*) and collapsed into a nominal category as well: "predicted violent" (concern ≥ 3) and "predicted nonviolent" (concern = 0).

Violence

Patient violence was measured based on information from patient and collateral follow-up interviews as well as reviews of records. To elicit information about

for the "predicted violent" sample. For individual ratings, each MHP must have significant concern to qualify the case as "predicted violent.")

violent incidents, interviewers administered an expanded version of the Conflict Tactics Scale (Straus & Gelles, 1990), which assesses multiple categories of specific acts of aggression ranging from pushing, grabbing or shoving, to using a weapon on someone (Lidz et al., 1993). When patients had engaged in violence, interviewers asked probe questions to determine the nature of the violent incident (e.g., the degree of co-combatant injury).

A patient was judged to have been involved in a *violent* incident if the patient, collateral, or an official record reported that the patient had laid hands on another person with intent to harm him or her, or had threatened someone with a weapon in hand. Verbal threats, incidents in which the patient was the victim of violence (e.g., the victim of a mugging) or did not engage in any retaliatory violent act, and parental discipline were excluded. A patient was judged to have been involved in a *serious violent* incident if the violent incident included rape, threats with a weapon in hand, use of a weapon, injuries that required medical attention, and attempted homicide. The measures of *violence* and *serious violence* were dichotomous, and reflected whether or not the patient had been involved in at least one incident during the full, 6-month follow-up period. Only violent incidents that occurred in the community (rather than closed treatment settings) were considered in these analyses. Of clinicians' 680 cases, 45% were involved in some form of violence, and 20% were involved in one or more serious violent incidents.

RESULTS

The chief aim of this study was to assess the relations among patient gender, MHP gender, and risk assessment accuracy. To help interpret the results of our main analyses, we began by assessing the basic accuracy of MHPs' individual concerns. In this section, MHPs' basic accuracy is presented, followed by its relation to gender. Only clinicians' results are considered to simplify the presentation.

Accuracy of Individual Concerns

We first assessed the basic accuracy of MHPs' nominal concerns ("predicted violent" vs. "predicted nonviolent") in predicting patience violence. As shown in Table 1, clinicians' nominal concerns were weakly ($\phi = .20$), but significantly predictive of patient violence, $\chi^2(1, N = 605) = 24.65$, $p < .001$. Notably, the finding that individual MHPs have modest predictive power is consistent with that found for MHPs functioning as teams (see Lidz et al., 1993).

Table 1. Clinicians' Basic Predictive Accuracy

	Not violent	Violent	Total
Predicted nonviolent	227	129	356
Predicted violent	108	141	249
Total	335	270	605

Sensitivity = .52. Specificity = .68. Positive predictive power = .57. Negative predictive power = .64.

Litwack and Schlesinger (1999) expressed concern that using dichotomous predictions of violence, as opposed to scaled predictions, may underestimate the accuracy of MHPs' predictions. Therefore, we also analyzed the accuracy of MHPs' predictions in the 0–5 scale using ROC curve methods. The results were consistent with those found using dichotomized predictions. Using Swets (1988) descriptive classifications, the accuracy of MHPs' predictions was poor (the area under the ROC curve was only 61%). The cutoff that balanced false negative and false positive errors (2) was similar to the value that we used to dichotomize the predictions. Therefore, for the remainder of the paper we consider "violent/non-violent" predictions only. Notably, a subset of analyses completed with scaled predictions produced findings parallel to those described below.

Relation Between Accuracy and Gender

The second aim of the study was to assess the relation between predictive accuracy and patient gender, MHP gender, and the patient gender × MHP gender interaction. Hierarchical loglinear analyses were performed to assess the relationship among four dichotomous variables: patient gender, MHP gender, MHPs' predictions, and whether the patient was violent during the follow-up. For these analyses, three interactions were of particular interest. The first was the four-way interaction (MHP prediction × patient gender × MHP gender × violence). If this interaction were significant, it would mean that predictive accuracy varied, depending upon specific MHP gender and patient gender combinations (that MHPs either were more accurate at predicting same gender patient violence or opposite gender patient violence). The second was a three-way interaction (MHP prediction × patient gender × violence), which, if significant, would indicate that predictive accuracy was associated with patient gender. The third was a three-way interaction (MHP prediction × MHP gender × violence), which, if significant, would indicate that predictive accuracy was associated with MHP gender.

In the first analysis, all variables were entered directly to test a saturated model (that included all main effects and all interactions). As shown in Table 2, the results of this conservative analysis indicated that patient gender, but neither MHP gender nor its interaction with patient gender, were associated with predictive accuracy. Specifically, the three-way interaction of interest that included patient gender was significant, but the three-way interaction of interest that included MHP gender, and the four-way interaction were not significant. The results indicate that MHPs' predictive accuracy varied significantly with patient gender, even when controlling for significant lower-order associations that reflected MHPs' basic accuracy (MHP prediction x violence) and women's greater involvement in violence (patient gender × violence).

In the second analysis, all variables were entered stepwise via backward deletion of effects to build a "best-fitting" model by statistical criteria. The substantive results of this analysis, particularly with respect to the three-way and four-way interactions of interest, were consistent with those reported above. The final model included only the main effect of MHP gender ($\Delta\chi^2[1, 600] = 23.39$, $p < .001$) and the three-way interaction among MHPS' predictions, patient gender, and patient

Table 2. Predictive Accuracy and Gender: Loglinear Results for a Saturated Model

Effect	χ^2
Main effects	
Clinician	23.35***
Patient	24.16***
Prediction	18.11***
Violence	6.01**
Two-way interactions	
Patient × clinician	2.33
Patient × prediction	.11
Clinician × prediction	1.14
Clinician × violence	.81
Patient × violence	4.30*
Prediction × violence	24.45*
Three-way interactions	
Patient × clinician × prediction	2.88
Patient × clinician × violence	.22
Clinician × prediction × violence	.20
Patient × prediction × violence	3.86*
Four-way interaction	.27

Clinician: clinician gender; Patient: patient gender.
*$p < .05$; **$p < .01$; ***$p < .001$.

violence ($\Delta\chi^2[1, 600] = 4.14, p < .05$). As before, the three-way interaction among MHPs' predictions, MHP gender, and patient violence was not significant, nor was the single four-way interaction. Notably, in these analyses, cell sizes were adequate and there was sufficient power to detect at least a medium effect (Cohen, 1988, 1992).

These multivariate results were consistent with the results of univariate tests. As shown in Table 3, the association between patient gender and predictive accuracy (accurate/inaccurate) was weak, but significant ($\phi = .10, p < .05$). Table 3 presents the sensitivity, specificity, negative predictive power (NPP) and positive predictive power (PPP) of clinicians' predictions for women and men. Clinicians' predictions were poorer for women than men, although the difference was statistically significant ($p < .05$) only for negative predictive power.

Table 3. Clinicians' Predictive Accuracy and Patient Gender

		Not violent	Violent	Total
Women	Predicted nonviolent	79	64	143
	Predicted violent	45	56	101
	Total	124	120	244
Men	Predicted nonviolent	148	65	213
	Predicted violent	63	85	148
	Total	211	150	361

Sensitivity for women vs. men = .47 vs. .57. Specificity for women vs. men = .64 vs. .70. Positive predictive power for women v. men= .55 vs. .57. Negative predictive power for women vs. men= .55 vs. .70.

Ruling Out Alternative Hypotheses

In summary, these results suggest that MHPs are less accurate at predicting violence involving women patients, and that MHP gender and its interaction with patient gender have little effect on predictive accuracy. We completed several final analyses to rule out possible methodological explanations for these findings.

Are Predictions More Accurate for Serious Violence?

One might argue that the modest rates of basic predictive efficiency found here reflect the fact that MHPs do not view minor patient violence as worthy of concern. Because MHPs focus narrowly on serious patient violence, only serious violence should be considered when assessing accuracy. We therefore completed the analyses of predictive utility using serious violence as the criterion. These analyses did not reveal greater predictive accuracy than those that focused on violence in general. For example, ROC analyses of clinicians scaled concerns in relation to serious violence produced an AUC of .63 ($SE = .03$) for serious violence, which is highly similar to the AUC of .61 ($SE = .02$) found for violence overall.

Another issue is whether MHPs' have poorer predictive accuracy for women because men are involved in more serious violence. This argument is wrong in its premise: there was no significant relation between patient gender and the occurrence of serious violence, $\chi^2(1) = 0.05$, *ns.*

Does Predictive Accuracy Vary Substantially Across MHP Subgroups?

As noted earlier, the results reported above focus on clinicians rather than attendings. This is because results were highly similar for the two groups. For example, ROC analyses of attendings' scaled concerns in relation to violence produced an area under the curve of .59 ($SE = .02$), which is only 2% lower than that for clinicians. As with clinicians, loglinear analyses revealed no significant association between predictive accuracy and attending gender, nor an attending by patient gender interaction. Relative to clinicians, there were less pronounced gender-related differences in accuracy for attendings: their predictions were accurate for 45% of women and 62% of men, ($\chi^2[1, 561] = 2.63$, $p < .10$). Nevertheless, as with clinicians, attendings' false negative rate for women (29.4%) was substantially higher than that for men (16.0%). In summary, the key findings seem applicable to both clinicians and attendings.

An additional concern is whether the results reflect the influence of a few individuals who evaluated an extreme number of cases in the emergency room. To examine this possibility, analyses parallel to those above were completed, omitting these "extreme" individuals. The results obtained were similar to those reported above. For example, an ROC analysis completed without the four outlying clinicians who completed 27–32 cases each produced an area of the curve of .61 ($SE = .02$), identical to that obtained with the entire sample.

How Might Dependent Samples Affect Predictive Accuracy?

As noted earlier, although the use of individual rather than team-based predictions destroyed the original study's matched design, univariate and multivariate analyses indicated that the individually-based "predicted violent" and "predicted nonviolent" cases were still similar on all matching variables (age, sex, race, and admission status). To ensure that the results obtained were not due to dependency between the originally matched samples, we completed analyses of MHPs' predictions using statistics appropriate for nonindependent samples (for example, McNemar's test of dependent sample proportions). The results obtained were parallel to those reported above.

DISCUSSION

In contemporary clinical settings, the prototypic violence risk assessment is completed by an MHP who chiefly relies on his or her professional judgment. This study was among the first to systematically assess the relation between violence risk assessment accuracy and MHP gender, patient gender, and their interaction. The results indicate that, as others had found (Coontz et al., 1994; Elbogen et al., 2001; McNeil & Binder, 1995; Nicholls, Ledwidge, & Ogloff, 2003) MHPs underestimate the risk of future violence in female psychiatric patients. We also found that male and female MHPs are equally poor at assessing women's violence potential.

We tested a number of alternative hypotheses for our finding that MHPs were poorer in assessing womens' risk of future violence than mens.' We found that the results were not limited to a particular professional group (they applied to clinicians and attending psychiatrists; and MHPS who saw more and fewer patients). Similarly, the results could not be attributed to any gender-related differences in seriousness of the violence, because women in this study were as likely to be involved in serious violent as men.

MHP's underestimation of violence in women may be based partially on two factors. First, in society as a whole, women have a lower rate of violence than men (Litwack & Schlesinger, 1999; Robbins, Monahan, & Silver, 2003), and MHPs may adjust their risk assessments accordingly. Substantial evidence, however, suggests that the gender gap in violence is absent in psychiatric populations (Elbogen et al., 2001; Lamb, McNiel, & Binder, 2000; Newhill, Mulvey, & Lidz, 1995). Based on data for 1,136 psychiatric patients involved in the MacArthur Violence Risk Assessment Study, Robbins et al. (2003) found that the prevalence rate of "violence or other aggressive acts" during the first year after discharge was essentially the same for men (59.8%) and women (61.6%). Second, MHPs' underestimation of women's violence potential may reflect the fact that women's involvement in violence is less visible than men's, since it occurs disproportionately in the home with family members (Hiday, Swartz, Swanson, Borum, & Wagner, 1998; Kruttschnitt, Gartner, & Ferraro, 2002; Newhill, Mulvey, & Lidz, 1995; Robbins et al., 2003). Robbins and her colleagues (2003) found that women's co-combatants were significantly more likely than men's to be family members, and men's co-combatants were significantly

more likely to be friends, acquaintances, or strangers. Violence and other aggressive acts were more likely for women than men to occur in the home.

Despite the clear finding that risk assessment accuracy was associated with patient gender, we found no relation between risk assessment accuracy and MHP gender or the interaction of MHP gender with patient gender. Because few studies have addressed this issue, and the results of these studies are mixed (compare Coontz et al., 1994; Elbogen et al., 2001), it is difficult to contextualize this finding. Nevertheless, the results suggest that MHPs' own gender may have little to do with their accuracy in assessing men and women's risk. Male and female MHPs seem to share whatever judgment processes there are that limit their ability to assess women's violence potential.

A secondary finding of this study is that the results on MHPs' accuracy originally presented by Lidz and his colleagues (1993) were not an artifact of the use of team-based predictions or a decision to dichotomize MHPs' predictions (see Litwack & Schlesinger, 1999). The AUC for clinicians' ratings of concern suggested there was only a 61% chance that a patient who later became violent elicited more concern from clinicians than a randomly chosen patient who did not become violent. Clinicians' poor levels of accuracy are lower than, but comparable to, the fair-good levels typically found for the violence risk assessment tools (approximately .76–.80; for example, Douglas, Ogloff, Nicholls, & Grant, 1999; Monahan et al., 2001; Quinsey, Harris, Rice, Cormier, 1998).

Limitations and Implications

The strength of this study lies in its large sample size, strong design for testing risk assessment accuracy, and provision of data to thoroughly test the specific aims against competing hypotheses. Although there was no effort in this study to correct estimates of the patient's community violence for the amount of time they actually spent in the community during the 6-month follow-up period, all cases (regardless of admission status) were followed from the time they left the hospital for the community.

This study has strong implications both for future research and clinical practice. First, although this study clearly suggests that MHPs underestimate women's violence potential, it sheds little light on why they do so. As suggested earlier, we expect that MHPs are mistaken about female patients' baserates of violence and serious violence. In a related sense, MHPs may weight the utilities of potential errors in prediction differently for men and women. To the extent that they are less concerned about false negative errors for women than men, they may less often deem women as at risk for future violence. A third possibility is that MHPs are limited in their ability to predict the *type* of violence in which women are involved (regardless of whether a man or woman is involved in this type of violence). Determining whether MHPs underestimate women's risk of violence, weight errors of prediction with women differently, or have a "blind spot" for the type of violence in which woman usually engage will have direct implications for developing interventions that might best improve MHPs risk assessments with women. Such interventions could focus on informing MHPs about the nearly equal baserates of violence for

men and women, or might emphasize a particular form of violence that often goes undetected by MHPs (and perhaps even risk assessment tools).

Improving MHPs risk assessments may, in turn, help to better address the needs of a potentially underserved population. If women are as likely to be violent (and seriously violent) as men following discharge, but are not identified as high risk cases, they will not be provided with services designed to manage risk. The determination of gender-based bias in judgment is thus linked to the larger inquiry of potential gender bias in service provision or the application of mandated treatment.

ACKNOWLEDGMENT

Data collection for this project was supported by the National Institute of Mental Health (grant #R01-MH400030). We would like to thank Candice Odgers for comments on an earlier draft of this paper.

REFERENCES

Binder, R. L., & McNiel, D. E. (1990). The relationship of gender to violent behavior in acutely disturbed psychiatric patients. *Journal of Clinical Psychiatry*, *51,* 110–114.

Blashfield, R. K., & Herkov, M. J. (1996). Investigating clinician adherence to diagnosis by criteria: A replication of Morey and Ochoa (1989). *Journal of Personality Disorders*, *10*(3), 219–228.

Bloch, S., & Chodoff, P. (Eds.). (1991). *Psychiatric ethics*. London: Oxford University Press.

Borum, R. (1996). Improving the clinical practice of violence risk assessment: Technology, guidelines, and training. *American Psychologist*, *51*(9), 945–956.

Coontz, P. D., Lidz, C. W., & Mulvey, E. P. (1994). Gender and the assessment of dangerousness in the psychiatric emergency room. *International Journal of Law and Psychiatry*, *17*(4), 369–376.

Cohen, J. (1988). Set correlation and contingency tables. *Applied Psychological Measurement*, *12*(4), 425–434.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Council of State Governments (2002). *Criminal Justice/Mental Health Consensus Project.* Retrieved August 15, 2002, from http://www.consensusproject.org

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243,* 1668–1674.

Dershowitz, A. M. (1969). The psychiatrist's power in civil commitment: A knife that cuts both ways. *Psychology Today*, *2*(9), 43–47.

Douglas, K., & Skeem, J. (in press). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law*.

Douglas, K. S., Ogloff, J. R., Nicholls, T. L., & Grant, I. (1999). Assessing risk for violence among psychiatric patients: The HCR-20 violence risk assessment scheme and the psychopathy checklist: Screening version. *Journal of Consulting and Clinical Psychology*, *67*(6), 917–930.

Elbogen, E. (2002) The process of violence risk assessment: A review of descriptive research. *Aggression and Violent Behavior*, *7,* 591–604.

Elbogen, E. B., Williams, A. L., Kim, D., Tomkins, A. J., & Scalora, M. J. (2001). Gender and perceptions of dangerousness in civil psychiatric patients. *Legal and Criminological Psychology*, *6,* 215–228.

Gardner, W., Lidz, C. W., Mulvey, E. P., & Shaw, E. C. (1996). A comparison of actuarial methods for identifying repetitively violent patients with mental illnesses. *Law and Human Behavior*, *20*(1), 35–48.

Grove, W. M., Zald, D. H., Lebos, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*(1), 19–30.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy and Law*, *2*(2), 293–323.

Heilbrun, K., O'Neill, M. L., Strohman, L. K., Bowman, Q., & Philipson, J. (2000). Expert approaches to communicating violence risk. *Law and Human Behavior*, *24*(1), 137–148.
Hiday, V. A., Swartz, M. S., Swanson, J. W., Borum, R., & Wagner, H. R. (1998). Male-female differences in the setting and construction of violence among people with severe mental illness. *Social Psychiatry and Psychiatric Epidemiology*, *33*(Suppl 1), S68–S74.
Kruttschnitt, C., Gartner, R., & Ferraro, K. (2002). Women's involvement in serious interpersonal violence. *Aggression and Violent Behavior*, *7,* 529–565.
Lam, J. N., McNeil, D. E., & Binder, R. L. (2000). The relationship between patients' gender and violence leading to staff injuries. *Psychiatric Services*, *51*(9), 1167–1170.
Lidz, C., Mulvey, E., & Gardner, W. (1993). The accuracy of predictions of violence to others. *Journal of the American Medical Association*, *269,* 1007–1011.
Litwack, T. R. (2001). Actuarial versus clinical assessments of dangerousness. *Psychology, Public Policy, and Law*, *7*(2), 409–443.
Litwack, T. R., & Schlesinger, L. B. (1999). Dangerousness risk assessments: Research, legal, and clinical considerations. In A. Hess & I. Weiner (Eds.), *The handbook of forensic psychology* (pp. 171–217). New York: Wiley.
McNiel, D. E., & Binder, R. L. (1991). Clinical assessment of the risk of violence among psychiatric inpatients. *American Journal of Psychiatry*, *148,* 1317–1321.
McNiel, D. E., & Binder, R. L. (1995). Correlates of accuracy in the assessment of psychiatric inpatients' risk of violence. *American Journal of Psychiatry*, *152*(6), 901–906.
McNiel, D. E., Binder, R. L., & Greenfield, T. K. (1988). Predictors of violence in civilly committed acute psychiatric patients. *American Journal of Psychiatry*, *145*(8), 965–970.
McNiel, D. E., Sandberg, D. A., & Binder, R. L. (1998). The relationship between confidence and accuracy in clinical assessment of psychiatric patients' potential for violence. *Law & Human Behavior*, *22*(6), 655–669.
Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the literature*. Minneapolis, MN: University of Minnesota Press.
Monahan, J. (1981). The clinical prediction of violent behavior. *Crime & delinquency issues: A monograph series,* ADM 81-921. Rockville, MD: DDHS.
Monahan, J. (1996). Violence prediction: The past twenty and the next twenty years. *Criminal Justice and Behavior*, *23*(1), 107–120.
Monahan, J. (1997). Actuarial support of the clinical assessment of violence risk. *International Review of Psychiatry*, *9,* 167–169.
Monahan, J., & Steadman, H. J. (Eds.). (1994). *Violence and mental disorder: Developments in risk assessment*. Chicago, IL: University of Chicago Press.
Monahan, J., Steadman, H. J., Silver, E., Appelbaum, P. S., Clark Robbins, P., Mulvey, E. P., Roth, L. H., Grisso, T., & Banks, S. (2001). *Rethinking risk assessment: The MacArthur study of mental disorder and violence*. New York: Oxford University Press.
Newhill, C. E., Mulvey, E. P., & Lidz, C. W. (1995). Characteristics of violence in the community by female patients seen in a psychiatric emergency service. *Psychiatric Services*, *46*(8), 785–789.
Nicholls, T. L., Ogloff, J. R., & Ledwidge, B. (2003). *Inpatient aggression and clinicians' assessments of violence risk*. Paper present at the meeting of the American Psychology and Law Society and European Association of Psychology and Law, Edinburgh, Scotland.
Otto, R. K. (1992). Prediction of dangerous behavior: A review and analysis of "second-generation" research. *Forensic Reports. Special Issue: Psychopathology and Crime*, *5*(1), 103–133.
Otto, R. K. (1994). On the ability of mental health professionals to "predict dangerousness": A commentary on interpretations of the "dangerousness" literature. *Law & Psychology Review*, *18,* 43–68.
Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). Violent offenders: Appraising and managing risk. *The Law and public policy: Psychology and the social sciences series*. Washington, DC, US: American Psychological Association.
Robbins, P. C., Monahan, J., & Silver, E. (2003). Mental disorder, violence, and gender. *Law and Human Behavior*, *27*(6), 561–571.
Skeem, J. L., & Mulvey, E. P. (2002). Assessing the violence potential of mentally disordered offenders being treated in the community. In A. Buchanan (Ed.), *Care of the mentally disordered offender in the community* (pp. 111–142). Oxford: Oxford University Press.
Slovenko, R. (1992). The right of the mentally ill to refuse treatment revisited. *Journal of Psychiatry and Law*, *20*(3), 407–434.
Spector, R. (2001). Is there racial bias in clincians' perceptions of the dangerousness of psychiatric patients? A review of the literature. *Journal of Mental Health*, *10*(1), 5–15.

Straus, M. A., & Gelles, R. J. (Eds.). (1990). *Physical violence in American families: Risk factors and adaptations to violence in 8,145 families.* New Brunswick, NJ: Transaction Publishers.
Swets, K. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240,* 1285–1293.
Ward, T., & Eccleston, L. (2000). The assessment of dangerous behaviour: Research and clinical issues. *Behaviour Change*, *17*(2), 53–68.