מאיירס - ג'וינט - מכון ברוקדייל
MYERS - JDC - BROOKDALE INSTITUTE
مايرس- جوينت- معهد بروكديل

CENTER FOR RESEARCH ON DISABILITY AND SPECIAL POPULATIONS

Instruments for Assessing the Risk of Violent Behavior toward Others and Suicidal Behavior

Review of the Literature

Assaf Kron

The review was commissioned by the Department of Forensic Psychiatry, Mental Health Services, Ministry of Health, and the National Social Work Service at the Ministry of Health and was funded with the assistance of the National Social Work Service at the Ministry of Health

RESEARCH REPORT

RR-494-07

Executive Summary

1. Introduction

The aim of this review is to present procedures for assessing violent behaviors toward others and procedures that focus on evaluating suicidal behavior. It was carried out at the request of the Department of Forensic Psychiatry at the Ministry of Health and focused on the possibility of applying certain instruments as decision-support instruments in four clinical decisions:

1. Assessing violent behavior toward another person for the purpose of a compulsory psychiatric intervention (forced hospitalization)
2. Assessing violent behavior toward another person for the purpose of issuing or revoking a firearms license
3. Assessing violent behavior toward a wife or partner for the purpose of intervention in the family
4. Assessing suicidal behavior to support a decision for forced/compulsory hospitalization.

The term "dangerousness" is today a central aspect of legal argumentation regarding forced psychiatric interventions in Israel (the Mental Health Treatment Law, 1991) and North America (Monahan et al., 2001). For example, in order to bring a person to an "emergency forced examination," it has to be proven, inter alia, that "he is at immediate risk of endangering himself or another person physically." Hence the assessment of the "extent of dangerousness" carries great weight in decisions regarding non-voluntary interventions and the revoking of rights and has attracted much attention (Borum, 1996; Mossman, 1994). While various systems (courts, police, health, etc.) require professionals to assess the "dangerousness" of a person, there is no clear professional standard (psychological or otherwise) or legal standard for this procedure (Whittemore and Kropp, 2002). Furthermore, there is no special training for the assessment of dangerousness either in research or practice (Borum, 1996). The growing demand of the various systems to assess dangerousness and, on the other hand, the obstacles impeding professionals from doing so constitute the context in which the discussion about prediction and assessment takes place.

One of the basic problems confronting us in the current review is to give meaning to the statistical findings in a way that will help reach a decision about policy on the use of a particular assessment instrument. The empirical findings cannot be completely translated into operative conclusions, as the latter take into account moral, financial, and other considerations in addition to the empirical findings. However, in order to permit discussion of the instruments and the conclusions and to facilitate the decision-making process, we shall define, as far as possible, the criteria for "successful prediction" of violent or suicidal behavior. We shall briefly discuss several fundamental concepts, basic problems, and possible solutions in predicting violent or suicidal behavior and clarify the statistical meaning of indices.

An "assessment of dangerousness" can be classified as a "clinical assessment" or an "actuarial assessment." In the professional arena today, most of the decisions about the "extent" that a person is dangerous are based on clinical assessment. In this report, "clinical assessments" are defined as procedures in which the relation between information gathered during the assessment and the final decision is not a known mathematical relation. This is important for two reasons: first, the professionals' reasoning and judgment carry great weight. Second, the way that the conclusion is reached is not known to outside parties nor, on many occasions, to the professionals themselves and they may not necessarily be aware of the variables and their relative importance to their decision. The term "actuarial assessment" unlike "clinic assessment" refers to a mathematical decision-making model (Buchanan, 1999). The actuarial instruments cope with both the above points: they offer better control over the variables and their weights and they make it possible to track the assessment process.

The distinction between actuarial and non-actuarial is not clear-cut and the two forms of assessment – clinical and actuarial – can be seen as the two ends of a continuum. In addition to the essential difference between the presence and absence of an explicit mathematical calculation, there are also differences in the types of instruments that are utilized. Thus, for example, clinical assessments are frequently based on clinical theory and refer to personality traits (despite the fact that the "personality = clinical" connection is not inevitable) and similarly, actuarial assessments tend to be based on "stable" variables (such as demographic and historical variables) although this is not inevitable either. We will review relatively recent instruments that reflect a synthesis between the two approaches (the clinical and the actuarial). These instruments do not rule out the inclusion of personality variables from the clinical theories about violence alongside "classical" actuarial (demographic and historical) variables, while defining a mathematical function that links the information gathered (the items) and the final decision (scores or final scores).

In order to use a certain procedure as a decision-support instrument, it is recommended that its accuracy and validity be tested. The validity of a test concerns the association between the test score and the evidence that supports the conclusions drawn from it (Cronbach, 1971). This is a most important matter when it comes to making a decision about the use of a particular instrument. There are various types of validity. The current review focuses upon criterion-related validity, i.e., the association between the test score and an external criterion. Criterion-related validity can be divided into two sub-types. Predictive validity is the validity of a test score in relation to the criterion – in our case, a certain behavior – in the future. Research designs in which the predictor variables are measured before the criterion are called predictive designs and they make it possible to estimate the predictive validity. Concurrent validity concerns the validity of the test scores in relation to the criterion in the present or past. Study designs in which the predictor variables are measured concurrently or after the criterion are called postdictive designs.

To examine the validity of a procedure, a range of indices can be used. In the current review we will, where possible, employ an accuracy index called ROC (Receiver Operating Characteristic). A ROC analysis is accepted in many disciplines that engage in diagnostics and its use is now

increasing. The ROC's advantage is that it allows a description of the accuracy of the prediction that is unaffected by the base rate or by the tendency of professionals to use a certain decision threshold and thus to prefer a certain type of error (Swets, 1988; Rice and Harris, 1995; Dolan and Doyle, 2000; Borum, 1996; Mossman, 1994). The ROC curve is a function of sensitivity and specificity, a ratio that indirectly expresses the interchange between the two. In addition, the analysis makes it possible to obtain detailed information about the behavior of the instrument at various decision thresholds. In order to compare the various instruments, we will use the AUC (Area Under the Curve) as an index that averages the predicted accuracy of the instrument at various decision thresholds. The AUC ranges from 0.50 to 1. An area of 0.50 shows that the test does not add information that contributes to the prediction of the criterion (Hanley and McNeil, 1982), while an area of 1 reflects perfect prediction ability. In order to make it possible to compare the different studies and instruments, we have transformed the various indices into the AUC index.

To reach conclusions about the validity of a procedure, we must ensure that the population on which it is tested in the study corresponds to the population to which it will be applied. Target populations will be defined as those for which there is an interest in applying the dangerous assessment procedures in Israel, while the study populations will be defined as those on which the procedures have been validated in practice. It goes without saying that the more that the study population's characteristics differ from those in the target population, the greater the concern that there will be differences both in the distribution and the validity. In the current review, we try to classify the populations and concentrate on those that are as close as possible to the target populations.

2. Instruments for Assessing Dangerousness toward Others

2.1 Instruments for Assessing Dangerousness toward Others in the Psychiatric Population

This part of the report describes four instruments for assessing violent behavior toward others in psychiatric and forensic psychiatric populations. The instruments were selected after a review of the literature, a search of databases, and consultation with professionals and researchers in this field. We have not included the instrument known as VRAG (Violence Risk Assessment Guide) (Quionsey et al., 1998), which was originally examined with criminal populations (which are of less concern to us) nor will we discuss instruments used to predict sexual assault.

1. PCL-R (Psychopathy Checklist–Revised) (Hare, 1991), one of the most important and well-known instruments for predicting violent behavior, is an improved version of the PCL. The revised version (PCL-R) shows sufficient psychometric qualities that also enable clinical application (Hare, 1991; Dolan and Doyle, 2000). Although it is an instrument for assessing the structure of a psychopathic personality, the PCL-R has become a key predictor variable of violent behavior and its general score is today the strongest single predictor (e.g., Monahan and Steadman, 1994). The list consists of twenty items coded on the basis of a semi-structured

interview and/or data in the person's dossier. The final score is produced after about two hours of interviewing and a further hour of encoding and it provides an evaluation of the extent to which a subject fits the prototype of a psychopath. The current review examined the association between the PCL-R and the violent behavior criterion among criminal/forensic psychiatric populations.

2. PCL-SV (Screening Version) is a shortened version of the former and facilitates a quick assessment using a clinical interview. The review examined the validity of the shortened version among forensic psychiatric populations and civil psychiatric populations while hospitalized and while in the community. The shortened version takes approximately 75 minutes to administer (45 minute interview and 30 minutes encoding).

3. HCR-20 (Historical/Clinical/Risk management: 20 items) (Webster et al., 1997), is a list of 20 risk factors of violent behavior. The structure of the HCR-20 is based on a division into three types of variable – from the past, present, and future (Douglas, 2001). It includes the PCL (the item that assesses psychopathic characteristics and is based on the final score of one of the two versions of PCL). The current review includes studies of forensic psychiatric populations and civil psychiatric populations, both hospitalized and in the community. The interview lasts at least 45 minutes and a further 30 minutes are required to code to PCL-SV (if the screening version is chosen) excluding the time for the remaining items on the HCR-20. Altogether between an hour and an hour-and-a-half are devoted to the PCL-SV and a minimum of half-an-hour to the remaining items on the HCR-20.

4. MacArthur VRAS (Violence Risk Assessment Study). The VRAS examines violence toward others among psychiatric populations (Monahan et al., 2000). The VRAS is software based on a process of several consecutive classifications in the form of a classification tree (CT) and a comparison and combination of several classification trees. The review explains the concepts of a "simple CT" and an "iterative CT" and a comparison and combination of different CTs to increase the statistical validity of the prediction (comparison of models). There are several obvious advantages to this instrument: (1) In contrast with earlier models, the subjects are divided into three risk groups using two decision thresholds: Below one threshold are the "low risk" subjects and above the other are the "high risk" subjects. This produces an "unclassified" group of subjects who do not belong in either group. At the end of the process of combining the models, there are five risk groups with different probability rates for violent behavior; (2) The model shows a particularly high predictive validity (the highest published to date); (3) It includes variables that are accessible to the clinician while interviewing; (4) It takes much less time to conduct an assessment with this instrument than with the others (approximately 15 minutes compared with one-to-three hours using the PCL-SV/R and the HCR-20). Despite the clear advantages of the instrument, there is one great disadvantage – it is based only on a postdictive study.

Validity

In populations of civil psychiatric patients who were examined during hospitalization (the assessment was conducted and the criterion measured during hospitalization), a low validity of AUC=.60 and .61 was found for the PCL-SV and the HCR-20, respectively. When the civil psychiatric population was examined after discharge into the community, a medium validity was found and HCR-20 had a slight advantage over PCL-SV – areas under the curve in the order of 0.73 and 0.69, respectively. The meaning of the areas is that if the sample were divided into violent and non-violent and if two subjects were selected at random (one from each group) and the process were repeated many times, then in 73% and 69% of the cases – based on the HCR-20 and PCL-SV, respectively – the subjects would indeed be classified in the appropriate groups (i.e., the violent and the non-violent).

"Forensic psychiatric" patients constitute a more "distant" population from the target population with regard to the relevant variables. In an examination of the assessment of violence within the institution, in the case of forensic psychiatric patients, mean areas under the curve of AUC= 0.67, 0.74, and 0.70 were found for PCL-SV, HCR-20, and PCL-R, respectively. When forensic psychiatric patients were examined after they were discharged into the community, mean areas under the curve of AUC= 0.71 and 0.72 were found for PCL-SV and HR-20, respectively. In conclusion, in the case of forensic psychiatric patients both in an institution and after their discharge into the community, the HCR consistently shows a somewhat higher validity than the PCL-SV and the PCL-R.

The VRAS shows higher validity than others (AUC=.88), gives a more complex classification, and requires less time for administration (15 minutes compared with an hour for the PSL-SV, and 1.5–2 hours for the HCR-20). Yet, the VRAS has two main drawbacks: (1) Only one study has been conducted; and (2) There are no data from a predictive design. Findings of a predictive study are expected to be published in the future.

2.2 Assessment and Prediction of Assault by Spouse or Partner

Out of dozens of existing instruments, regarding spousal assault we found evidence of predictive/concurrent validity for only two (Hart, 2003; Roehl and Guertin, 1998: (1) SARA (Spousal Assault Risk Assessment) (Kroop et al., 1999; Kroop and Hart, 2000) and (2) DA2/DAS2 (Danger Assessment) (Campbell, 1986).

1. SARA (Spousal Assault Risk Assessment – Kroop et al., 1999; Kroop and Hart, 2000). This instrument is defined as a checklist for conducting an assessment to predict violence toward (but not murder of) a wife or partner. The list includes twenty risk factors that are obtained from a semi-structured interview with potential wife beaters. Half of the risk factors assess "general violence" and half specifically address "violence towards one's spouse." The ten risk factors in the latter category include three variables concerning violence that has occurred near the time of the assessment. The interview and coding take about fifteen minutes. In the United

States, SARA assessments are today accepted as decision-support instruments in the courts in the states of Colorado and Vermont (Roehl and Guertin, 2000).

2. DA2 (Danger Assessment – Campbell, 1986). This instrument belongs to a group of instruments designed to assess the risk of spousal murder. The goal of the procedure is to predict the cases where spouses are at risk of their lives (as opposed to predicting violence towards the spouse, in the case of SARA). The development of the list was based on several postdictive studies of murder or severe harm inflicted by husbands or male partners. The first part of the DA2 includes questions that consist of a sort of preface to the questionnaire, the goal of which is to reduce denial of the violence by the women tested, where applicable. The second part comprises fifteen "yes/no" answers to be completed by the subjects, which represent the risk factors for spousal murder. The studies in the review include populations of potential victims of serious spousal injury (generally women exposed to violence during the period they responded to the questionnaire). It takes approximately 15 minutes to administer the questionnaire and encode it.

The review describes three more instruments that, despite the absence of data as to their validity and reliability, are in extensive clinical use and are at present the subject of a predictive study: (1) DVSI (Domestic Violence Screening Instrument); (2) Kingston Screening Instrument for Domestic Violence and (3) Mosaic-20.

Validity

Validity of SARA: Two studies, one postdictive (Kropp and Hart, 2000) and one predictive (Heckert and Gondolf, 2002) examined the validity of SARA (Whittemore and Kropp, 2002). Kropp and Hart (2000) examined a population of 2,681 people consisting of 1,010 prisoners and 1,671 persons on probation.

The validity of the instrument, i.e., its association with spousal violence, was examined in a postdictive design in three ways:

1. A comparison was made between prisoners with a history of violence against their spouses and prisoners without such a history, and a significant difference was found between the final score for SARA in both groups, AUCt=0.85.

2. A comparison was made between a population of recidivists (those who resumed violent behavior toward their partners after they had taken part in a treatment program) and non-recidivists (who did not exhibit violence toward their partners after treatment). No significant difference was found in the overall score of the list in this comparison. A significant difference was found in the indices in the second part (specifically addressing spousal violence): AUC=66.

3. In a somewhat different analysis, the subjects were divided according to SARA into three categories: "high risk," "medium risk," and "low risk." The association between the SARA classification and the recidivism criterion was medium: Cohen's d=0.76, AUC=0.70 (Kropp and Hart, 2000).

Heckert and Gondolf (2002) examined 804 women whose husbands were in a rehabilitation program following violent behavior toward them. The examination was conducted in a predictive design after fifteen months. The researchers reported 43% sensitivity and a false positive of 27%.

Validity of the DA2: Three postdictive studies (Campbell et al., 2003; Tolman and Saunders, 2000; Goodman et al., 1999) and one predictive study (Heckert and Gondolf, 2002) examined the predictive validity of the DA2.

Campbell et al. (2003) examined 220 relatives of women who had been murdered by their spouses/partners and a control group of 343 women who reported that they had been abused by their partners. They report an area under the curve of AUC= 0.90.

In the same study that examined SARA, Heckert and Gondolf (2002) performed the procedure for 804 women whose partners had been in a rehabilitation program for 15 months. They reported 66% sensitivity and a false positive of 33%.

Weisz et al. (2000) reported a lack of ability to predict when the DA2 was examined for 177 wives of men who had been convicted in the courts of violent behavior toward their partner.

Assessment of Violent Behavior in Israel

In Israel, very limited use has been made of semi-actuarial instruments as decision-support instruments for assessing violent behavior toward others. The interviews have given the impression that in most cases an intuitive "clinical assessment" is made on the basis of the professional's experience. Two main reasons have emerged for the minimal use of structured instruments: (1) Budget difficulties: developing an instrument, along with research and the training of professionals to use it, requires appropriate funding; and (2) In some cases there is opposition to the use of a structured instrument. This opposition may be the result of the training given to professionals, which does not familiarize them with the use of structured procedures, and the assumption that the validity of the instruments is insufficient and therefore it is not worth investing in using them. Two instruments are being developed in Israel: one by the Israel Police, the other by the Ministry of Social Affairs in collaboration with the University of Haifa.

1. The instrument developed by the police (Dr. Morag, 2003, personal correspondence) is designed to assess the danger of spousal violence. It can be administered by various professionals (not only mental health professionals). No data have been published about the reliability and validity of the instrument. Furthermore, neither the items nor the instrument itself are given to outside parties. The instrument is in the final stages of predictive validation and is applied on a regular basis as a decision-support instrument at two police stations.

2. The instrument developed by the Ministry of Social Affairs and the University of Haifa (Dr. Innes, 2003, personal correspondence) is also designed to assess the danger of spousal violence. As in the previous case, no data have been published about the reliability and validity of the instrument and neither the items nor the instrument itself are given to outside parties. The instrument has been validated postdictively vis-à-vis convictions in the courts for four kinds of assault, from murder through minor assault where no injury is sustained.

3. Prediction and Assessment of Suicidal Behavior

This section of the report discusses seven instruments for predicting and assessing suicidal behavior with the goal of using them to support a decision for forced hospitalization. As will be recalled, this review includes only those instruments for which either predictive validity or concurrent validity has been reported.

1. The SSI (Scale for Suicide Ideation) (Beck et al., 1979) is one of the most important instruments for assessing suicidal behavior (Range and Knott, 1997; Brown, 2003). It comprises 19 items that examine the strength of attitude, behavior, and planning of suicidal behavior on the day of the interview. The review includes studies of populations in ambulatory psychiatric care and/or hospitalized.

2. The SSI-W (Scale for Suicide Ideation-Worst) is unique in that it examines different aspects of suicide ideation during the period when, in the subject's opinion, these suicidal thoughts were at their peak ("worst"). This is in contrast to the SSI and other instruments, which examine suicidal thinking at the time of the interview (SSI-Current). The interviewer asks the interviewees to recall the period when they experienced the strongest suicide ideation and asks them to refer to that time when answering the questions. The period could be several years in the past. The logic underlying this instrument is based on earlier findings that showed a link between the strength of suicide ideation in the past and the carrying out of a suicidal act subsequently. Following this reasoning, the instrument asks specifically about the period when the suicidal thoughts were at their worst. The instrument includes 19 items. The studies included in the review were conducted on a population of people seeking ambulatory treatment.

3. The ASIQ (Adult Suicide Ideation Questionnaire) is intended to assess the frequency of suicidal thoughts among adults. The items on the questionnaire examine different aspects of suicidal thought from unconscious suicidal fantasies such as the wish "not to have been born" to specific thinking such as how, when, where, writing suicide notes, etc. (Reynold, 1991). The review includes a study on a student population and civil psychiatric populations.

4. The SIS (Suicide Intent Scale) (Beck et al., 1974) examines the seriousness of the suicide attempts of subjects who have made such an attempt. It includes 15 items that assess behavior and attitudes before, during, and after the suicide attempt. The first part of the SIS comprises eight items that examine what the authors call "objective aspects" of the suicide attempt, such

as preparations, execution, and clues that the subject gave to those around him/her. This part can sometimes be completed on the basis of records. The second part concerns the subject's perception of the finality of the suicide attempt and includes items such as expectations of being saved or the declared intent of the suicide attempt. The studies in the review include an examination of the criterion in the psychiatric population.

5. The BHS (Beck Hopelessness Scale) (Beck and Steer, 1988) was developed in order to examine positive and negative beliefs about the future during the week before the questionnaire is completed. The questionnaire comprises 20 statements to which the respondent has to respond "yes" or "no." The questionnaire is completed by the subjects themselves with no need for a preliminary interview and it takes five minutes to do so. The review includes studies on hospitalized populations with or without a history of suicidal ideation or behavior and an ambulatory population.

6. RFL/LRFL (the Linehan Reason for Living Inventory) (Linehan et al., 1983) assesses the will and motivation to live. In an attempt to assess the potential for suicidal behavior, the questionnaire examines the forces pushing the subject toward suicidal behavior and the barriers impeding it. The questionnaire comprises 48 Likert-type items, such as "I would like to accomplish my plans or goals in the future." The respondent is asked to what extent each item is a reason "for not committing a suicidal act." Administration takes about 20 minutes. The studies in the review include the following populations: hospitalized psychiatric patients, hospitalized psychiatric patients suffering from major depression, shoppers at a mall, and psychology students.

7. The MAST (Multi-attitude Suicide Tendency) (Orbach et al., 1991) was developed in Israel for the purpose of studying adolescent suicide (not for the purposes of predicting at the individual level) and is based on a theoretical assumption that four forces influence suicidal behavior: attraction to life, repulsion by life, attraction to death, and repulsion by death. Each of the thirty Likert-type items on the questionnaire represents one of these four forces that influence suicidal behavior. The MAST has been examined on the following populations: adolescent high school pupils, adolescents receiving psychiatric treatment, and hospitalized adolescents (with or without suicidal behavior). Some of the studies employed female adolescents only. The studies have been conducted in Israel and in the United States.

Validity

Altogether, with regard to the ability to predict suicidal behavior using the seven instruments (as examined in 25 studies), the area under the curve was between AUC= 0.62 and 0.82, with a mean of 0.73. It should be noted that there is a sample overlap in three of the instruments (BHS, SSI-W, and SSI). However, we made an effort to avoid overlap and over-reporting (which happens in the literature) with regard to each individual instrument. In other words, the review does not include different publications that address the same sample or an overlapping sample for the same instrument.

Three instruments demonstrated medium to high validity. The SSI (Scale for Suicide Ideation) was examined in four studies and has an AUC of 0.77. It is important to note that the ability of the instrument to distinguish between those who have committed a suicidal act and those who have not, among subjects who are already engaged in suicide ideation, is negligible and amounts to little more than a guess. Apparently such people, with the high base rate of suicide ideation, are closer to our target population: in many cases there is an a priori apprehension of suicidal behavior. The SSI-W (Scale for Suicide Ideation-Worst) was examined in one study and showed an AUC of 0.82. Note that it is important to consider the performances of the instrument on populations with a high base rate of suicidal ideation. The idea underlying the SSI-W is interesting in itself, because clinical and psychiatric interviews generally examine suicidal behavior at the time of the interview and not the severity of the behavior and attitudes when the ideation was at its peak. Such an examination, whether it is conducted using the SSI-W or during a clinical interview, could, as we have seen, indicate the potential severity of suicide ideation in the future. The third instrument that demonstrated medium to high validity is the ASIQ (Adult Suicide Ideation Questionnaire). The instrument is interesting because it also examines fantasies such as "not wanting to have been born," which in certain circumstances could incorporate a subconscious wish to die. The ASIQ showed the highest mean validity (AUCt=.86). However, note that this result is based solely on postdictive designs.

The BHS (Beck Hopelessness Scale), SIS (Suicide Intent Scale), RFL/LRFL (the Linehan Reason for Living Inventory), and the MAST (Multi-attitude Suicide Tendency) showed below average performance. It is interesting that for subjects with suicide ideation (Beck et al., 1985), the BHS showed medium-low prediction ability (0.70) but it was higher than that of the SSI. When the BHS was examined on a population of subjects who had been hospitalized due to a suicide attempt, the instrument did not predict suicidal behavior for a period of five to ten years. The LRFL, which assesses barriers to suicidal action, was examined in six postdictive studies and showed a medium-low validity of 0.68.

The MAST (Multi-attitude Suicide Tendency), an instrument constructed in Israel and validated in Israel and the United States, was tested on adolescents and showed a medium-low validity (MAUCt=70). However, it is important to note that the overall score was calculated by averaging all four indices. This action is not actually performed according to the instructions in the questionnaire and was only done to make it possible to sum up the findings for a single validity measure (i.e., specific scales had a higher validity).

When we come to assess the performance of the instruments, attention must be paid to the time span over which the criterion is measured. In postdictive assessments the time span is largely not noted. With the exception of one study, which examined suicidal behavior over a period of three months (Osman et al., 1999) and did not report predictive validity, the studies examined the criterion over a time span of five to ten years. Extreme caution must be taken in drawing conclusions from these findings with regard to the time span over which it had been intended to apply the instruments for the current purposes. The instrument has been validated for a five-to-ten

year time span and is apparently more effective for long-term planning purposes than for use as a decision support tool regarding forced hospitalization, where a short-term estimate about risk for suicidal behavior is required.

4. Conclusions and Recommendations

Compulsory Hospitalization

It was our intention to find an instrument that could assist in evaluating the civil psychiatric population, preceding hospitalization, as a decision support instrument regarding the decision about compulsory hospitalization. No specific studies of this population were found. Most of the studies conducted on civil psychiatric populations were conducted during hospitalization (forced or otherwise) or afterwards, during rehabilitation in the community. In accordance with the review of the literature in this report, the recommendation is to consider using the HCR-20 or VRAS, while stressing the importance of adjusting the instrument to Israel and to the target population by means of a preliminary study and examination of the predictive validity.

Firearms License

The goal of assessing the risk of violent behavior as a decision support instrument regarding the issuing of a firearms license is to assess violent behavior among people who have declared that they are, or have been, receiving psychological and/or psychiatric treatment and are applying for a firearms license. While there is a debate as to whether the population receiving psychiatric treatment is more violent than the rest of the population, we are not familiar with data linking the fact that a person is receiving psychological treatment with an inclination to violence. It is, perhaps, possible to apply the recommendations for forced/compulsory hospitalization and use an instrument based on HCR-20 or VRAS. It is important to note that the instrument can be used as a decision support instrument regarding the potential for violent behavior but not as a decision support instrument regarding a firearms license. The decision about issuing a firearms license must, necessarily, include broader criteria than the potential for violent behavior (for example with regard to people who are in a lower-than-average risk group for violence, but suffer from severe difficulties of impaired reality testing and judgment).

Spousal Violence

It is difficult to compare the four instruments reviewed in the context of using a procedure as a decision support in the case of spousal violence. This is because there are still no validity and reliability data for the instruments developed in Israel and because the instruments reviewed do not predict the same criterion (the DA2, as stated, predicts murder). The police instrument is in the advanced stages of development, toward the end of a predictive study, and the findings are expected to be published in the coming months. In addition, the two instruments reviewed, SARA and the DA2, are supported by a small number of studies that indicate medium validity.

Meanwhile, for the time being we recommend considering a procedure based on the instrument developed by the Israel Police. However, it is likely that there will be difficulty transferring information from one agency to another (using the procedure outside of the police). This

difficulty could become a real problem because some of the items are based on prior information known to the police. A solution could be to develop a partial version that does not include the items based on such information. However, it has to be remembered that the validity may be lower without these items. In addition, it is important to ascertain that the validity of this instrument will not fall below that of SARA. Whether it is decided to use the instrument developed by the police or to use the Israeli version of SARA, it is possible – and worthwhile – to add the DA2, which could constitute an additional and important measure of the potential for serious physical harm and a threat to life. In addition, consideration should be given to using the first part of the DA2 or an alternative procedure designed to reduce the tendency of battered women to denial, which would be performed before the assessment.

Suicidal Behavior

Based on the review of the literature in this report, the recommendation is to conduct a comprehensive study of two of the instruments, the ASIQ and the SSI-W, which have in the past demonstrated a medium-high validity and to test them on the target population in Israel over a time span relevant to the decision in question. (Since the decision concerns hospitalization in the immediate future, it is essential to examine the instruments' ability to predict suicidal behavior for the immediate future rather than for more distant periods in time, as prevalent in various studies). We would note that even if it is decided not to use a semi-actuarial instrument, such as the two aforementioned instruments, consideration may be given to using the rationale underlying these two instruments also for a regular clinical assessment. That is, it is worthwhile to consider assessing the seriousness of the suicidal ideation and behavior when they were at their peak (and not only during the interview and the period preceding it). It is further useful to be aware of the negative impact of the questions in the clinical interview, and to employ the assessment of barriers to suicidal behavior.

Table of Contents

List of Tables

List of Graphs