Using Graphs to Improve Violence Risk Communication

N. Zoe Hilton, Elke Ham,
University of Toronto and Waypoint Research Institute

Kevin L. Nunes,
Carleton University

Nicole C. Rodrigues,
Royal Ottawa Health Care Group

Cairina Frank,
Carleton University

and Michael C. Seto,
Royal Ottawa Health Care Group

Version  Post-print/accepted manuscript


How to cite TSpace items

Always cite the published version, so the author(s) will receive recognition through services that track citation counts, e.g. Scopus. If you need to cite the page number of the author manuscript from TSpace because you cannot access the published version, then cite the TSpace version in addition to the published version using the permanent URI (handle) found on the record page.

This article was made openly accessible by U of T Faculty. Please tell us how this access benefits you. Your story matters.
Using Graphs to Improve Violence Risk Communication

N. Zoe Hilton, Elke Ham,
University of Toronto and Waypoint Research Institute
Kevin L. Nunes,
Carleton University
Nicole C. Rodrigues,
Royal Ottawa Health Care Group
Cairina Frank,
Carleton University
and Michael C. Seto,
Royal Ottawa Health Care Group


Author Note

N. Zoe Hilton, University of Toronto Department of Psychiatry and Waypoint Research Institute, Penetanguishene, Ontario, Canada; Elke Ham, Waypoint Research Institute, Penetanguishene, Ontario, Canada; Kevin L. Nunes and Cairina Frank, Department of Psychology, Carleton University, Ottawa, Ontario, Canada; Nicole C. Rodrigues and Michael C. Seto, Royal Ottawa Health Care Group, Ottawa, Ontario, Canada.

Cairina Frank is now at Department of Medicine, McGill University, Montreal, Quebec, Canada.

Please address correspondence to N. Zoe Hilton, Ph.D., Waypoint Research Institute, 500 Church Street, Penetanguishene, ON, Canada, L9M 1G3. Email: zhilton@waypointcentre.ca
Abstract

We examined the use of graphs as an aid to communicating statistical risk among forensic clinicians. We first tested four graphs previously used or recommended for forensic risk assessment among 442 undergraduate students who made security recommendations about two offenders whose risk differed by one actuarial category of risk for violent recidivism (Study 1). Effective decision making was defined as actuarially higher risk offenders being assigned to greater security than lower risk offenders. The graph resulting in the largest distinction among less numerate students was a probability bar graph. We then tested this graph among 54 forensic clinicians (Study 2). The graph had no overall effect. Among more experienced staff, however, decisions were insensitive to actuarial risk in the absence of the graph and in the desirable direction with the addition of the graph. Further research into the benefit of graphs in violence risk communication appears viable.

Key Words: forensic, graphs, risk assessment, risk communication, violence
Using Graphs to Improve Violence Risk Communication

The communication of violence risk using numeric information has received growing attention over the past 15 years. Researchers stress the need to conceptualize risk as a continuous variable, and to communicate risk using precise statistical estimates to aid communication (e.g., Babchishin & Hanson, 2009; Hanson, Lloyd, Helmus, & Thornton, 2012; Hilton, Carter, Harris, & Sharpe, 2008; Scurich & John, 2012; Vrieze & Grove, 2010). There are several reasons for this emphasis. For example, structured assessment of risk is more accurate than unaided clinical judgment (e.g., Ægisdóttir et al., 2006; Grove, Zald, Lebow, Smith, & Nelson, 2000). Actuarial data can help evaluate the proficiency of forensic decisions and the costs and benefits of detention policies (e.g., Blasko, Jeglic, & Mercado, 2011; Harris & Rice, 2007, 2013; Lindsay et al., 2010), and actuarial tools appear at least as effective as structured clinical judgment for predicting violent recidivism (e.g., Campbell, French, & Gendreau, 2009; Singh, Grann, & Fazel, 2011; Yang, Wong, & Coid, 2010).

Statistical information, however, can be confusing, especially for less numerate persons (e.g., Wegwarth & Gigerenzer, 2011). Clinicians appear uncomfortable with statistical information and rarely use actuarial tools to assess violence risk (e.g., Heilbrun et al., 2004; Kwartner, Lyons, & Bocaccini, 2006; Taylor, Goldberg, Leese, Butwell, & Reed, 1999; Vrieze & Grove, 2009; though see Vrieze & Grove, 2010). The use of pictograms and other graphical representations of risk has been suggested in order to aid risk communication (e.g., Babchishin & Hanson, 2009; Galesic, Garcia-Retamero, & Grigerenzer, 2009; Lipkus & Hollands, 1999; Waters, Weinstein, Colditz, & Emmons, 2006), but their benefit to violence risk communication remains unknown. The present research investigated the effect of graphs on violence risk communication. We first tested a variety of graphs in a large student sample, then tested the most promising graph in a sample of forensic clinicians.
The Use of Graphs to Communicate Risk

Studies on the communication of personal risks have shown that careful use of graphs can improve the accuracy of risk perception and aid health-related decisions (e.g., Schapira, Nattinger, & McHorney, 2001; Stoff & Swerlick, 2013; Stone, Yates, & Parker, 1997; Waters et al., 2006), especially when the proportion of a population at risk is depicted (e.g., Ancker, Senathirajah, Kukafka, & Starren, 2006; Wilhelms & Reyna, 2013). Icon arrays, in which the proportion of the population at risk is shaded in a different color, can improve risk communication even among less numerate adults (e.g., Galesic et al., 2009; Garcia-Retamero & Cokely, 2013; Visschers, Meertens, Passchier, & de Vries, 2009), although less numerate people may still find it difficult to process the numeric information contained in graphs (e.g., Hess, Visschers, Siegrist, & Keller, 2011; Hess, Visschers, & Siegrist, 2011).

It is not yet clear whether graphs aid the communication of offenders’ risk of violence. Babchishin and Hanson (2009) suggested that probabilistic and categorical risk communication could be improved by explicitly linking risk categories with numerical information and graphs. Their illustrations included a thermometer-style gauge for portraying an offender’s percentile rank within a specified reference class (i.e., offenders with a similar offense type). Elsewhere, histograms or bar graphs of the probability of recidivism, and overlaying lines representing percentile rank, have been used to portray the results of violence risk assessment tools (e.g., Harris, Rice, Quinsey, & Cormier, 2015; Hilton, Harris, & Rice, 2010) and icon arrays and pie charts in which the proportion of the population at risk is shaded have been recommended for use with victims (Hilton et al., 2010). We pilot-tested a number of these recommendations to communicate the results of an existing actuarial risk assessment, the Violence Risk Appraisal Guide (VRAG; Quinsey, Harris, Rice, & Cormier, 2006; Harris et al., 2015).
Measuring the Effectiveness of Violence Risk Communication

In research investigating the communication of risk for violent recidivism, Hilton and colleagues (Hilton, Harris, Rawson, & Beach, 2005; Hilton et al., 2008) evaluated the quality of forensic decisions using a performance standard in which assigning greater security to higher risk offenders was considered desirable. Risk was measured by the VRAG, a 12-item actuarial risk assessment that yields a total score that places the assessed offender into one of nine categories of ascending risk for violent recidivism. Hilton et al. (2008) gave their participants brief case descriptions either with or without a numerical probability statement, and with or without the addition of non-numerical risk category labels (e.g., “high risk”). They reported that participants assigned greater security to the case with the higher VRAG score when the risk assessment result was stated (particularly as a percentage rather than a frequency) but adding a categorical judgment (e.g., “high risk”) did not improve the decision. In these studies, the higher and lower risk cases used differed substantially in their estimated risk of violence, corresponding to an actuarial risk of 64% and 24%, respectively. Scurich, Monahan, and John (2012) subsequently reported that a case description and statistical information was sufficient to produce a desirable distinction between offenders with an 8%, 26% or 76% risk. In the present research we aimed to extend this research further, examining whether forensic decisions can be made that are sensitive to differences between risk levels as small as consecutive categories of violence risk. The risk categories in Study 1 differed in likelihood of recidivism by as little as seven percentage points.

Strategies to improve appropriate use of actuarial risk need to consider the limiting effect of numeracy because the benefit of any risk communication method could differ for more or less numerate people (e.g., Fagerlin, Ubel, Smith, & Zikmund-Fisher, 2007). Numeracy affects
performance on tasks involving calculating or applying risks (e.g., Chapman & Liu, 2009; Keller & Siegrist, 2009; Scurich et al., 2012) resulting in poor decision making (Peters, Hibbard, Slovic, & Dieckmann, 2007). Even highly educated adults show poor statistical reasoning (e.g., Lipkus, Samsa, & Rimer, 2001) and statistical innumeracy is commonplace among clinicians (e.g., Wegwarth & Gigerenzer, 2011) who make little distinction between different risk statistics (e.g., Hilton et al., 2005). Numeracy also limits the efficiency and accuracy of use of numeric information depicted in graphs (e.g., Hess, Visshchers, & Siegrist, 2011), leading some researchers to consider implications for designing and testing risk communication graphs with the less numerate in mind (e.g., Hess, Visschers, Siegrist, & Keller, 2011). In the present research, we included a brief test of statistical numeracy and treated it as a covariate in a test of factors associated with performance on a forensic decision task. Also, we identified less numerate participants as in some previous research on numeracy and risk communication (Scurich, 2015; Scurich et al., 2012) to examine which condition was most associated with their performance.

**Study 1**

Several different graphical formats have been suggested for use to communicate offenders’ risk of recidivism, but there is little existing empirical literature indicating that graphs can aid forensic decision making that can guide such a study with forensic clinicians. Evidence from medical risk communication studies shows that graphs that research participants say they prefer (including simple vertical bar charts) are not necessarily those that improve the accuracy of the risk perception or the extent to which their decisions are associated with the portrayed risk (e.g., Ancker et al. 2006; Brown et al., 2011; McCaffery, Dixon, Hayen, Jansen, Smith, & Simpson, 2012; Price, Cameron, & Butow, 2007). Therefore, we conducted a pilot study to
examine whether graphs can influence forensic risk communication, and to investigate the most effective format for improving forensic decisions from the suggestions made in the existing offender risk communication literature.

We selected graphs to represent probability of recidivism or percentile rank because both are normally communicated in violence risk assessment. Absolute probability of recidivism has been portrayed in a bar chart indicating probability (or percentage) of recidivism as a function of risk category by the developers of a number of actuarial risk assessment tools (e.g., Harris et al., 2015; Hilton, Harris, & Rice, 2010). Bar charts are considered suitable for portraying categorical data (e.g., Lyons, 2010), although they do not necessarily lead to the most accurate risk-related decision making (e.g., Brown et al., 2011; McCaffery et al., 2012). Pie charts representing percentage recidivism across risk categories have also been used, particularly for communicating risk to laypersons (e.g., Hilton et al., 2010), although there is evidence that pie charts do not help people compare risks (e.g., Stone et al., 2003).

Percentile rank is arguably more important when making decisions about more than one case but this statistic is poorly understood in forensic risk assessment (e.g., Hilton et al., 2005). We created a second bar graph illustrating percentile rank based on the Harris et al. data instead of their superimposed line graph in order to more clearly illustrate each offender’s place within the distribution. We also used a thermometer style graph to illustrate percentile rank, based on a recommendation by Babchishin and Hanson (2009). Babchishin and Hanson also suggested icon arrays to portray the confidence interval around the probability of recidivism. We did not test this particular suggestion because the confidence interval is not normally reported with the VRAG and it would require an additional statistical explanation.
In sum, we created four graphs, two (a bar chart and a series of pie charts) to represent absolute risk and two (a bar chart and a thermometer) to represent percentile rank. A fifth condition involved no graph. We tested which condition led to security level recommendations that most strongly distinguished between the two offenders within each pair in terms of the security level chosen. We had no a priori standard for the “most suitable” security levels and anticipated that there could be substantial disagreement between participants on appropriate security. Using a within-participants design controlled for this disagreement and enabled a difference score to be calculated that represented each participant’s use of the security options to distinguish between offenders of different risk levels.

**Method**

The pilot study was reviewed and approved by the Research Ethics Boards of the first author’s institution and the university at which the study took place. Students in undergraduate psychology courses signed up for the study for class credit via an online recruitment system. Informed consent, study materials, and debriefing information were all administered through an online survey system subscribed to by the university. After eliminating students who gave no response \((n = 250)\) or participated in the wrong condition due to a hyperlink error \((n=60)\), the total sample was 442 adults (70% female) whose mean age was 20.5 years \((SD = 2.8)\).

**Design and Materials.** We tested the effect of graphs on forensic decision-making in a 5 between-participants (absolute risk bar chart, absolute risk pie chart, percentile rank bar chart, percentile thermometer, or no graph; Figure 1) by 2 between-participants (high or low risk level pair of cases) by 2 within-participants (high or low risk level within each case pair) design. High and low risk level was determined by constructing offender case summaries in which the stated actuarial risk was either above or below the base rate in the normative population. Case
descriptions involved hypothetical offenders who had committed a violent offense and were undergoing psychiatric assessment. Offenders were described in terms of the 12 items comprising the Violence Risk Appraisal Guide (VRAG; Harris et al., 2015), and were presented in pairs, representing either VRAG risk categories 3 and 4 (24% and 31% risk of violent recidivism within 10 years, respectively), or VRAG categories 7 and 8 (64% and 82% risk of violent recidivism within 10 years, respectively). These risk categories were selected in order to provide comparable results with previous research which used categories 3 and 7 (Hilton et al., 2005). With the aim of avoiding a ceiling effect whereby most participants would readily distinguish between offenders with a 4-category difference in risk, and to test whether participants could distinguish between offenders with more similar risk, we paired each case summary with another in a consecutive risk category. The order of presentation cases within each pair was counterbalanced but not used as an experimental condition.

Within risk levels, each participant read each pair of offender case summaries twice. Presentation 1 included the description of case characteristics only. After each case, participants were asked to select a forensic security placement from one of four options: 1 – release to the community with psychiatric treatment, 2 – commit to a psychiatric hospital with access to the community, 3 – commit to a hospital with no access to the community, 4 – commit to a maximum security hospital for at least 12 months. Presentation 2 was preceded by the instructions, “On this page you will read about the two violent offenders again. This time we have added information about the results of a violence risk appraisal instrument (VRAG).” Presentation 2 included the same information as Presentation 1 plus a) a statement of statistical results of the VRAG assessment, including the risk category, the percentile rank in terms of the percent of offenders scoring lower (low risk condition) or higher (high risk condition) and the
percentage likelihood of recidivism within 10 years after release, as recommended by the VRAG authors and as used in previous risk communication research (Hilton et al., 2005), and b) one of the four graphs or no graph. After each case participants were again asked to select a forensic security placement from one of the four options.

**Dependent Measures.** The main dependent measure was the difference in forensic security recommendation between the higher risk case and the lower risk case within each pair, with a positive value indicating that the higher risk case of the two was selected for greater security (a desirable result). This difference score does not yield a meaningful number translatable into actual security placements, but it does provide a measure of the distinction that participants made between the two cases. This approach was established by Hilton et al. (2005, 2008) as a measure of effective risk communication. It is similar to a method used by Hess, Visschers, and Siegrist (2011) who interpreted the difference between the ratings of risk on a six point scale from 1 (very low) to 6 (very high) for higher and lower risk cases. This forensic decision score was calculated once for Presentation 1 and again for Presentation 2, and we conducted a 2-within (Presentation 1 vs. 2) by a 2-between (high vs. low risk level) by 5-between (no graph or one of four graphs at Presentation 2) analysis of variance (ANOVA) to test the effect of graph condition and risk level on the difference in the forensic decision measure from Presentation 1 to 2.

Participants also responded to questions about their experience with mathematics and statistics, and completed a short test of numeracy derived from Lipkus et al. (2001). We used the total correct score as a continuous measure of numeracy in our main analysis entering it as a covariate in the ANOVA. We then divided participants into more or less numerate groups based on scores above or below the sample median on the numeracy test, as in previous research
(Scurich et al., 2012), to further explore which type of graph was the most beneficial for less numerate participants. Mean forensic decision scores falling outside the 95% confidence interval of another were considered statistically different.

**Results**

In the 2- x 2 x 5 mixed ANOVA on the forensic decision score with numeracy score as a covariate, there was a main effect of risk level, \( F(1,417) = 48.81, p < .001, \text{partial } \eta^2 = .105 \), a main effect of Presentation, \( F(1,417) = 4.71, p < .031, \text{partial } \eta^2 = .011 \), and a three-way interaction, \( F(4,417) = 3.03, p = .018, \text{partial } \eta^2 = .028 \), but no effect of numeracy, \( F(1,417) = 0.19, p = .667 \). The cases in the high risk pair were distinguished in the desirable direction (positive forensic decision scores). Low risk cases were distinguished in the undesirable direction (i.e., participants selected higher security for the risk category “3” case than the risk category “4” case), especially at Presentation 1, but this distinction improved at Presentation 2.

Comparing the benefit of graphs for less numerate participants across risk levels, the bar chart representing absolute probability of recidivism produced the largest positive improvement in the forensic decision score, \( M = 0.63 (SD = 0.90), 95\% \text{ CI } = [0.20, 1.06] \), and produced a significantly better forensic decision score than both the no graph condition, \( M = -0.14 (SD = 1.02), 95\% \text{ CI } = [-0.48, 0.21], d = 0.80 \), and the pie charts, \( M = 0.05 (SD = 1.43), 95\% \text{ CI } = [-0.62, 0.72], d = 0.49 \).

**Discussion**

The bar chart of absolute probability of recidivism resulted in the most effective risk communication, and it adhered to principles of graphical representation of categorical data more closely than the other graphs (e.g., Lyons, 2010; Stone et al., 2003). Therefore, we selected this bar chart for our main study. In Study 1 we used a student sample because there was no previous
research in forensic risk communication to guide our choice of graphs and we wished to test multiple conditions, requiring a larger sample than we anticipated we could recruit from professionals. Study 2 attempted to replicate the benefit of a graph among forensic clinicians.

Study 2

The clinicians who staff forensic psychiatric units are the principal decision makers concerning the security and supervision of forensic patients. Daily decisions about grounds privileges and community access are made in-hospital on the basis of clinical factors, within the limits set by review boards (e.g., Quinsey, Coleman, Jones, & Altrows, 1997). Review board decisions, in turn, are strongly correlated with clinicians’ recommendations (e.g., Hilton & Simmons, 2001; McKee, Harris, & Rice, 2007; Wilson, Crocker, Nicholls, Charette, & Seto, 2015) as are comparable judicial and conditional release decisions (e.g., Blais, 2015; Guy, Kusaj, Packer, & Douglas, 2015; Konečni & Ebbesen, 1984). Consequently, Study 2 concentrated on the use of graphs to improve violence risk communication among practicing forensic clinicians.

Early research indicated that clinicians had objections to statistical risk communication (e.g., Heilbrun et al., 2004; Kwartner et al., 2006; Taylor et al., 1999) but more recent work suggests that the use of statistical or mechanical risk assessment increases with more recent graduation from clinical training, mediated by the discussion of such tools in graduate school (Vrieze & Grove, 2009). There is also evidence suggestive of change from initial research finding that forensic decisions were unrelated to actuarial risk (e.g., Hilton & Simmons, 2001) towards a small association between decisions and measured risk (e.g., McKee et al., 2007), and more recent evidence of clinicians’ use of evidence-based tools and risk factors (e.g., Crocker, Nicholls, Charette, & Seto, 2014; Hilton, Simpson, & Ham, 2016; Wilson et al., 2015). In the
present study, therefore, we also measured years of forensic clinical experience and attitudes towards evidence based practices.

**Method**

This study was conducted in two psychiatric hospitals, among staff of forensic clinical units (maximum and medium security). It was reviewed and approved by the Research Ethics Boards of each hospital.

**Participants.** One of the study sites was in central Ontario, Canada and had a forensic clinical staff of approximately 200, primarily registered nursing and registered practical nurses but also including psychiatrists, psychologists, social workers and other professions. The other hospital in eastern Ontario had a forensic clinical staff of approximately 95 with a similar professional make up. Invitations to participate in the survey were initially emailed to forensic staff to participate in an online version of the survey, but so few participants completed the survey within the first week ($n = 6$) that we attended staff meetings on the units at both hospitals to recruit participants until our goal of 60 participants was met. The refusal rate at these recruitment meetings was approximately 12%. Questionnaires were provided along with written and oral instructions. Participants were informed that their participation was voluntary, their responses would be treated anonymously and confidentially, and they could contact specific persons regarding the research study and/or ethics with any questions or concerns. To preserve anonymity we requested no identifying information such as professional discipline.

**Materials and Procedure.** The main experimental condition was the presence or absence of a graph used to illustrate the actuarial risk of recidivism posed by two hypothetical forensic patients. Each participant read a pair of cases, presented in counter balanced order, who scored at the third or fourth categories of the VRAG. We chose to use only one pair of cases and
not test higher and lower risk cases (as was done in Study 1) to reduce the number of variables for what we expected to be a much smaller sample. We chose VRAG risk categories three and four because they did more poorly in Study 1 in the absence of a graph and we wished to see whether the benefit of the graph could be replicated. This choice likely meant subjecting our hypothesis that a graph would improve risk communication to a relatively stringent test. Each case description consisted of a single paragraph describing a patient who had committed a violent act, was detained in a maximum security hospital, and was about to be reviewed by the review board. The patient’s characteristics were described according to the 12 items that comprise the VRAG (Harris, et al., 2015): whether he lived with his biological parents to age 16, his elementary school adjustment, history of alcohol problems, marital status, nonviolent criminal history, any conditional release failures, his age at the time of the index offense, any victim injury, victim gender, whether he met the DSM-III criteria for a personality disorder or for schizophrenia, and his score on the Hare Psychopathy Checklist – Revised. Both case descriptions were the same length and reading level (Flesch-Kincaid Reading Level 8.3 – 8.5, measured using Microsoft Word 2007).

Each description was followed by a statement of the VRAG results, including the risk category, the percentile rank as a percentage of offenders scoring lower, and the percentage likelihood of violent recidivism within 10 years after release, as recommended by the VRAG authors and used in previous risk communication research (e.g., Hilton et al., 2008). In half the questionnaires, the results included a probability bar chart and the statement, “The graph below shows the probability of violent recidivism for the nine VRAG categories. Mr. [M or P]’s probability of violent recidivism is highlighted. The graph used for the category 3 case (Mr. P.) is shown in Figure 1.
Participants were asked three questions after each case. The first two served as manipulation checks to ensure that participants attended to the actuarial risk information in both cases: “If Mr. X were released into the community today, what do you think is his likelihood of violent offending within 10 years?” and “If 100 male violent offenders were standing in a row in order of risk of violently offending, and the first man had the lowest risk and the 100th man had the highest risk, where would Mr. X stand?” These two questions were intended to reflect absolute risk and percentile rank, respectively, and served to check whether participants attended to the stated risk communicated in the case summary.

The third question asked participants to recommend a security level for each case. The response options reflected the normal possible recommendations to make about a patient to the review board: 1 - absolute discharge, 2 - transfer to open setting, minimum security hospital, 3 - transfer to medium security with community access, 4 - transfer to medium security with no community access, 5 - remain in maximum security but with an early review, 6 - remain in maximum security with the normal 12 month review. We tested whether participants’ security recommendations distinguished between patients in consecutive risk categories using a repeated measures t-test of the security recommendations for the two cases.

**Dependent Measures.** The main dependent measure was the forensic decision score, equal to the difference in security level recommended for the two patients. Following Hilton et al. (2008), we defined a good forensic decision as one in which the category 4 case was recommended higher security than the category 3 case, which would yield a positive value for the forensic decision score. We tested the effect of the graph on the forensic decision score in a 2-within (the two cases) by 2-between (graph or no graph) analysis of variance.
Participants then completed the Evidence Based Practice Attitudes Questionnaire (EBPAS; Aarons, 2004), stated how many years’ experience they had working in forensic mental health, and answered three questions pertaining to risk and probability. Two of these questions were drawn from Lipkus et al. (2001): “Which of the following represents the biggest risk of getting a disease - 1/10, 1/100, 1/1000” and “If the probability of getting a viral infection is 0.0005, how many people out of 10000 will get infected?” Questionnaire length precluded using the whole Lipkus et al. test, but we added a third question intended to yield greater discrimination: “If the odds of an event happening is 3:1, what is the probability? 25%, 33%, 66%, 75%.” We used EBPAS total score, experience, and numeracy test total score as continuous measures in a series of correlational analyses to test their bivariate association with the forensic decision score, and then added those that were associated as factors to the analysis of variance of forensic decision scores.¹

Results

Participant Characteristics. There were 62 participants, but only 54 provided a security recommendation for both cases, our primary dependent measure. The mean length of experience working in forensic mental health was 13 years (SD = 9.46). In the numeracy test, the fraction question was answered correctly by 76% of participants, the probability question by 55%, and the odds ratio question by 42%. On the EBPAS, the mean total was 3.78 (SD = 0.55), above the normative average of 2.30 (Aarons, 2004), indicating moderate-to-great acceptance of evidence-based practice. Experience and numeracy were not substantially associated with each other, r (50) = .115, p = .420, but total mean EBPAS scores were higher among participants with greater numeracy, r (51) = .299, p = .031, and with less forensic experience, r (50) = -.370, p = .007,
indicating that more numerate and less experienced clinicians had more positive attitudes towards evidence based practice.

**Manipulation Checks.** Manipulation checks indicated that, on average, the perceived risk was consistent with the stated probability of recidivism for both cases. For Mr. M, the stated probability was 31%, and the risk perceived by participants was $M = 34.41$ ($SD = 17.07$); one-sample $t$ test of difference from 31, $t(50) = 1.43$, $p = .160$, 95% CI of the difference = [-1.39, 8.21]. For Mr. P, the stated probability was 24%, perceived risk $M = 23.46$ ($SD = 13.09$); $t(51) = -0.297$, 95% CI of the difference = [-4.18, 3.12], $p = .768$. The stated percentile rank for Mr. M. was the 40th percentile, and the perceived percentile rank was $M = 44.18$, ($SD = 17.59$); $t(51) = 1.71$, $p = .093$, 95% CI of the difference = [-.717, 9.078]. Mr. P’s percentile rank was stated as the 22nd percentile and the perceived percentile rank was $M = 29.10$ ($SD = 19.26$), $t(51) = 2.66$, $p = .010$, 95% CI of the difference = [1.74, 12.26]. Thus, the manipulation appeared effective on average, as participants’ mean reported perception of risk in the cases was negligibly different from the stated risk; however, some variability in perceptions is noted.

**Forensic Security Recommendations.** The median security recommendation for both cases was 3 (transfer to medium security hospital with community access). The mean security recommendation for Mr. P (VRAG category 3) was 2.78 ($SD = 1.15$), close to “transfer to medium security with community access,” and the mean security level for Mr. M (VRAG category 4, was 3.52 ($SD = 1.22$), falling between the “community access” and “no community access” options for transfer to medium security hospital. The mean difference between these two decisions was 0.72 ($SD = 0.98$), which was significantly above zero, $t(53) = 5.42$, $p < .001$, 95% CI of the difference = [0.45, 0.99], $d = .64$, indicating on average desirable forensic decision making. Most (30, 56%) participants made a distinction in the desirable direction, including 56%
in the graph condition and 11 (44%) in the no-graph condition, \( \chi^2(2,54) = 3.71, p = .156 \). Five (9%) made recommendations in the “wrong” direction and 19 (35%) made the same recommendation for both offenders.

In the test of the effect of the graph on forensic decision scores, the 2-within (the two cases) by 2-between (graph vs. no graph) ANOVA yielded only a main effect of case, \( F(1, 52) = 28.63, p < .001 \), partial \( \eta^2 = .355 \). There was no overall effect of the provision of a graph, \( F(1, 52) = 1.33, p = .255 \), and no interaction effect, \( F(1, 52) = 0.92, p = .161 \).

Variability in the forensic decision measure was not associated with correctly answering the fraction question \( r(53) = .003 \), the probability question \( r(53) = .018 \), or the odds ratio question, \( r(53) = .123, ps \geq .375 \), or with mean total EBPAS scores, \( r(51) = -.20, p = .157 \). Forensic decisions were associated with years of forensic experience, \( r(50) = -.438, p = .001 \); participants with less experience made a more positive distinction between the two cases (i.e., “better” forensic decision making). A 2-within (the two cases) by 2-between (graph vs. no graph) ANOVA with years of experience as a covariate yielded a main effect of case, \( F(1, 48) = 39.09, p < .001 \), partial \( \eta^2 = .449 \), and an interaction of case and experience, \( F(1, 48) = 11.94, p = .001 \), partial \( \eta^2 = .199 \), whereby less experienced staff made a more positive distinction between the two cases, \( M = 1.18, (SD = 0.86) \), 95% CI = [0.84, 1.16], than more experienced staff, \( M = 0.26 (SD = 0.92) \), 95% CI = [-0.14, 0.66], \( d = 1.03 \). There was no main effect of graph, \( F(1, 48) = 1.57, p = .216 \); however, among more experienced staff, based on a median split at 10 years, the provision of a graph was associated with significantly improved forensic decision making, \( F(1, 21) = 5.38, p = .031 \), partial \( \eta^2 = .451 \), reaching a level comparable to less experienced staff whose forensic decision was not significantly affected by the provision of a graph (Figure 2).
Discussion

In this experimental study of decision making by forensic clinicians, security recommendations were on average consistent with an objective standard of good decision making, whereby actuarially higher risk offenders were assigned to greater security than lower risk offenders. Forensic clinicians made this distinction between cases in consecutive categories of violence risk, which differed in probability of violent recidivism by seven percentage points. Providing clinicians with a graph, selected from a pilot study as one that appeared most effective in violence risk communication, had no overall effect on decision making. Among more experienced staff, however, the provision of a graph had a positive effect on decision making, and without the graph forensic decisions were not in the desirable direction. These findings indicate that there is value in further investigating the benefit of using graphs to improve violence risk communication.

Results also point to the possibility that a newer generation of forensic clinicians may be more likely to make security-related decisions in accordance with actuarial risk, although it might also be that more clinical experience detracts from good decision making. Such an effect could be attributable to decreased attention to new evidence or resistance to developments in forensic risk assessment, similar to the resistance to new statistical techniques observed among researchers by Sharpe (2013). Indeed, we observed that less experienced forensic clinicians had more positive attitudes towards evidence based practice. Future studies of risk communication among practicing clinicians could examine their familiarity with and beliefs about formal risk assessment and their relation to the influence of risk assessment instruments on their decision making. Clinical training in risk assessment, and its evaluation, could be tailored for participants’ levels of clinical experience and comfort with evidence based practice.
Limitations

In order to protect anonymity, we did not ask participants to indicate their professional discipline and we did not measure participants’ familiarity with or experience of using actuarial measures. Our finding that experience was associated with forensic decision making could be explained more parsimoniously if less experienced participants were more familiar with the VRAG as a result of more recent forensic training. Not knowing participants’ discipline also meant we could not test professional group differences, determine the representativeness of our sample, or compare our sample with those from related studies. Also, our small sample likely restricted statistical power to detect small effect sizes, and we were not able to study more than one graph or one pair of risk categories.

The hypothetical nature of the risk communication exercise raises the question of generalizability to actual clinical practice, and the potential demand characteristics inherent in presenting participants with two cases of different stated actuarial risk might have exaggerated participants’ ability to make forensic decisions that distinguish between adjacent risk categories. Future studies could attempt to replicate the present results using full length reports that are more typical of the material in psychological reports of violence risk and that might lead to more conservative decision making overall (e.g., Hilton et al., 2008). Other more naturalistic conditions could include using graphs during expert testimony or mock mental health tribunals, or evaluating the effects of a controlled implementation of incorporating graphs into actual risk assessment reports. Outcome measures could move beyond the security level distinction score and include time to reduced security, prioritization for treatment, or conditions imposed.

General Discussion
Violence risk communication research is a fledgling field compared with medical risk communication (e.g., Hilton, Scurich, & Helmus, 2015), in which researchers have explored such questions as how much detail is helpful in a graph (e.g., Zikmund-Fisher, Fagerlin, & Ubel, 2010) and created computer assisted dynamic presentations of graphical information (e.g., Okan, Garcia-Retamero, Cokely, & Maldonado, 2015). Comparable research still needs to be conducted on the use of graphs in forensic contexts, in which the decisions are not so much about which treatment a patient might choose but about how much supervision an offender might require during treatment, and the cost-effectiveness of decisions is not so much about relative mortality and survival rates but about balancing the protection of potential victims with the rehabilitation and recovery of offenders. It remains to be tested whether these somewhat different questions require different communication methods in the way that different message framing promotes preventative actions compared with treatment choices in medical health (e.g., Edwards, Elwyn, Covey, Matthews, & Pill, 2001).

General Limitations and Future Research

We made decisions about which graphs to test based on their previous use and recommendations for violence risk communication, but had to limit the number of different graphs out of necessity in order to keep to a practical sample size in the pilot study. We elected not to include the confidence-interval icon array previously suggested in the forensic risk communication literature, but icon arrays are considered by some to be the graph of choice for medical risk communication and they can now be readily generated using online technology (e.g., http://www.iconarray.com). Future research could take advantage of this resource to compare the effectiveness of icon arrays with the simpler bar graph, or to experiment with optimal icon arrays for communicating violence risk.
We used a simple rule for determining “good” forensic decisions. Strengths of this measure of the effectiveness of risk communication include that it was derived from previous research in this field, it represented security options available for forensic patients in our jurisdiction, and it was a simple and brief task suitable for use with students as well as busy professionals. Limitations of this measure include that we had no expectation for the exact security level that should be allocated to an offender in any given category of risk, as risk is not the sole determinant of such decisions and we did not manipulate or measure other pertinent variables such as the resources available in a given jurisdiction. Relatedly, failing to distinguish between two VRAG categories out of nine may not always be a concern if there are only five security options. However, this measure did provide a strong test of clinicians’ ability to use actuarial risk information with the aid of a graph for the purpose for which it was intended. More nuanced investigations of forensic decision making could include the process of allocating treatment or other risk management resources to offenders. For such research lay populations are unlikely to be a proxy for forensic clinicians.

Existing forensic risk communication research has relied on a combination of student samples and small professional samples. Similar to our research, mail-outs to professional association members are subject to low participation rates (e.g., Heilbrun et al., 2004) and researchers often resort to student populations (e.g., Heilbrun, Wolbransky, Shah, & Kelly, 2010; Scurich & John, 2011). Studying forensic or medical students might be a suitable compromise but it might lack ecological validity and obscure the effects of work experience such as we detected. Public recruitment systems like Amazon’s Mechanical Turk have also been used in this field (e.g., Scurich, Monahan, & John, 2012) and have some advantages for behavioral research taking precautions to ensure validity (e.g., Chandler, Mueller, & Paolacci, 2014), but
again this kind of sampling does not represent forensic professionals. An alternative professional group to consider in future research is probation and parole officers, who are more populous than forensic mental health professionals and frequently assess and communicate offender risk. Incentives for participating might also be an appropriate consideration and would not be a disproportionate expense compared with the costs expended in developing and conducting risk assessments.
References


doi:10.1016/j.pec.2010.04.027


doi:10.1037/a0014474
doi:10.1177/0963721413491570


http://dx.doi.org/10.1027/hb0000111


doi:10.1177/0093854807307029

doi:10.1002/bsl.2048


Footnote

Footnote 1: It was our intention to examine whether the provision of an evidence-based security recommendation, the Forensic Institutional Decisions or Recommendations Scheme (FIDORS; McKee et al., 2007), would improve forensic decision making on a second presentation of the forensic patient cases following all of these measures. However, due to lower rates of completion of this part of the questionnaire, and uneven questionnaire completion patterns such that the FIDORS condition was unsuccessfully factored with the graph vs. no-graph condition, this experimental condition yielded no interpretable results and all results presented below are based on the first presentation without this manipulation.
Figure 1. Graphs tested in Study 1 Presentation 2. Originals in color.
Figure 2. Interaction of experience and graph on forensic decision-making score. A positive score indicates the higher risk case was recommended to greater security than the lower risk case. Less experienced staff (open circles) had more positive scores, and the provision of a graph in risk communication improved scores among more experienced staff (closed circles).
Appendix

Example Case Description

Mr. P. is a 27 year old forensic patient in maximum security at [name of hospital removed for masked review] who was found NCR two years ago. Mr. P. was born in Toronto, Ontario. He was well behaved and academically successful throughout his elementary and high school years. He left home at age 19 to attended McMaster University in Hamilton and lived with his girlfriend for 2 years. He experienced major depressive episodes on and off since leaving university. Mr. P. has never had an alcohol problem himself and there is no history of alcohol or substance abuse in any of his family members. He has no prior criminal record. Mr. P. was admitted to Oak Ridge in 2011 on charges of assault following an attack on his girlfriend in their home; the victim required treatment at a hospital emergency clinic. Mr. P. had not been using alcohol at the time. He does not meet the criteria for schizophrenia. He does not have any personality disorder and his score on the Psychopathy Checklist (PCL-R) is 5.

On the basis of his score on a violent offender risk appraisal instrument (the VRAG) constructed from the variables used to predict violent recidivism, Mr. P.'s category of risk of violent recidivism is in the 3rd, or third lowest, of 9 categories on the VRAG. Among offenders in a standardized sample, 22% obtained lower scores, and 24% of offenders in Mr. P.’s category reoffended violently within an average of 10 years after release.

The graph below shows the probability of violent recidivism for the 9 categories. Mr. P.’s probability of violent recidivism is highlighted.
N. Zoe Hilton is an associate professor in the Department of Psychiatry at the University of Toronto, Senior Research Scientist in the Waypoint Research Institute, and a Registered Psychologist. Her primary research interests concern domestic violence risk assessment, risk communication, and forensic psychiatric services.

Elke Ham is a research psychometrist at the Waypoint Centre for Mental Health Care. She holds a Research Analyst diploma from Georgian College. She is the project coordinator for “ODARA 101: The Electronic Training Program” for the Ontario Domestic Assault Risk Assessment. Her research interests are in risk assessment and communication, and clinical outcome indicators for forensic patients and violent offenders.

Kevin L. Nunes is an associate professor in the Department of Psychology and director of the Aggressive Cognitions and Behaviour Research Lab at Carleton University. The main focus of his research is on the conceptualization and measurement of cognitions thought to be relevant to sexual and nonsexual violent behavior, and the role these cognitions may play in violent behavior.

Nicole C. Rodrigues, M.Ed., CCC, is a research coordinator at the Brockville Mental Health Centre. She is also a Canadian Certified Counsellor with a focus on forensic psychiatric populations.

Cairina Frank received an undergraduate degree in Integrated Sciences with a concentration in Forensic Science at Carleton University, and a master’s degree in Epidemiology and Biostatistics at The University of Western Ontario. She is currently a student within the Department of Medicine at McGill University, and has research interests in human behavior and perinatal epidemiology.
Michael Seto, PhD, is a forensic research director at the Royal Ottawa Health Care Group, with cross-appointments at the University of Toronto, Ryerson University, Carleton University, and the University of Ottawa. He has presented and published extensively on paraphilias, sexual offending, and risk appraisal. He is Editor-in-Chief of Sexual Abuse: A Journal of Research and Treatment.