



FEATURE ARTICLE

Kennedy Axis V: Clinimetric properties assessed by mental health nurses

Margo D. M. Faay,^{1,2} Roland van de Sande,^{4,6} Floor Gooskens² and Thóra B. Hafsteinsdóttir^{1,3,5,7}

¹Clinical Health Sciences, Faculty of Medicine, Utrecht University, ²Department of Psychiatry, Division of Neurosciences, ³The Department of Rehabilitation, Nursing Science and Sport, Rudolf Magnus Institute, University Medical Centre Utrecht, ⁴Faculty of Health, University of Applied Sciences Utrecht, ⁵Research Centre for Innovations in Health Care, University of Applied Sciences Utrecht, Utrecht, ⁶Mental Health Centre Bavo-Europoort, Rotterdam, the Netherlands; and ⁷Faculty of Nursing, University of Iceland, Reykjavik, Iceland

ABSTRACT: *The Kennedy Axis V is a routine outcome measurement instrument which can assist the assessment of the short-term risk for violence and other adverse patient outcomes. The purpose of this study was to evaluate the interrater reliability and clinical utility of the instrument when used by mental health nurses in daily care of patients with mental illness. This cross-sectional study was conducted in inpatient and outpatient adult psychiatric care units and in one adolescent inpatient unit at a university hospital in the Netherlands. Interrater reliability was measured based on the independent scores of two different nurses for the same patients. The clinical utility of the instrument was evaluated by means of a clinical utility questionnaire. To gain a deeper understanding of rating difficulties at the adolescent unit, additional data were collected in two focus group interviews. The overall results revealed a substantial level of agreement between nurses (intraclass correlation coefficient and Pearson 0.79). Some rating challenges were identified, including difficulties with scoring the instrument and using tailor-made interventions related to the scores. These challenges can be resolved using refined training and implementation strategies. When the Kennedy Axis V is accompanied by a solid implementation strategy in adult mental health care, the instrument can be used for short-term risk assessment and thereby contribute in efforts to reduce violence, suicide, self-harm, severe self-neglect, and enhanced objectivity in clinical decision-making.*

KEY WORDS: *Kennedy Axis V, mental health nursing, psychometric properties, risk assessment, violence, instrument.*

INTRODUCTION

Early recognition of emerging agitation is crucial to prevent or minimize severe violent incidents to self or

others at acute psychiatric units. Violence often results in seclusion in some European countries (Janssen *et al.* 2011; Kaltiala-Heino *et al.* 2003; Keski-Valkama *et al.* 2010). Various studies reveal that seclusion rates differ substantially regionally in the same type of clinical populations. The prevalence of seclusion varies 20–74% (Janssen *et al.* 2011; Keski-Valkama *et al.* 2010; Tunde-Ayinmode & Little 2004). Worldwide, many psychiatric hospitals are involved in seclusion and restraint reduction programs (Husum *et al.* 2010; Hyde *et al.* 2009; Steinert *et al.* 2010). The prevention of violent incidents is crucial as they are highly associated with the start of seclusion

Correspondence: Margo D. M. Faay, Division of Neurosciences, Department of Psychiatry, University Medical Centre Utrecht, HP A.00.241, PO Box 85500, 3508 GA Utrecht, the Netherlands. Email: M.D.M.Faay@umcutrecht.nl

Margo D. M. Faay, MSc, RN.

Roland van de Sande, MSc, RN.

Floor Gooskens, MSc, RN.

Thóra B. Hafsteinsdóttir, PhD, RN.

Accepted September 2012.

episodes (Foster *et al.* 2007). Therefore, early recognition of increased agitation and deterioration needs constant attention to enhance proactive risk management strategies. Risk assessment instruments are important aids to facilitate effective risk management in clinical psychiatry (Webster & Hucker 2006).

Risk assessment instruments are designed to capture high-risk features in individual cases and are normally accompanied by severity scores (Kelly *et al.* 2002). Most of these instruments include either static factors, which change rarely, or dynamic factors, which can change frequently (Kumar & Simpson 2005). Because agitation and deterioration are complex and dynamic phenomena, risk assessment instruments should include biological, social, and psychopathological aspects (Amore *et al.* 2008; Duxbury & Whittington 2005), and preferably include situational and environmental factors (National Institute for Clinical Excellence [NICE] 2005). Several experts argue that clinical judgement combined with the use of structured risk assessment instruments will result in more efficient clinical decision-making processes (Blok *et al.* 2010; Ruiter 2007; Webster *et al.* 2002).

Risk assessment instruments

We conducted a systematic published work review concerning violence risk assessment instruments to select clinically useful instruments to incorporate in our seclusion reduction program in three psychiatric units in one of the largest university hospitals in the Netherlands and found 19 available instruments (M.D.M. Faay, unpubl. data, 2010). The methodological quality of the studies was appraised by using the Quadras standards (Whiting *et al.* 2004). Based on the reliability, validity, clinical utility, and the methodological quality of the studies, the three following instruments were identified that showed consistently adequate results and could be considered as useful for frequent short-term risk assessment in our units: (i) the Brøset Violence Checklist (BVC; Almvik *et al.* 2000); (ii) Violence Screening Checklist (VSC; McNiel and Binder 1995); and (iii) Historical Clinical Risk Management-20 (HCR-20; Webster *et al.* 1997). These three instruments have some disadvantages. The BVC has been studied extensively and is clinically useful but relies dominantly on short-term (within 24 hours) dynamic factors and is most predictive during the first days of hospitalization (Almvik & Woods 2003). The VSC mainly captures static factors. The HCR-20 appeared to be less useful in our setting due to the time-consuming clinical interview protocols and the requirement of having access to comprehensive patient records to complete this instrument.

We also identified the Kennedy Axis V (Kennedy 2003a). This routine outcome measurement instrument was developed to refine the Global Assessment of Functioning (GAF) which is incorporated in Axis V of the DSM-IV (American Psychiatric Association 1994). The Kennedy Axis V can be used, like the GAF, for the global assessment of patients' functioning, as it captures the clinician's impression and can be used as a treatment outcome measure (Kennedy 2003a). The instrument captures not only dynamic strength and risk factors in patient functioning, but also important static factors, like previous violence towards self and others and substance misuse (van de Sande, 2007). Therefore, the Kennedy Axis V can be used for a wider range of patient-related risk factors (Kennedy 2003a). Because the instrument uses dynamic and static factors, assesses multiple risk factors, and can be used for outcome measurement, the Kennedy Axis V has many benefits compared with contemporary risk assessment instruments.

Only a few studies systematically evaluated the clinimetric properties of the Kennedy Axis V. Bilezikian (1998; as cited in Kennedy 2003a) found good interrater reliability, varying 0.80–0.91 on different subscales. Ebben (2006) found high interrater reliability of the instrument by comparing scores of psychologists and physicians with an average kappa of 0.85 and concurrent validity of 0.62 as compared to the GAF in a population of psychiatric patients with severe mental illness. However, both studies dealt with small sample sizes and included specific populations, which limits the generalizability of these findings. An Italian study found high interrater reliability between the scores of clinicians rating the Kennedy Axis V, varying 0.63–0.81 as measured with kappa (Mundo *et al.* 2010). The Kennedy Axis V is one of the key scales used in a short-term risk assessment model, the crisis monitor (van de Sande *et al.* 2011a), and is in combination with the BVC recommended as the best practice for short-term risk assessment by the Dutch Inspectorate of Health (Inspectorate of Health 2011). One comprehensive cluster randomized study has been published which focussed on the predictive validity of the crisis monitor and the risk of being exposed to seclusion. This study showed promising results concerning the Kennedy Axis V for short-term risk assessment (van de Sande *et al.* 2011b). The instrument was also used in a study aimed to determine risk factors for seclusion and restraint which found the likelihood of being coerced significantly associated with the Kennedy Axis V (Georgieva *et al.* 2012). Although the Kennedy Axis V is, among other instruments, recommended as best practice in the Netherlands, there are specific aspects that have not been

studied before: the use of the instrument in adolescent populations and the utility of the instrument for nursing practice.

According to the Kennedy Axis V manual (Kennedy 2003a), the instrument is applicable for everyone above the age of 5 years. However, published studies in child and adolescent mental health-care samples are lacking. Risk assessment in adolescents, who are in a vulnerable and dynamic episode of their lives, appears to be rather challenging (Hage *et al.* 2009). Studies revealed that 37–42% of the adolescents became violent during hospitalization (Blake & Hamrin 2007; Lodewijks *et al.* 2008). Patients suffering from psychosis are also associated with a higher risk of violent incidents and in some samples there appeared to be a significant correlation between psychosis and violent incidents (Brendel *et al.* 2010). Approximately 12–39% of first-episode psychosis patients are involved in violent incidents (Harris *et al.* 2010). Therefore, adolescents with mental health problems and adult patients suffering from psychosis need to be monitored frequently by means of structured short-term risk assessment.

The previously described Kennedy Axis V studies did not evaluate the clinimetric properties of the tool in a sample of nurse raters. In clinical practice, mental health nurses are often the frontline health-care professionals and likely to be exposed to impulsive and dangerous behaviour (Foster *et al.* 2007; Woods & Ashley 2007). Due to their frontline position, they are also likely to recognize the early signs of agitation and violent behaviour to self or others. The quality of this proactive work can be enhanced by a more structured approach (Fluttert *et al.* 2008). Identifying high agitation levels, poor control over psychiatric symptoms, and temporary lack of adequate social skills are essential for nurses to promote a safe environment (Duxbury *et al.* 2008). On the other hand, sufficient interrater reliability is crucial for the adequate clinical use of risk assessment tools in day-to-day practice, otherwise consistent risk communication will impede clinical decision-making in unacceptable ways. Therefore, the aim of this study was to determine the interrater reliability and clinical utility of the Kennedy Axis V assessed by mental health nurses in clinical practice at three psychiatric units in a university hospital in the Netherlands.

MATERIALS AND METHODS

Participants

This cross-sectional study was conducted from February until May 2011 in three psychiatric units in a university

hospital in the Netherlands. Unit A and B cater for patients with first-episode psychosis and other psychotic disorders. Unit A is focussed on inpatient care and unit B on outpatients. Unit C is a locked unit for diagnosis and treatment of adolescents with a variety of mental health problems.

A total of 28 mental health nurses worked on the three units. All licensed and registered nurses working day and evening shifts were included in this study, resulting in a final sample of 23 nurses. This sample included nine nurses from unit A, two from unit B, and 12 from unit C. All 23 nurses were invited to participate in the different data collection methods after they were informed face-to-face and by newsletters about their role in the study protocol. The study protocol was approved by the local medical ethical review committee.

Measures

Kennedy Axis V

The Kennedy Axis V is constructed in seven subscales: (i) psychological impairment, including mood changes and psychotic symptoms; (ii) social skills, including interpersonal and communication skills; (iii) violence, including threatening, assaultive, and suicidal behaviour; (iv) activities of daily living and occupational skills, including job or school skills; (v) substance misuse, including alcohol, drugs, and nicotine misuse; (vi) medical impairment, including hypertension and weight gain; and (vii) ancillary impairment, including financial, legal, and housing problems (Kennedy 2003b). In the Dutch version, an additional scale is used: (viii) motivation for treatment, including the patient's cooperation with clinicians and medication adherence (Mulder 2000). Each subscale can be scored in the range 0–100. A score above 50 indicates a strength in the patient's functioning on the specific subscale and, subsequently, a score of 50 or lower indicates problematic functioning in a specific area (Kennedy 2003a). The Kennedy Axis V manual provides brief descriptions of the severity scores to support the completion of the subscales. These descriptions are anchor points and examples of patients' functioning which should generally agree with the clinical impression of the professional (Kennedy 2003a). For every subscale, a dangerousness score can be calculated which is used to determine the most dangerous rating on the subscales. The lowest score indicates the patient is at risk for adverse outcomes in a specific area of functioning. The dangerousness level for each subscale varies 30–50. Those cut-off scores indicate the likelihood of adverse situations (Kennedy 2003b).

Interrater reliability

The interrater reliability of the Kennedy Axis V was calculated by comparing the agreement in scores of two nurses for the same patients in similar timeframes. Both daily and weekly scores were used for the calculation of the interrater reliability. The researcher selected patients randomly using a random number generator on the Internet and monitored if the nurses scored these patients independently of each other.

To ensure sufficient power, we calculated that for an estimated population correlation of 0.80, based on the studies conducted by Ebben (2006), Bilezikian (1998), and Mundo *et al.* (2010), a two-tailed alpha of 0.05 and power of 0.80, the instrument should be scored 10 times to calculate Pearson's *r* (Polit 2010) and 51 times to calculate the intraclass correlation coefficient (ICC) (Bonnet 2002). This is consistent with Terwee *et al.* (2007) who state that for determining interrater reliability, there should be at least 50 screenings conducted twice.

Clinical utility

The clinical utility of the Kennedy Axis V was measured using a questionnaire, which was originally developed to investigate the clinical utility of depression screening instruments for patients with stroke using the criteria of Harris and Warren (1995). Clinical utility reflects aspects such as cost of procuring, reproducing, and retrieving information, and issues on how to score and interpret the data as well as the time needed to administer the instrument (Harris & Warren 1995). The clinical utility questionnaire was constructed using the two step method of Lynn (1986). The content validity was determined in a five round Delphi survey by an expert panel of four clinical nursing researchers. The clarity of the terminology in relation to clinical practice was further assessed by four expert neuroscience nurses. For this study, the questionnaire was adjusted by eliminating questions focusing on specific clinical aspects of stroke patients. Five expert psychiatric nurses evaluated the content, formulation and layout of the questionnaire for its use in psychiatric settings. Only minor changes were made. The questionnaire includes 10 items focusing on nurses' opinions of the questions stated. The answers are provided on a dichotomous scale ('yes' and 'no') with open spaces where the nurses could explain their experiences. Two questions consider the time needed to fill in the questionnaire. Because of the diversity of the questions, no total score is calculated for the questionnaire.

Procedure

The Kennedy Axis V was introduced on the units in January 2011 using five different strategies:

- 1 A plenary lecture on the research protocol and procedures in the team meetings. Also, the Kennedy Axis V manual was provided to all participating professionals.
- 2 A plenary and interactive lecture considering the basic principles of the instrument and a demonstration of the instrument by means of case vignettes. This training was provided twice on unit A and B, once on unit C, and strongly recommended for all professionals. Those who were not present were individually trained by the researcher (unit A and B). Because of organizational problems, the nurses at the adolescent unit C that were unable to attend this plenary lecture were instructed by other colleagues.
- 3 Training on the job while nurses scored patients by means of the Kennedy Axis V.
- 4 Frequent consensus meetings to address scoring problems in specific cases.
- 5 A newsletter was published every second week, which included various aspects of working with the Kennedy Axis V and the proceeding of the study.

The first four subscales of the Kennedy Axis V were scored every day, whereas all eight subscales were scored every week for each patient. All scores were noted in the patient's file and were accompanied by short descriptive reports explaining the scores. The dangerousness cut-off scores mentioned in the Kennedy Axis V manual (Kennedy 2003b) were not formally used. The ratings were incorporated in the daily reports and multidisciplinary meetings. The principal researcher monitored the instrument fidelity on the units and provided coaching on the job during the research period.

Nurses received the clinical utility questionnaire by an email after using the instrument for 1 month (unit C) or 3 months (unit A and B). The data on interrater reliability were collected throughout the study period.

Data analysis

The interrater reliability was calculated using the ICC and the Pearson's *r*. The two-way random ICC model was used because the nurses and patients were both considered to be random (Shrout & Fleiss 1979). Although Pearson's *r* is considered unsuitable for calculating interrater reliability (Bland & Altman 1986; Fleener *et al.* 1996), we calculated this coefficient to compare our findings with previous studies. An ICC score of 0.49–0.59 indicates moderate agreement, a score of 0.60–0.79 substantial agreement, and a score of 0.80 or higher indicates a high level of agreement (Landis & Koch 1977). Pearson's *r* should be 0.80 or higher when used for interrater reliability (Polit & Beck 2008). The demographic data and

the clinical utility questionnaire data were analyzed with descriptive statistics. SPSS ver. 16.0 for Windows (SPSS, Chicago, IL, USA) was used for all data analysis.

Additional data collection and analysis

The nurses at the adolescent unit C experienced considerable rating problems and therefore it was decided to stop using the instrument after 1 month. To explore the impeding factors in the rating process in this adolescent setting, additional data were collected. Two groups of three nurses were formed and each group participated in a focus group interview. These interviews were based on the qualitative research principles of Morse and Field (1986) and the focus was twofold:

- 1 What difficulties did you experience?
- 2 What are the positive factors you experienced while using the instrument?

The issues the nurses described during the interviews were further explored by in-depth questions, clarifying underlying arguments, consequences, and relationships between issues. In the final stage of the interviews, the issues were clustered into categories.

The data were analyzed using deductive content analysis as described by [Elo and Kyngäs \(2007\)](#). The interviews were transcribed and coded with the categories from the interviews. The categories and accompanying issues from both interviews were combined and presented to the respondents to allow a member check. The issues were clustered by two researchers. Finally, the most frequently named positive and negative issues for each category were presented in a figure.

RESULTS

Sample characteristics of nurses and patients

The sample consisted of 23 nurses who participated in different parts of this study. The mean age was 31.3 years and 15 nurses (65.2%) were female (Table 1). The mean age of the patients was 24.6 years at unit A, 24.7 years at unit B, and 16.3 years at unit C (Table 2). Twenty-five patients on unit A (67.5%) and seven patients (58.4%) at unit B suffered from psychosis. Nine patients at unit C (90.0%) suffered from multiple psychiatric disorders.

Interrater reliability

Within the entire sample of 23 nurses, 17 nurses (73.9%) participated in scoring patients in order to measure the interrater reliability. Six nurses were unable to participate because of absence related to illness or holiday leave. In some cases, not all subscales were scored by both nurses

TABLE 1: *Nurses characteristics*

	Total (n = 23) n (%)
No. of nurses	
Unit A	9 (39.1)
Unit B	2 (8.7)
Unit C	12 (52.2)
Age, years	
Mean (SD)	31.3 (7.8)
Range	23–50
Sex	
Male	8 (34.8)
Female	15 (65.2)
Education	
Master's degree	3 (13.0)
Registered nurses (Bachelor degree)	16 (69.6)
Licensed practical nurses	4 (17.4)
Work experience in psychiatry, years	
Mean (SD)	7.6 (6.4)
Range	1–25

SD, standard deviation.

TABLE 2: *Patients' characteristics*

	Unit A (n = 37) n (%)	Unit B (n = 12) n (%)	Unit C (n = 10) n (%)	Total (n = 59) n (%)
Age, years				
Mean (SD)	24.6 (8.0)	24.7 (5.8)	16.3 (2.2)	23.2 (7.5)
Range	16–55	16–36	12–20	16–55
Sex				
Male	30 (81.1)	8 (66.7)	5 (50.0)	43 (72.9)
Female	7 (18.9)	4 (33.3)	5 (50.0)	16 (27.1)
Diagnosis†				
Psychotic disorder	25 (67.5)	7 (58.4)	1 (10.0)	33 (55.8)
Mood disorder	8 (21.6)	5 (41.7)	3 (30.0)	16 (27.1)
Autism spectrum disorder	4 (10.8)	2 (16.7)	5 (50.0)	11 (18.6)
Other disorders	1 (2.7)	0 (0.0)	4 (40.0)	5 (8.4)
Multiple diagnosis	4 (10.8)	2 (16.7)	9 (90.0)	15 (25.4)
Substance abuse	21 (56.8)	6 (50.0)	1 (10.1)	28 (47.5)
Ethnicity: Dutch	30 (81.1)	10 (83.3)	9 (100)	49 (83.0)
Involuntary commitment	11 (29.7)	0 (0.0)	4 (40)	15 (25.4)

†More than one diagnosis possible. SD, standard deviation.

for each patient. Table 3 reveals that there were missing scores because the first four subscales should have been scored 78 times and the subscales 5–8 should have been scored 65 times. Those missing scores were excluded from analysis.

The interrater reliability for the total instrument was 0.79 as measured with both ICC and Pearson (Table 3). The ICC and Pearson values were similar on all subscales. All correlation coefficients of the subscales were more

than 0.60, except for the subscales social skills (0.56) and ancillary impairment (0.06). The mean difference between the scores of two nurses for the same patients differs among units; especially on these subscales, there are outliers in the data from unit C (Fig. 1). Therefore, we repeated the analysis, leaving out the data from unit C.

TABLE 3: Interrater reliability of the Kennedy Axis V used by nurses

	Interrater reliability	
	ICC	Pearson
1. Psychological impairment ($n = 78$)	0.78	0.78**
2. Social skills ($n = 78$)†	0.56	0.56**
3. Violence ($n = 77$)	0.69	0.69**
4. ADL-occupational skills ($n = 76$)	0.71	0.71**
5. Substance abuse ($n = 62$)	0.79	0.79**
6. Medical impairment ($n = 62$)	0.79	0.79**
7. Ancillary impairment ($n = 56$)‡	0.06	0.06
8. Motivation for treatment ($n = 64$)	0.78	0.79**
Total ($n = 553$)§	0.79	0.79**

** $P < 0.01$. †When corrected for the data from unit A and only data from unit A and B is used ($n = 69$), ICC 0.64, Pearson 0.65**. ‡when corrected for the data from unit A and only data from unit A and B is used ($n = 52$), ICC 0.18, Pearson 0.18. §376 ratings were conducted on unit A, 124 ratings on unit B, and 53 ratings on unit C. ADL, activities of daily living; ICC, intraclass correlation coefficient; n , number of patients scored twice by nurses.

For the subscale social skills, this resulted in an ICC of 0.64 and Pearson 0.65. Both correlation coefficients for the subscale ancillary impairment are 0.18.

Clinical utility

From the total sample of 23 nurses, 20 nurses (87%) completed the clinical utility questionnaire (Table 4). All 20 nurses (100%) believed that patients could benefit if the staff would frequently use risk assessment instruments. The mean time of scoring the eight subscales was 15.2 min and, when nurses had scored four times or more, it took 11.6 min to score the instrument and to write a short report. Fourteen nurses (73.7%) considered this time needed to score the instrument unacceptable because they had other priorities and tasks and experienced little profit from using the instrument, as they explained in the questionnaire. Eighteen nurses (90.0%) had trouble scoring one or more subscales because they found the content description of the subscales not clear enough. On unit C, the nurses could not link the content of the manual to the adolescent population. Thirteen nurses (72.2%) could not relate the scores with accompanying interventions to reduce the risk factors. Fourteen nurses (70.0%) found that the instrument provided useful information for the (nursing) care plan. Thirteen nurses (72.2%) considered

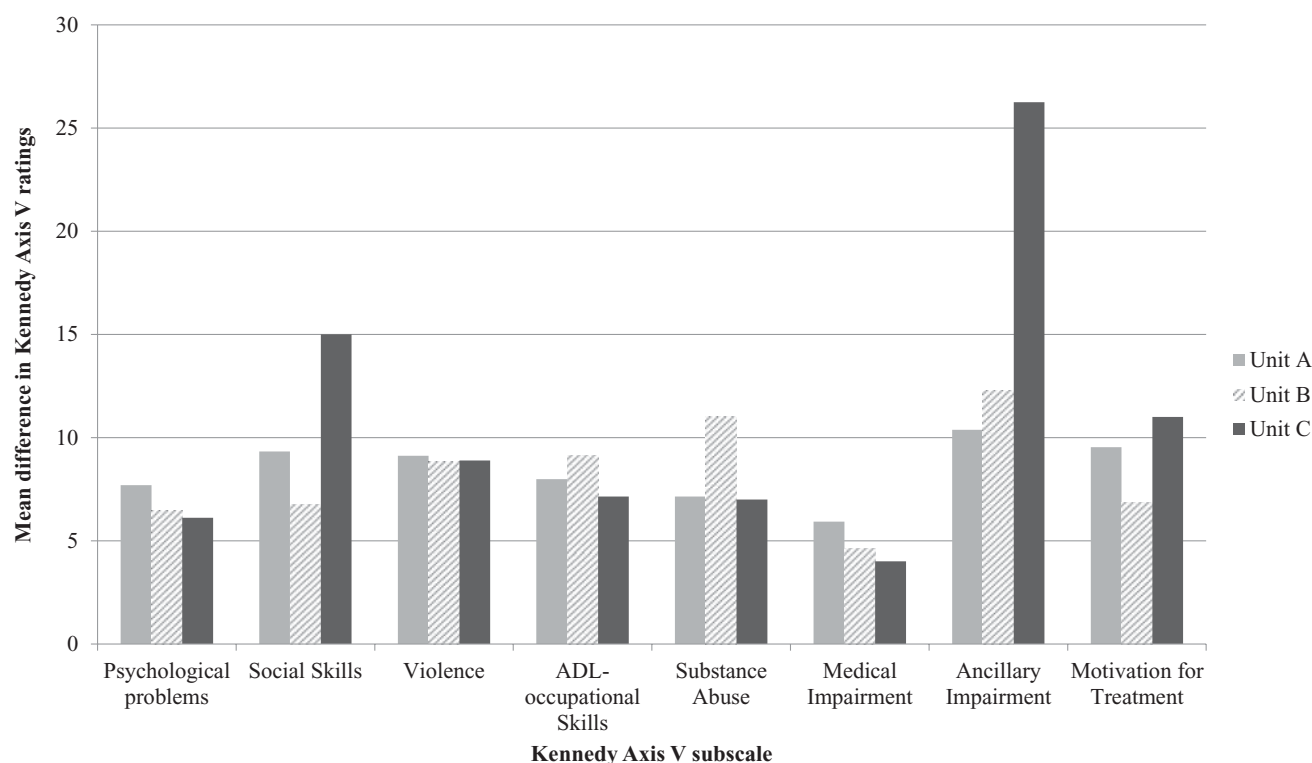


FIG. 1: Mean differences in nurses' ratings on the subscales among units

TABLE 4: *Results of the clinical utility questionnaire of the Kennedy Axis V used by nurses*

Items on the clinical utility questionnaire	Total (n = 20)	
	Mean (SD)	n (%)
1. Do you think patients benefit from the use of a risk assessment instrument?		
Yes		20 (100)
No		0 (0.0)
2. Is the instrument is easy to score without extra activities?		
Yes		8 (40)
No		12 (60)
3. How much time does it take to score subscales 1–4?		
Time in min	6.3 (3.1)	
Time when scored <10 times (n = 9)	7.8 (2.6)	
Time when scored ≥10 times (n = 10)	4.9 (2.9)	
4. How much time does it take to score subscales 1–8?		
Time in min	5.2 (9.2)	
Time when scored <4 times (n = 8)	17.9 (11.2)	
When scored ≥4 times (n = 6)	11.6 (4.2)	
5. Do you consider the time needed to score to be acceptable?		
Yes		5 (26.3)
No		14 (73.7)
6. Do you think the subscales are clear?		
Yes		10 (50.0)
No		10 (50.0)
7. Are the subscales easy to score?		
Yes:		2 (10.0)
No:		18 (90.0)
7a. Which subscales do you have trouble scoring:		
(i) Psychological impairment		7 (38.9)
(ii) Social skills		5 (27.8)
(iii) Violence		12 (66.7)
(iv) ADL–occupational skills		5 (27.8)
(v) Substance abuse		5 (27.8)
(vi) Medical impairment		2 (11.1)
(vii) Ancillary impairment		8 (44.4)
(vii) Motivation for treatment		3 (16.7)
8. Is the instrument easily scored?		
Yes		5 (26.3)
No		14 (73.7)
9. Do you know which interventions to use according to the score?		
Yes		5 (27.8)
No		13 (72.2)
10. Is the instrument easy to use (layout, scores)?		
Yes		14 (70)
No		6 (30)
11. Does the instrument provide information that is clinically useful for the (nursing) care plan?		
Yes		14 (70.0)
No		6 (30.0)
12. Do you think the instrument is useful within:		
Nursing verbal report		6 (33.3)
Nursing written report		13 (72.2)
Daily medical verbal report		4 (23.5)
Weekly multidisciplinary meeting		12 (70.6)

ADL, activities of daily living; SD, standard deviation.

the instrument useful within the nursing written report and 12 nurses (70.6%) considered the instrument to be useful for the weekly multidisciplinary meetings.

Focus group interviews on the adolescent unit

A sample of 12 nurses from the adolescent unit participated in this study, of whom six nurses (50%) participated in the focus group interviews. Four thematic categories emerged from the interviews (Fig. 2). The nurses had difficulties in distinguishing the psychiatric impression from developmental and childhood issues. The nurses found that they were insufficiently trained because formal training was provided only once on unit C and they needed more training on the job. Moreover, they experienced that the Kennedy Axis V manual did not provide enough information for them on how to rate the scale in an age-specific way. The nurses also reported positive experiences. Some described the anchor points in the instrument to observe violent behaviour and found the instrument efficient to use.

DISCUSSION

This is one of the few studies that investigates the interrater reliability and clinical utility of the Kennedy Axis V from a mental health nursing perspective in a sample of adolescent patients and adults with psychotic disorders. The overall interrater reliability of the Kennedy Axis V (0.79) was found to be close to outstanding. Among the challenges experienced, the nurses found it difficult to score the instrument and use interventions related to the scores. On unit C, the nurses found it difficult to use the instrument for the adolescent population.

There are some limitations to this study. The study was relatively small because there were only two wards participating and, even though almost all nurses were included in the study, the number of nurses remains low. Also, because the adolescent unit dropped out of the study, there is little data on interrater reliability on this unit and the results, considering that clinical utility data were collected after using the instrument for only 1 month. This limitation may also be discussed as a strength of this study because different types of wards were involved. Although the adolescent unit was not able to finish the study, additional data were collected exploring the nurses' opinion of using the instrument with adolescents. As far as we know, this aspect has not been studied before. The results of this study indicate promising options to generalize the findings to comparable adult psychiatric units. However, the data on clinical utility revealed that there are many unit-specific factors

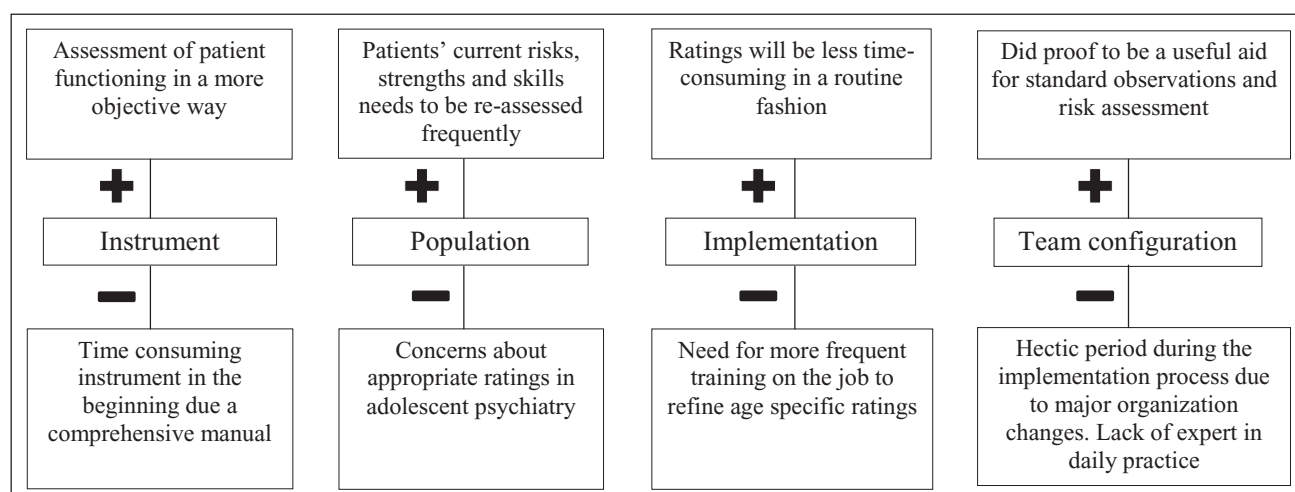


FIG. 2: Results of the group focus group interviews on the adolescent unit C.

contributing to the clinical utility of an instrument, such as previous education of nurses and the number of staff. We learned that clinical utility is essential for an adequate implementation process and it is therefore recommended that the instrument is evaluated for every specific unit prior to implementation.

The interrater reliability found in this study is comparable with the findings of Mundo *et al.* (2010), where the subscales social skills and ancillary impairment also have the lowest agreement (respectively, ICC 0.633 and 0.667). However, in the current study the interrater reliability of the subscale ancillary impairment is much lower than these earlier findings and this may possibly be caused by a lack of information about the patient by the nurses or disagreement considering which problems should be rated. Pearson's *r* from the study of Ebben (2006) and Bilezikian (1998) was higher than in the current study. However, because of the small sample size ($n = 24$) in the study of Ebben (2006) and the inappropriateness of Pearson's *r* for measuring interrater reliability, these results could overestimate the actual agreement (Bland & Altman 1986). In this study, a larger sample size was used and the correlation coefficients were similar, indicating that our findings reflect the actual agreement (Polit 2010). Furthermore, little is clear about how the scoring procedures took place in previous studies. The raters were possibly more experienced using the scales but these articles did not mention this.

This study has several implications for the use of the instrument in adult mental health-care facilities. The nurses on the adult units A and B felt that they captured useful clinical information with the Kennedy Axis V but

sometimes struggled with scoring and were unsure which complementary interventions they should use related to the assessed level of patient functioning. These challenges during the implementation processes have also been reported in other structured risk assessment studies (Watts *et al.* 2004; Webster & Hucker 2006). Various authors have emphasized that implementation of risk assessment instruments is a time-consuming process which needs careful planning and the use of multifaceted, tailored implementation strategies (Grol & Grimshaw 2003; Hoge 2010). Several more specific strategies are therefore recommended to adequately implement the Kennedy Axis V in adult mental health-care facilities:

1. Multiple training sessions and workshops which address how to refine the nursing care plan supported by an assessment tool are highly recommended. The need for adequate training is also emphasized in British risk management best practice evaluations (Department of Health 2007). Continuous education in this area is needed as structured risk assessment is often lacking in basic and post-graduation nurse education curricula (McKenna *et al.* 2003). Moreover, risk assessment is considered especially challenging for junior registered nurses because they have experienced few risk scenarios and are in the process of learning to prioritize their observations (Whitley *et al.* 1996). This legitimates prolonged and intensive coaching for junior nurses in the course of the career trajectory, but it is also essential to emphasize the mastery of risk assessment scales in the basic and post-Bachelor education of nurses.

2. Structured and individual de-escalation plans can facilitate the use of interventions for risk management. These plans should incorporate the specific cut-off scores and corresponding preventive interventions. The Kennedy Axis V cut-off scores for dangerousness were not used in this study but these could facilitate the use of tailored interventions when a patient reaches a certain high-risk level cut-off rating. It should be noted that the relation between risk assessment and risk management is difficult with most instruments (Gilbert *et al.* 2011). Like most assessment scales, the Kennedy Axis V manual also exclusively focuses on assessment issues rather than recommending corresponding interventions. However, in this study, the Kennedy Axis V treatment plan handbook (Kennedy 2003c) was not incorporated into the training because treatment planning was not the aim of this study. When the scores of the Kennedy Axis V are adequate, valid, and reliable, the use of treatment plans incorporating ratings should be considered.
3. Planning which nurses score which patients should prevent missing scores. The missing scores appear to be caused by nurses who may not have been adequately informed about the present mental state of patients or have not been able to observe the patient more specifically when they started rating. This should not be a problem when nurses have specific patients under their supervision or otherwise nurses should inform each other more in detail to enable adequate Kennedy Axis V scores. This phenomenon also indicates the importance of being aware of what the quality of patient–nurse interaction has been during the shift. This is specifically important with the ancillary impairment subscale because nurses should be informed about the present mental state of patients and their finances, legal, and housing issues, which are required. We assume that the nurses in this study were generally not well informed about these issues which may have caused lower correlation coefficients on this subscale.
4. Select an adequate timeframe to score patients on each unit. We learned that on the outpatient unit B, the daily scores change minimally. On the other hand, we identified that the ratings did fluctuate frequently on the acute units and that the ratings can really support short-term decisions.
5. Careful planning of the implementation period, including conducting a pilot test where some patients are rated. We learned that nurses needed some time to adequately use the instrument and therefore it may be necessary to plan more than 3 months for the implementation.

During this study, the Kennedy Axis V was used alongside the BVC (Almvik & Woods 2003), which had been implemented on the units prior to this study and was used structurally. Although this was not the scope of this study, we recommend the use of both instruments because of their complementary qualities. For scoring the Kennedy Axis V, information about the background of the patient is required. The BVC can be scored using observations only and is appropriate during the first days of hospitalization (Almvik & Woods 2003). Van de Sande *et al.* (2011b) has presented further details concerning the combination of these instruments.

The author of the instrument (Kennedy 2003a) claims that the instrument can be used for all patients including children over 5 years of age. However, we believe that this needs to be studied in a broader sense because the results on the adolescent unit C show that the nurses experienced problems with using the instrument for adolescents. The Kennedy Axis V captures violence (to self and others) in a dynamic way and provides a clinical impression of the level of self-control regarding anger, which is a developmental challenge in the adolescent phase. Adolescents admitted to acute psychiatric units are in the process of developing self-control competencies that they did not have before. Using the findings of the focus group interviews, the rating problems need to be explored more specifically in the future with use of case vignettes and refined training.

There is a need for similar studies on a larger scale. More specifically, all clinimetric properties of the instrument should be structurally evaluated in multicentre longitudinal studies after the instrument is adequately implemented using the recommended strategies. Especially, the predictive validity of the instrument should be evaluated by comparing the dangerousness scores with patients' outcomes, such as violence, seclusion, suicide, self-harm, self-neglect, and recovery, to learn more about the instrument and its use for short-term risk assessments. Because of the problems encountered in the implementation process, the effects of various implementation strategies should be studied to maximize effectiveness and cost-efficiency.

CONCLUSION

This study has shown that the Kennedy Axis V has substantial interrater reliability when used by mental health nurses. However, accurate and thorough implementation of the instrument is essential because of the challenges experienced with scoring the instrument. Another common challenge with risk assessment is the adequate use of specific interventions related to scores which was

also encountered in this study. This challenge may be addressed by: (i) specific structured and continuous education for nurses; (ii) structured incorporation of the instrument in the treatment planning meetings on the units; and (iii) the use of multidisciplinary risk management plans with preventive interventions. Considering that the low amount of data from the adolescent unit is a limitation of this study, the rating problems on the adolescent units need to be explored more specifically in the future in terms of specific case vignettes and refined training. The Kennedy Axis V requires a solid implementation strategy and it takes time for nurses to become accustomed to scoring. However, when the instrument is adequately implemented, it may help with objectivity in clinical decision-making. The Kennedy Axis V can be used for short-term risk assessment and thereby contribute to a reduction of severe self-neglect, suicide, self-harm, and violence.

ACKNOWLEDGEMENTS

We would like to thank the nurses and medical staff of the participating units for their contribution and efforts with the use of the instrument and data collection process.

REFERENCES

- Almvik, R. & Woods, P. (2003). Short term risk prediction: The Broset Violence Checklist. *Journal of Psychiatric and Mental Health Nursing*, 10, 231–238.
- Almvik, R., Woods, P. & Rasmussen, K. (2000). The Broset Violence Checklist: Sensitivity, specificity, and interrater reliability. *Journal of Interpersonal Violence*, 15, 1284–1296.
- American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Health Disorders*, 4th edn. Washington DC: American Psychiatric Association.
- Amore, M., Menchetti, M., Tonti, C. *et al.* (2008). Predictors of violent behaviour among acute psychiatric patients: Clinical study. *Psychiatry and Clinical Neurosciences*, 62, 247–253.
- Bilezikian (1998). In: Kennedy, J.A. *Mastering the Kennedy Axis V: A New Psychiatric Assessment of Patient Functioning*. (pp. 2.30). Arlington, TX: American Psychiatric Publishing inc.
- Bland, J. M. & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 8476, 307–310.
- Blake, C. S. & Hamrin, V. (2007). Current approaches to the assessment and management of anger and aggression in youth: A review. *Journal of Child and Adolescent Psychiatry*, 20, 209–221.
- Blok, G. T., de Beurs, E., Ranitz, A. G. S. & Rinne, T. (2010). Psychometrische stand van zaken van risicotaxatie-instrumenten voor volwassenen in Nederland. *Tijdschrift Voor Psychiatrie*, 52, 331–341.
- Bonnet, D. G. (2002). Sample size requirements for estimating intraclass correlations with desired precision. *Statistics in Medicine*, 21, 1331–1335.
- Brendel, R. W., Wei, M. H. & Edersheim, J. G. (2010). An approach to the patient in crisis: Assessments of the risk of suicide and violence. *Medical Clinics of North America*, 940, 1089–1102.
- Department of Health (2007). Best practice in managing risks. [Cited 20 Jun 2011]. Available from: URL: http://webarchive.nationalarchives.gov.uk/+www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_076512.pdf
- Duxbury, J. & Whittington, R. (2005). Causes and management of patient aggression and violence: Staff and patient perspectives. *Journal of Advanced Nursing*, 50, 469–478.
- Duxbury, J., Hahn, S., Needham, I. & Pulsford, D. (2008). The Management of Aggression and Violence Attitude Scale (MAVAS): A cross-sectional comparative study. *Journal of Advanced Nursing*, 62, 596–606.
- Ebben, C. (2006). Validiteit en betrouwbaarheidsonderzoek naar de Kennedy As V. [Cited 30 Sep 2010]. Available from: URL: http://oathesis.eur.nl/ir/repub/asset/4111/13944_495d68ed3439c78ae5c9a30ca6609155.pdf
- Elo, S. & Kyngäs, H. (2007). The qualitative content analysis process. *Journal of Advanced Nursing*, 62, 107–115.
- Fleenor, J. W., Fleenor, J. B. & Grossnickle, W. F. (1996). Interrater reliability and agreement of performance ratings: A methodological comparison. *Journal of Business and Psychology*, 10, 367–380.
- Fluttert, F., van Meijel, B., Webster, C., Nijman, H., Bartels, A. & Grypdonk, M. (2008). Risk management by early recognition of warning signs in patients in forensic psychiatric care. *Archives of Psychiatric Nursing*, 22, 208–216.
- Foster, C., Bowers, L. & Nijman, N. (2007). Aggressive behaviour on acute psychiatric wards: Prevalence, severity and management. *Journal of Advanced Nursing*, 58, 140–149.
- Georgieva, I., Vesselinov, R. & Mulder, C. L. (2012). Early detection of risk factors for seclusion and restraint: A prospective study. *Early Intervention in Psychiatry*. Early view ahead of publication.
- Gilbert, E., Adams, A. & Buckingham, C. D. (2011). Examining the relationship between risk assessment and risk management in mental health. *Journal of Psychiatric and Mental Health Nursing*, 18, 862–868.
- Grol, R. & Grimshaw, J. (2003). From best evidence to best practice: Effective implementation of change in patients' care. *Lancet*, 32, 1225–1230.
- Hage, S., van Meijel, B., Fluttert, F. & Berden, G. F. M. G. (2009). Aggressive behaviours in adolescent psychiatric settings: What are risk factors, possible interventions and implications for nursing practice? A literature review. *Journal of Psychiatric and Mental Health Nursing*, 16, 661–669.

- Harris, A. W. F., Large, M. M., Redoblado-Hodge, A., Nielssen, O., Anderson, J. & Brennan, J. (2010). Clinical and cognitive associations with aggression in the first episode of psychosis. *The Royal New Zealand College of Psychiatrists*, 44, 85–93.
- Harris, M. D. & Warren, J. J. (1995). Patient outcomes: Assessment issues for the CNS. *Clinical Use Specialist*, 9, 82–86.
- Hoge, R. D. (2010). Youth level of service/case management inventory. In: R. K. Otto & K. S. Douglas (Eds). *Handbook of Violence Risk Assessment*. (p. 86). New York: Routledge.
- Husum, T. L., Bjørngaard, J. H., Finset, A. & Ruud, T. (2010). A cross-sectional prospective study of seclusion, restraint and involuntary medication in acute psychiatric wards: Patient, staff and ward characteristics. *BMC Health Services Research*, 10, 1–9.
- Hyde, S., Fullbrook, P., Fenton, K. & Kilshaw, M. (2009). A clinical improvement project to develop and implement a decision-making framework for the use of seclusion. *International Journal of Mental Health Nursing*, 18, 398–408.
- Inspectorate of Health (2011). *Terugdringen separeren stagneert, normen vereist rondom insluiting psychiatrische patiënten. Uitkomsten inspectieonderzoek naar de preventie van separeren 2008–2011*, pp.37. [Cited 28 Mar 2012]. Available from: URL: http://www.igz.nl/Images/2011-12a%20Rapport%20incl.%20bijlagen%20Terugdringen%20separeren%20psychiatrische%20patienten_tcm294-315452.pdf
- Janssen, W., van de Sande, R., Noorthoorn, E. *et al.* (2011). Monitoring the use of restrictive measures: Methodological issues in data collecting, analysis and outcome. *International Journal of Law and Psychiatry*, 34, 419–428.
- Kaltiala-Heino, R., Tuohimäki, C., Korkeila, J. & Lehtinen, V. (2003). Reasons for using seclusion and restraint in psychiatric inpatient care. *International Journal of Law and Psychiatry*, 26, 139–149.
- Kelly, T., Simmons, W. & Gregory, E. (2002). Risk assessment and management: A community forensic mental health practice model. *International Journal of Mental Health Nursing*, 11, 206–213.
- Kennedy, J. A. (2003a). *Mastering the Kennedy Axis V: A New Psychiatric Assessment of Patient Functioning*. Arlington, TX: American Psychiatric Publishing inc.
- Kennedy, J. A. (2003b). *Kennedy Axis V*. [Cited 4 Apr 2012]. Available from: URL: <http://www.ct.gov/dmhas/LIB/dmhas/MRO/Kaxis.pdf>
- Kennedy, J. A. (2003c). *Fundamentals of Psychiatric Treatment Planning*. Arlington, TX: American Psychiatric Publishing inc.
- Keski-Valkama, A., Sailas, E., Eronen, M., Koivisto, A., Lönnqvist, J. & Kaltiala-Heino, R. (2010). The reasons for using restraint and seclusion in psychiatric inpatient care: A nationwide 15-year study. *Nordic Journal of Psychiatry*, 64, 136–144.
- Kumar, S. & Simpson, A. I. F. (2005). Application of risk assessment for violence methods to general adult psychiatry. *Australian and New Zealand Journal of Psychiatry*, 39, 328–335.
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Lodewijks, H. P. B., Doreleijers, T. A. H., de Ruiter, C. & Borum, R. (2008). Predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) during residential treatment. *International Journal of Law and Psychiatry*, 31, 263–271.
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35, 382–385.
- McKenna, B. B., Poole, S. J., Smith, N. A., Coverdale, J. H. & Gale, C. K. (2003). A survey of threats and violent behaviour by patients against registered nurses in their first year of practice. *International Journal of Mental Health Nursing*, 12, 56–63.
- McNiel, D. E. & Binder, R. L. (1995). Correlates of accuracy in the assessment of psychiatric inpatients' risk of violence. *American Journal of Psychiatry*, 152, 579–586.
- Morse, J. & Field, P. (1986). *Nursing Research, the Application of Qualitative Approaches, Second Edition*. (p. 26). Cheltenham: Nelson Thornes Publishers.
- Mulder, C. L. (2000). *Kennedy As V: Vertaling En Bewerking*. Rotterdam: Erasmuscentrum.
- Mundo, E., Bonalume, L., Del Corno, F., Madeddu, F. & Lang, M. (2010). L'applicazione dell'Asse V di Kennedy a un campione clinic italiano. *Rivista Di Psichiatria*, 45, 214–220.
- National Institute for Clinical Excellence (2005). *Violence: The short-term management of disturbed/violent behaviour in psychiatric in-patient settings and emergency departments*. [Cited 23 Feb 2012]. Available from: URL: <http://guidance.nice.org.uk/CG25/NICEGuidance/pdf/English>
- Polit, D. F. (2010). *Statistics and Data Analysis for Nursing Research*. Upper Saddle River, NJ: Pearson Education.
- Polit, D. F. & Beck, C. T. (2008). *Nursing Research. Generating and Assessing Evidence for Nursing Practice*. 8th edn. Philadelphia, PA: Lippicott Williams & Wilkins.
- de Ruiter, C. (2007). Wat risicotaxatie wel en niet is. *Tijdschrift Voor Psychotherapie*, 33, 142–149.
- van de Sande, R. (2007). Het beoordelen en bewaken van acute gezondheidsbedreigende risico's. In: M. Jansen, M. Kuiper, J. Ettema & R. van de Sande (Eds). *De Expertverpleegkundige*. (pp. 128–150). Houten: Bohn Stafleu van Loghum.
- van de Sande, R., Nijman, H. L. I., Noorthoorn, E. O., van der Staak, C. & Mulder, C. L. (2011a). Risicotaxaties in de voorspelling van separaties: Uitkomsten van dagelijkse scoringen. *Tijdschrift Voor Psychiatrie*, 53, S–62.
- van de Sande, R., Nijman, H., Noorthoorn, E. *et al.* (2011b). Aggression and seclusion on acute psychiatric wards: Effects of short term risk assessment. *British Journal of Psychiatry*, 6, 1–234.
- Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations: Using in assessing rater reliability. *Psychological Bulletin*, 86, 420–428.
- Steinert, T., Lepping, P., Bernhardsgrütter, R. *et al.* (2010). Incidence of seclusion and restraint in psychiatric hospitals: A literature review and survey of international trends. *Social Psychiatry and Psychiatric Epidemiology*, 45, 889–897.

- Terwee, C. B., Bot, S. D. M., Boer, M. R. *et al.* (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, 60, 34–42.
- Tunde-Ayinmode, M. & Little, J. (2004). Use of seclusion in a psychiatric acute inpatient unit. *Australasian Psychiatry*, 12, 347–351.
- Watts, D., Bindman, J., Slade, M., Holloway, F., Rosen, A. & Thornicroft, G. (2004). Clinical assessment of risk decision support (CARDS): The development and evaluation of a feasible violence risk assessment for routine psychiatric practice. *Journal of Mental Health*, 13, 569–581.
- Webster, C. D. & Hucker, S. J. (2006). *Violence Risk Assessment and Management*. West Sussex: John Wiley & Sons.
- Webster, C. D., Douglas, K. S., Eaves, D. & Hart, S. D. (1997). *HCR-20: Assessing the Risk for Violence*. (Version 2). Vancouver: Mental Health, Law, and Policy Institute, Simon Fraser University.
- Webster, C. D., Hucker, S. J. & Bloom, H. Y. (2002). Transcending the actuarial versus clinical polemic in assessing risk for violence. *Criminal Justice and Behaviour*, 29, 659–665.
- Whiting, P., Rutjes, A. W. S., Dinnes, J., Reitsma, J. B., Bossuyt, P. M. M. & Kleijnen, J. (2004). Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technology Assessment*, 25, 1–234.
- Whitley, G. G., Jacobson, G. A. & Gawrys, M. T. (1996). The impact of violence in the health care setting upon nursing education. *Journal of Nursing Education*, 35, 211–218.
- Woods, P. & Ashley, C. (2007). Violence and aggression: A literature review. *Journal of Psychiatric and Mental Health Nursing*, 14, 652–660.