International Journal of Law and Psychiatry 64 (2019) 205–210

Contents lists available at ScienceDirect

International Journal of Law and Psychiatry

journal homepage: www.elsevier.com/locate/ijlawpsy

Structured versus unstructured judgment: DUNDRUM-1 compared to court decisions

Inge Jeandarme[a,*], Petra Habets[a], Harry Kennedy[b]

[a] Knowledge Centre for Forensic Psychiatric Care (KeFor), OPZC, Rekem, Belgium
[b] National Forensic Mental Health Service, Central Mental Hospital, Dundrum, Ireland

ARTICLE INFO

Keywords:
Forensic psychiatry
Court decision
Security level
DUNDRUM-1

ABSTRACT

Background: Criteria to determine in which level of security forensic patients should receive treatment are currently non-existent in Belgium. Courts largely rely on the evaluations of the prison psychiatrists and psychologists to form their decision. None of the few available instruments – e.g., the DUNDRUM-1 – is currently used to provide structured clinical judgment when determining security level.
Method: DUNDRUM-1 scores were collected for 150 forensic patients. Security levels according to DUNDRUM-1 assessment were compared to security levels as decided by the court.
Results: There was little agreement between DUNDRUM-1 scores and proposals for secure care made by the court. The DUNDRUM-1 predicted eventual admission to a high security setting, but not a medium security setting.
Conclusion: The DUNDRUM-1 is an instrument that can help clinicians and judges to make more reliable and transparent decisions regarding secure care. However, further research with regard to practical applicability is needed.

1. Introduction

According to the law on the internment of persons (Act of 5 May 2014, modified by the Potpourri III Act of 4 May 2016), in Belgium, in case of insanity (Not Guilty by Reason of Insanity, NGRI), an internment measure of indefinite duration can be imposed by investigation and judgment courts. An internment measure is a security measure with a twofold purpose, namely to protect society and to permit compulsory psychiatric treatment for the internees. The Chambers for the Protection of Society (part of the tribunal for the execution of sentences) are responsible for the execution of the internment measure. These Chambers (further referred to as "courts") decide under which conditions and security level treatment will be delivered, either under the judicial statute of "placement" or under conditional release. Remand prisoners who receive an NGRI verdict remain in so called 'psychiatric annexes' within prison, until the court decides on the appropriate treatment setting. The courts are thereby advised by the psychosocial team of the prison. In Flanders, placement occurs in a Forensic Psychiatric Center (FPC Gent or FPC Antwerp) or in a government-recognized facility (e.g., a general psychiatric hospital), with whom a placement agreement has been concluded (Casselman, De Rycke, & Heimans, 2015). The first FPC was implemented in 2014. Placement in an FPC is mandatory, i.e. the FPC cannot refuse to admit a patient, if a bed is available. Internees can also be treated while they are conditionally released from their internment measure under the strict conditions of ambulatory or residential forensic psychiatric or general psychiatric treatment (Moens & Pauwelyn, 2012). Treatment settings all have admission and exclusion criteria related to psychopathology, risk and security level. On December, 1th, 2017, 3421 persons had an internment measure in Belgium (Seynnaeve, Goyens, & Dheedene, 2018). Less than 20% of these internees resided in prison. Belgium has been censured several times by the European Court of Human Rights because of the imprisonment of internees that do not receive the appropriate care they need. More than 80% of the internees received treatment either in an in- or outpatient forensic or general psychiatric setting. Assessing which security level a forensic patient needs in order to safely undergo treatment is an important first step in the forensic recovery process. The security level should be compatible with the risks and needs of the patients and proportionate to the seriousness of the risks. Incorrect assessment of this can have a serious impact on society, on forensic services as well as on the patients. It is surprising that clear criteria and procedures to support this decision-making process are currently lacking in Belgium. Courts and judges therefore are inclined to rely on their "common sense" or unstructured judgment.

* Corresponding author.
E-mail addresses: Inge.jeandarme@opzcrekem.be (I. Jeandarme), Petra.habets@opzcrekem.be (P. Habets), kennedh@tcd.ie (H. Kennedy).

https://doi.org/10.1016/j.ijlp.2019.04.006
Received 28 December 2018; Received in revised form 26 March 2019; Accepted 18 April 2019
Available online 01 May 2019

1.1. Court decisions

Clinicians provide expert opinion about treatment options, security level and risk of recidivism in forensic patients to review boards, parole boards and other legal decision makers. However, despite decades of research on topics such as risk assessment, scientific results seldom find their way into the courtroom. The question can thus be asked about how reliable, valid and transparent these decisions are in practice (Stredny, Parker, & Dibble, 2012). Like other human beings, judges use their "common sense" or "common understanding" as part of the judicial decision making process (Burns, 2016). However, decision-making short-cuts (heuristics) and cognitive biases can unconsciously have an impact on decision making. For example, the availability heuristic suggests that a risk is perceived as more serious if it can be readily called to mind (Burns, 2016). Kahneman and Klein (2009) have contrasted the naturalistic decision making approach to research and development of expert decision making with the so-called heuristics and biases approach. They concluded that noisy and highly complex situations are difficult to predict and algorithms may outperform humans in such low-validity situations because of their advantage of consistency. Structured professional judgment instruments can enhance the quality of the clinician's judgment as measured by consistency and reliability, to ensure that scientifically valid items are not forgotten, to make the decision-making process transparent and to reduce the chance of serious error.

Canadian research indicated that few empirically based risk factors (such as in the Historical Clinical Risk Manangement-20; HCR-20 (Webster, Douglas, Eaves, & Hart, 1997)) were mentioned during the hearing process or in the reasons cited for the disposition (Cote, Crocker, Nicholls, & Seto, 2012). Likewise, Wilson, Crocker, Nicholls, Charette, and Seto (2015) found that only a subset of HCR-20 or Violence Risk Assessment Guide (VRAG; Quinsey, Harris, Rice, & Cormier, 2006) risk factors were used by review boards in decisions and risk management factors were seldom discussed. Although there was no association with empirically based risk factors (e.g., no association with VRAG scores), there was a strong, positive correlation of tribunal decision with senior clinical testimony (Hilton & Simmons, 2001). Although some studies indicated a trend in the appropriate direction (Hilton, Simpson, & Ham, 2016; McKee, Harris, & Rice, 2007), most studies suggest that empirically based risk factors and risk assessment instruments are not actively used in courts' dispositions. In addition, risk assessment cannot simply be equated with the determination of a security level (Kennedy, O'Neill, Flynn, Gill, & Davoren, 2016). Security refers to an essential aspect of forensic care and consists of three components: (1) material security, (2) procedural security and (3) relational security (Kennedy, 2002). Relational security closely matches the individual risk profile of the patient, but is not entirely compatible with it. *Risk* is defined as the degree of violent tendencies and the risk of (violent) recidivism which can be low, moderate or high. This risk is usually measured using structured risk assessment instruments. Risk is a dynamic concept that can fluctuate over time, situation and context. The security level, on the other hand, relates to the infrastructural security level of an institution or department (low, moderate or high), is therefore static and related to factors other than the risk level of the patient. A patient has a low, moderate or high *security need*, while the department provides the *security level* (e.g., an escape-resistant building, high relational security). Security need relates to the seriousness of the risk and related factors, rather than probability (Eckert, Schel, Kennedy, & Bulten, 2017; Flynn, O'Neill, & Kennedy, 2011; Kennedy, 2002; Kennedy et al., 2016; Scott, 1977). Relational security characteristics are related to quality of care (e.g., quality of the therapeutic relationships, usage of risk management strategies) and include higher patient-to-staff ratios, intensive treatment programs and specific therapeutic skills to deal with patients with severe mental disorders at risk of committing further violence. Examples of procedural security are supervised visits and limited access to contraband such as drugs, weapons or pornography, or cash money (Kennedy, 2002).

Research in Belgium into the determination of security levels is minimal. Jeandarme (2016) investigated the profile of the patients admitted to a medium security unit (MSU) in Flanders and found that the demographic and diagnostic profiles were largely comparable with the profile of the MSU populations in the international literature. Surprisingly, the judicial antecedents and the scores on risk assessment instruments were rather concordant with scores of patients with a high security need. These findings were confirmed in another study. In a comparison between a Walloon high security population and Flemish MSU populations, higher risk profiles were found among the MSU population (Pham et al., 2019). However, there is an inherent problem of circularity by trying to determine patient characteristics for a specific type of treatment unit or by determining the level of security by examining those people who are staying in those units at that time (De Page et al., 2018). In this respect, there remains a need for clear definitions and research into the various aspects of security levels within forensic psychiatry in Flanders.

1.2. Security levels

Internationally, to our knowledge, only three instruments are available to measure security level: The Security Needs Assessment Profile (SNAP; Collins & Davies, 2005), the Health of the Nations Outcome Scale-Secure (HoNOS-Secure; Royal College of Psychiatrists, 2013) and the Dangerousness Understanding, Recovery and Urgency Manual V1.0.30 (DUNDRUM-toolkit) (Kennedy et al., 2016). According to the review by Shinkfield and Ogloff (2015), only the DUNDRUM meets the desired criteria for routine outcome measurement for the aforementioned instruments. Keulen-de Vos and Schepers (2016) concluded in their review that of the four instruments they found suitable for mapping out needs in forensic patients (Camberwell Assessment of Need - Forensic version, DUNDRUM toolkit, HoNOS-Secure and International Resident Assessment Instrument - Correctional Facilities) only the DUNDRUM toolkit focuses specifically on the allocation of a security level to patients. The DUNDRUM toolkit is a set of structured professional judgment instruments that can be used to support structured professional judgment (SPJ) in determining security level (DUNDRUM-1), admission priority (DUNDRUM-2), treatment progress (DUNDRUM-3) and degree of recovery (DUNDRUM-4). The DUNDRUM SPJ instruments are qualitatively different from risk assessment instruments, as they measure complimentary domains (Kennedy et al., 2016). Since the current study focusses on security level, only DUNDRUM-1 will be used.

1.3. Current study

Literature reviews have shown promising results regarding psychometric properties of the DUNDRUM. Furthermore, the decision process is clearly outlined in the instrument. It is however an Irish instrument, until recently not validated in a Flemish population. A previous study using the same population as included in the current study found that the final judgment of the DUNDRUM-1 had an excellent Interrater Reliability (IRR) (Habets, Jeandarme, & Kennedy, 2019), which was in line with an Irish study (Flynn, O'Neill, McInerney, & Kennedy, 2011). The internal consistency of the DUNDRUM-1 was slightly lower (Cronbach's Alpha = 0.67; Jeandarme & Habets, 2019) compared with international studies (Adams, Thomas, Mackinnon, & Eggleton, 2018: $\alpha = 0.84$; Flynn, O'Neill, McInerney, & Kennedy, 2011: $\alpha = 0.95$; Freestone et al., 2015: $\alpha = 0.77$). In comparison to the HoNOS-Secure, both the IRR and internal consistency were more favorable for the DUNDRUM (Jeandarme & Habets, 2019). In the current follow-up study, the DUNDRUM-1 is compared with court decisions.

2. Method

2.1. Study outline

This retrospective cohort study ($N = 150$) combines two datasets using 1) a random selection of 50 male patients referred for admission to a medium security unit (OPZC Rekem) by the court (sample 1) and 2) a random selection of 100 male patients referred to various levels of therapeutic protection by the court (sample 2). The files from sample 1 were scored blind for current placement, but not for placement on admission. The files from sample 2 were blindly scored for court decision. For all files, the DUNDRUM-1 was scored by the second author (PH). The files consisted of psychosocial reports during detention, psychiatric insanity evaluation reports and in some cases also treatment reports.

2.2. Participants

The mean age of the participants was 41.5 years ($SD = 11.26$, range 21–71 years). The majority had Belgian nationality (86.6%) and were single or divorced (87.8%). Patients were found not guilty by reason of insanity and received a treatment order (internment measure) after committing physical violence (65.4%), or other offenses such as theft or stalking (34.7%). The primary diagnosis in 56.7% was an Axis I major mental disorder (in 90.6% of cases a psychotic disorder), in 29.3% a personality disorder and in 14% another psychiatric diagnosis (e.g., intellectual disability). Substance misuse was present as a co-morbid disorder in more than half of the sample (51.7%). Personality disorders were present as a co-morbid disorder in 46.2%. Most of the co-morbid personality disorders were cluster B disorders (92.5%); a minority were cluster A (6%) or not specified (1.5%).

2.3. Outcome measures

2.3.1. DUNDRUM-1

The DUNDRUM-1 consists of 11 Triage Security items, each scored using a 5-point scale ranging from 0 to 4, and is scored prior to admission. The instrument follows the method of SPJ, and provides support in clinical decision making when determining the security level. The items are clearly anchored and explicit criteria are matched for each security level. The items are mainly static in nature and measure something that co-varies only to a small extent with the historical items of the HCR-20 (Kennedy et al., 2016). For individual patients, it is recommended to first look at 'pattern recognition' and individual items. In the case of those who mostly score 0, no additional security is required; mostly scores 1 can be treated within an open setting; scores 2 require low security, score 3 medium security and for score 4 high security is necessary. For research purposes, it is recommended to add up the scores and divide them by the number of items. The items relating to risk of suicide (TS2 and TS4) are not included for research purposes, but they are included for the SPJ. The final judgment (total score DUNDRUM-1/9) yields a number between 0 and 4. Related to security level this yields the following classification: 0–0.5 = ambulatory, > 0.5–1 = open, > 1–2 = low, > 2–3 = moderate and > 3 = high. For individual patients this is a guide that does not bind the decision maker or expert. For research at system level the score indicates appropriateness of matching groups to needs.

2.3.2. Court decision

The courts decide under which conditions and security level treatment will be delivered and are advised about this by the psychosocial team of the prison. For the current study, court decisions were classified at following DUNDRUM-1 security level:

0. Ambulatory = ambulatory.
1. Open = unit general psychiatric hospital.
2. Low = unit for drug addiction, unit for patients with intellectual disabilities, unit for patients with severe conduct disorder.
3. Moderate = MSU, MSU for sex offenders, forensic unit for patients with intellectual disabilities.
4. High = FPC, high security unit for women, prison.

This classification was not made based on scientific research, but based on the structural and procedural security that the institutions currently use according to the researcher's assessment. The classification is used in this study to examine the relationship between the court decisions and DUNDRUM-1 scores and is not intended as a definitive classification of security level. Because it is assumed that there is reasonable consensus in practice regarding the concepts "MSU" and "FPC", the predictive validity is only examined for these variables.

2.4. Statistical analyses

SPSS Version 22 was used for the statistical analyses (IBM Corp, 2013). The data were processed anonymously. The IRR was calculated based on Kappa statistics. According to Landis and Koch (1977) these scores can be interpreted as follows: 0 = bad, 0.0–0.2 = weak; 0.21–0.40 = reasonable; 0.41–0.60 = moderate; 0.61–0.80 = strong; and 0.81–1.00 = (almost) perfect match. The non-parametric Kruskal-Wallis test was used to compare mean scores. Post hoc comparisons with the Mann Whitney test were performed using Bonferroni correction. The non-parametric Spearman's rho was used to calculate correlations between instruments in which the classification of Cohen (1988) was used to quantify the magnitude of the correlation: < 0.30 = weak; 0.30–0.50 = average; 0.50–0.70 = strong and > 0.70 = very strong. To evaluate the predictive validity of the DUNDRUM-1 for admission in a medium or high security unit, the Receiver Operating Characteristic (ROC) analysis was used, with the Area Under the Curve (AUC) value reflecting the accuracy of the test. The evaluation of the AUC values was described according to the standards of Rice and Harris (2005) where: AUC ≥ 0.56 = little effect, AUC ≥ 0.64 = moderate effect and AUC ≥ 0.71 = large effect.

3. Results

3.1. Descriptive analyses

Five participants were removed from the analyses because there was either no DUNDRUM-1 score available due to too many missing items ($n = 3$), or no court decision found in the files ($n = 2$). The mean DUNDRUM-1 score ($N = 145$) for the 9 security items was 2.3 ($SD = 0.59$). Further information at item level can be found in Table 1. Subdivided to security level, the recommended security level based on the DUNDRUM-1 indicated a need for high level of security in 8.3% of the cases; in 62.8% it referred to a medium level of security; in 26.2% to low security; and in 2.8% to an open security level. In contrast, the court decisions indicated a high level of security in 35.9% of the cases, a medium level of security in 51.0% of the cases, low security in 3.4%,

Table 1
Descriptive variables DUNDRUM-1.

	n	*M*	*SD*	Range	
TS1. Seriousness of violence	145	2.71	1.15	0	4
TS3. Immediacy of risk of violence	145	2.89	1.17	0	4
TS5. Specialist forensic need	144	2.22	1.05	0	4
TS6. Absconding/eloping	145	2.18	1.08	0	4
TS7. Preventing access	144	2.33	1.05	0	4
TS8. Victim sensitivity / public confidence issues	141	1.35	1.27	0	4
TS9. Complex Risk of Violence	145	2.40	0.79	0	4
TS10. Institutional behaviour	144	1.60	1.12	0	4
TS11. Legal process	139	3.05	1.10	0	4
DUNDRUM-1 total score (9 items)	145	20.52	5.43	6	31

n = sample size; M = mean; SD: Standard Deviation; Range = Minimum and Maximum score.

Table 2
Crosstabulation DUNDRUM-1 recommended security level and court decision.

		Court Decision					
		Ambulatory	Open	Low	Medium	High	Total
DUNDRUM-1 recommended security level	Ambulatory	0	0	0	0	0	0
	Open	0	0	0	2	2	4
	Low	4	4	0	22	8	38
	Medium	5	1	5	48	32	91
	High	0	0	0	2	10	12
	Total	9	5	5	74	52	145

open security level in 3.4% of the cases and ambulatory in 6.2% of the cases. Table 2 shows a cross table of the DUNDRUM-1 recommended security category and the court decision.

3.2. DUNDRUM-1 scores compared to court decisions

3.2.1. Agreement

The correlation between the recommended security category based on the DUNDRUM-1 versus the court decision was analyzed with the Spearman's rho test. Although the test was significant, it was a weak association ($r = 0.25$, $p < 0.05$). In 40% of the cases ($n = 58$), the recommended security category based on the DUNDRUM-1 was in line with the court security level. In 14.5% ($n = 21$) the proposed court security level was lower and in 45.5% of cases ($n = 66$) the court determination was higher than the recommended security category based on the DUNDRUM-1. Court decisions that were higher than DUNDRUM-1 recommended security categories most often differed only in one security level, but in 12 cases there was a difference of 2 ($n = 10$) or 3 points ($n = 2$). For example, the court advised a high-security institution or a medium-security unit, while the patient, according to the DUNDRUM-1, should stay in a low-security or an open facility. The recommended security category levels based on the DUNDRUM-1 versus the court decisions were compared with Kappa statistics. In general, the agreement between the two assessments was bad ($\kappa = 0.06$, $p = 0.15$). As is shown more in detail in Table 2, there was no agreement regarding ambulatory, open or low security settings. In cases where the DUNDRUM-1 indicated a high security level, court decisions largely agreed (83.3%). However, little agreement was found with respect to high security decisions by the court (19.2%). The largest concordance of judgments of court decisions and DUNDRUM-1 was at medium security level (52.7% and 64.9% respectively). Decisions which differed more than two points were analyzed in detail. In 12 files, the court's decision was higher compared to the recommended security category based on the DUNDRUM-1. In two files, no argument or explanation for the court's decision was given. In the other cases, the higher security level was related to the illegal status of the offender ($n = 5$), and to soft obstacles, i.e. either no setting with a lower security level could be found ($n = 3$) or the offender refused to be admitted ($n = 2$). In 10 files, the court's decision was lower compared to the recommended security category based on the DUNDRUM-1. No arguments or explanations for these decisions were given. However, in one file, the decision was not feasible due to soft reasons and in four files, the conditional release was revoked and hence, the court's decision was changed to a higher security level, more concordant with the DUNDRUM-1.

3.2.2. DUNDRUM-1 scores stratified for court decision

In Table 3 the mean DUNDRUM-1 score is stratified per court decision. The mean DUNDRUM-1 scores for the different court decisions were compared with the Kruskal Wallis test, which revealed that overall scores differed (Kruskal Wallis, $p < 0.05$). However, post-hoc tests only show a difference between the mean DUNDRUM-1 score for *high* ($M = 2.53$) and *open* ($M = 1.60$) ($p < 0.01$) and *high* ($M = 2.53$) and *medium* ($M = 2.19$) ($p < 0.01$). These analyses were repeated at the item level. The mean DUNDRUM-1 scores for each item were compared per court decision with the Kruskal Wallis test. Overall the test revealed that items scores differed for item TS 1, TS 3, TS5, TS 6, TS 8 and TS 11 (Kruskal Wallis, $p < 0.05$). The post-hoc comparisons showed that there was a difference between *ambulatory* and *open* ($p < 0.01$) for TS1 and TS2. For TS 6 post-hoc tests show a difference between *open* and *high* ($p < 0.01$) and *medium* and *high* ($p < 0.01$). For TS11 post-hoc tests show a difference between *ambulatory* and *medium* ($p < 0.01$), *ambulatory* and *high* ($p < 0.01$), *ambulatory* and *open* ($p < 0.01$) and *medium* and *high* ($p < 0.01$).

Table 3
DUNDRUM-1 nine-item mean score (recommended security category) stratified per court decision.

	n	*M*	*SD*	Min.	Max.
Ambulatory	9	2.11	0.38	1.67	2.78
Open	5	1.60	0.37	1.22	2.22
Low	5	2.42	0.38	2.11	2.89
Medium	74	2.19	0.57	0.67	3.22
High	52	2.53	0.59	0.86	3.44

n = sample size; M = mean; SD = standard deviation; Min. = minimum; max. = maximum.

3.2.3. Predictive validity DUNDRUM-1

We further examined whether the recommended security category based on the DUNDRUM-1 was predictive of the court decision to admit the patient in a high and a medium security unit by performing ROC analyses. We also repeated the significant analyses for each item. The DUNDRUM-1 predicted admission to a high security unit (AUC = 0.70; $p < 0.05$). Also, as shown in Table 4 at item level, some items were predictive: TS 5 Specialist forensic need (AUC = 0.64; $p < 0.05$), TS 6 Absconding/eloping (AUC = 0.71; $p < 0.05$) and TS11 Legal process (AUC = 0.81; $p < 0.05$). The DUNDRUM-1 score was not predictive for admission to a medium security unit (AUC = 0.37; $p < 0.05$). In fact, the analysis showed a negative predictive validity (i.e., the higher

Table 4
ROC curve DUNDRUM-1.

	n	AUC	*p*	CI	
TS1. Seriousness of violence	145	0.54	0.43	0.44	0.64
TS3. Immediacy of risk of violence	145	0.59	0.08	0.49	0.69
TS5. Specialist forensic need	144	0.64	0.01	0.54	0.73
TS6. Absconding/eloping	145	0.71	0.00	0.62	0.80
TS7. Preventing access	144	0.47	0.50	0.37	0.57
TS8. Victim sensitivity / public confidence issues	141	0.55	0.29	0.45	0.65
TS9. Complex Risk of Violence	145	0.50	1.0	0.40	0.60
TS10. Institutional behaviour	144	0.58	0.10	0.48	0.68
TS11. Legal process	139	0.81	0.00	0.74	0.89
DUNDRUM-1 total score (9 items)	145	0.70	0.00	0.60	0.79

n = sample size; TS = Triage Security; AUC: Area Under the Curve; CI = Confidence Interval.

the DUNDRUM-1 score, the lesser the admission to a medium security unit). Finally, it was examined whether the DUNDRUM-1 score was predictive of a court decision to admit either to a medium or to a high security unit. The DUNDRUM-1 score was not predictive for court decision to admit to a medium or a high security unit ($p = 0.13$).

4. Discussion

The aim of the current study was to expand our knowledge of court decision making, particularly with respect to attributing security level to forensic psychiatric patients. Dispositions should be least restrictive as possible while taking into consideration the need to protect the public from people at risk of committing further violence and the seriousness of the violence. Criteria to determine the level of security are currently non-existent in Belgium, and none of the instruments described above are used in the decision making process. This is surprising because secure care is expensive and there is a high cost associated with incorrect placements for the patient, society and institutions themselves. Although it may be logical that clinicians decide whether someone is included in secure care, there is a lot of discussion about such decisions and even experienced clinicians can reach different conclusions. At this moment, the DUNDRUM-1 (like other clinical judgment aids) provides a best-available basis for decision making around patient placement and can therefore serve as an evaluative framework within which to evaluate the adequacy (consistency) of court outcomes. With this study, we wanted to compare the DUNDRUM-1 to current practices in the Belgian judicial processes. For this purpose, a random selection of court dispositions to different security levels was drawn. In this study, it was investigated to what extent the assessment of the level of need for therapeutic security as determined by the DUNDRUM-1 corresponds with the decisions concerning security level made by the court in Belgium. Our research showed that there is a mismatch between measured security needs and attributed security level, with most patients being referred by the courts to a higher security level.

The mean DUNDRUM-1 score for admission to a medium level of security ($M = 2.19$) is consistent with scores in male medium security patients abroad (UK: $M = 2.5$, Freestone et al., 2015; Australia: $M = 2.1$, Adams et al., 2018; Ireland: $M = 2.6$, Davoren et al., 2015; The Netherlands $M = 2.5$–3.0, Eckert et al., 2017). For the high security units, however, the mean scores in this study were lower and for the other security levels higher than expected (Kennedy et al., 2016).

Compared to court decisions, recommended security category based on the DUNDRUM-1 more often pointed to a lower level of security (26.2% versus 3.4%) and less often to a high level of security (8.3% versus 35.9%). DUNDRUM-1 score only differed between patients who got the decision *high* from the court compared to the people who had received the security level *open* or *medium*. Furthermore, no logical pattern of increasing scores as court security level rises was found for the DUNDRUM-1 scores. This held true even for items such as the seriousness of violent act. There is only a weak association between the court decision and recommended security category based on the DUNDRUM-1. In other words, there is little agreement between assessments based on the definitions in the DUNDRUM-1 and assessments based on decision-making by the court. One could argue that at this point it is not clear whether the court's decision is the gold standard to oppose the DUNDRUM clinically based recommended security category. Consistency in mean DUNDRUM-1 scores for medium security has been observed in many different jurisdictions (Adams et al., 2018; Davoren et al., 2015; Eckert et al., 2017; Flynn, O'Neill, & Kennedy, 2011; Freestone et al., 2015). However, this issue can only be resolved in Belgium with follow-up studies of violence in hospital, absconding and unconfounded long-term re-offending studies. It is not known whether the court classifies the patients in the correct categories. As became evident when examining files that differed in two or three points, it is possible that the court refers a person to a FPC for practical reasons (no residence permit, no motivation, no willingness to accept the patient) or simply for convenience when a bed was available even though this person could be treated in a lower security level. The DUNDRUM item concerning the need to prevent absconding was one of the items that did predict the court disposition. It may be that courts exercise a heuristic bias based on an assessment of a limited number of relevant factors. While the current study methodology did not allow these analyses, it would be highly interesting to conduct more in-depth analyses in future research.

This is also reflected in the high number of high security decisions by the court. Of those high security decisions, only 19.2% received a high security score on the basis of the DUNDRUM-1 (low sensitivity). According to the DUNDRUM-1 final judgment score, the majority (61.5%) could be treated in a medium security institution and in 19.2% of the cases in an even lower security level. Similar figures were found in England where clinicians judged that only 22% of patients in a high security hospital were placed at the correct level (Shaw, Davies, & Morey, 2001).

A much larger agreement is found for the decisions of medium security between the court and DUNDRUM-1 recommended security category. Nevertheless, the DUNDRUM-1 was not predictive for admission in a medium security unit, but moderately predictive of admission in a high security unit. This is probably due to a difference in specificity (38.7% versus 96.8%). No agreement has been found for the lower levels. This may be explained by the classification made from the court decisions. At the time of the study, there was no overview of the different security levels for each institution. For this study, the court decisions were converted into a format that is used by the DUNDRUM-1 and in many jurisdictions. This classification was based on the structural and procedural security that the institutions currently use according to the researcher's assessment. For the high security institutions this classification is clear: prisons, FPCs and the high security unit for women. The medium security units for both general and sexual offenders and forensic units for people with intellectual disabilities received the classification medium security. The classification was more difficult for low security and open settings and will have to be further substantiated in the course of time through scientific research. A clearer distinction between general psychiatric units and forensic units with differing security levels may facilitate transfers between settings.

International research has already shown that the DUNDRUM-1 score differentiated between different levels of therapeutic security in England (Freestone et al., 2015) and in Ireland (Flynn, O'Neill, & Kennedy, 2011), Australia (Adams et al., 2018) and matched TBS units in The Netherlands (Eckert et al., 2017). The forensic landscape there is, of course, different from Belgium and although these results are promising we cannot yet say with certainty that the DUNDRUM-1 criteria for therapeutic security levels also apply in Belgium.

4.1. Limitations of research

The classification of the different institutions with respect to security level used in the current study is regarded as a temporary classification, which can influence the results. The For-Care project is currently working to classify different forensic institutions according to security level. This report is scheduled to be published in summer of January 2019 (Leys et al., 2017).

The court gives a decision based on a report from the psychosocial advisory team, but the advice of the psychosocial team is not binding. It is therefore possible that this advice would result in a better agreement with the DUNDRUM-1 scores. However, analyses showed that there was no difference in the DUNDRUM-1 total score between the group where the court followed the advice versus the group where the court judged that the psychosocial advice was not relevant (results not shown, available on request).

5. Conclusion

The substantiated attribution of the correct level of security in forensic psychiatric patients remains a precarious matter in Belgium. In such low-validity situations, use of a structured clinical judgment instrument can provide support in determining security level. In order to draw conclusions about the relationship between therapeutic security level and DUNDRUM-1 in Belgium, there must first be a clear definition of the various forensic institutions and their corresponding security levels.

References

Adams, J., Thomas, S. D. M., Mackinnon, T., & Eggleton, D. (2018). The risks, needs and stages of recovery of a complete forensic patient cohort in an Australian state. *BMC Psychiatry, 18*(1), 35. https://doi.org/10.1186/s12888-017-1584-8.

Burns, K. (2016). Judges, 'common sense' and judicial cognition. *Griffith Law Review, 25*(3), 319–351.

Casselman, J., De Rycke, R., & Heimans, H. (2015). *Internering: Nieuwe interneringswet en organisatie van de zorg.* Brugge: die Keure.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Collins, M., & Davies, S. (2005). The security needs assessment profile: A multidimensional approach to measuring security needs. *International Journal of Forensic Mental Health, 4*(1), 39–52. https://doi.org/10.1080/14999013.2005.10471211.

Cote, G., Crocker, A. G., Nicholls, T. L., & Seto, M. C. (2012). Risk assessment instruments in clinical practice. *Canadian Journal of Psychiatry, 57*(4), 238–244.

Davoren, M., Byrne, O., O'Connell, P., O'Neill, H., O'Reilly, K., & Kennedy, H. G. (2015). Factors affecting length of stay in forensic hospital setting: Need for therapeutic security and course of admission. *BMC Psychiatry, 15*, 301. https://doi.org/10.1186/s12888-015-0686-4.

De Page, L., Boulanger, M., De Villers, B., Dugauquier, A., Pham, T., Saloppé, X., & Thiry, B. (2018). Perspectives on triage of mentally disordered offenders in Belgium. *Acta Psychiatrica Belgica, 118*(1), 25–30.

Eckert, M., Schel, S., Kennedy, H., & Bulten, E. (2017). Patient characteristics related to length of stay in Dutch forensic psychiatric care. *Journal of Forensic Psychiatry and Psychology, 28*(6), 863–880.

Flynn, G., O'Neill, C., & Kennedy, H. G. (2011). DUNDRUM-2: Prospective validation of a structured professional judgment instrument assessing priority for admission from the waiting list for a forensic mental health hospital. *BMC Research Notes, 4*, 230. https://doi.org/10.1186/1756-0500-4-230.

Flynn, G., O'Neill, C., McInerney, C., & Kennedy, H. G. (2011). The DUNDRUM-1 structured professional judgment for triage to appropriate levels of therapeutic security: Retrospective-cohort validation study. *BMC Psychiatry, 11*, 43. https://doi.org/10.1186/1471-244X-11-43.

Freestone, M., Bull, D., Brown, R., Boast, N., Blazey, F., & Gilluley, P. (2015). Triage, decision-making and follow-up of patients referred to a UK forensic service: Validation of the DUNDRUM toolkit. *BMC Psychiatry, 15*, 239. https://doi.org/10.1186/s12888-015-0620-9.

Habets, P., Jeandarme, I., & Kennedy, H. G. (2019). Applicability of the DUNDRUM-1 in a forensic Belgium setting. *Journal of Forensic Practice*.

Hilton, N. Z., & Simmons, J. L. (2001). The influence of actuarial risk assessment in clinical judgments and tribunal decisions about mentally disordered offenders in maximum security. *Law and Human Behavior, 25*(4), 393–408.

Hilton, N. Z., Simpson, A. I., & Ham, E. (2016). The increasing influence of risk assessment on forensic patient review board decisions. *Psychological Services, 13*(3), 223–231. https://doi.org/10.1037/ser0000068.

IBM Corp (2013). *IBM SPSS statistics for windows, version 22.0.* Armonk, NY: IBM Corp.

Jeandarme, I. (2016). *Medium security units: Recidivism & risk assesment* (PHD)Tilburg: Tilburg University.

Jeandarme, I., & Habets, P. (2019). Determining Security Need in Forensic Psychiatric Patients: HoNOS-Secure compared to DUNDRUM-1 [Bepalen beveiligingsnood van forensisch psychiatrische patiënten: Een vergelijking tussen HoNOS-Secure en DUNDRUM-1]. *Tijdschrift voor Psychiatrie* (in press).

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *The American Psychologist, 64*(6), 515–526. https://doi.org/10.1037/a0016755.

Kennedy, H. G. (2002). Therapeutic uses of security: Mapping forensic mental health services by stratifying risk. *Advances in Psychiatric Treatment, 8*, 433–443. https://doi.org/10.1192/apt.8.6.433.

Kennedy, H. G., O'Neill, C., Flynn, G., Gill, P., & Davoren, M. (2016). *The DUNDRUM toolkit draft V1.0.30.* Dublin: Trinity College Dublin.

Keulen-de Vos, M., & Schepers, K. (2016). Needs assessment in forensic patients: A review of instrument suites. *International Journal of Forensic Mental Health, 15*(3), 283–300. https://doi.org/10.1080/14999013.2016.1152614.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Leys, M., Nicaise, P., Schoenaers, F., Vander Laenen, F., Pans, M., Bourmock, D., ... Meulewaeter, F. (2017). *For-care. Research project on implementation health care trajectories for interned persons: A "realist evaluation" of a reform program in a multisectoral framework (interim report).* In opdracht van de Federale Minister van Volksgezondheid, uitgevoerd door VUB, UCL, ULG en UGent.

McKee, S. A., Harris, G. T., & Rice, M. E. (2007). Improving forensic tribunal decisions: The role of the clinician. *Behavioral Sciences & the Law, 25*(4), 485–506. https://doi.org/10.1002/bsl.768.

Moens, I., & Pauwelyn, L. (2012). *Geen opsluiting, maar sleutels tot re-integratie. Voorstellen voor een gecoördineerd zorgtraject voor geïnterneerden [stop incarceration, give way to reintegration. Proposals for a coordinated health trajectory for forensic psychiatric patients].* Brussel: Zorgnet Vlaanderen.

Pham, T., Habets, P., Saloppé, X., Ducro, C., Delaunoit, D., & Jeandarme, I. (2019). Violence risk profile of medium- and high-security NGRI offenders in Belgium. *The Journal of Forensic Psychiatry & Psychology*. https://doi.org/10.1080/14789949.2019.1570540 (Advanced online publication).

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk* (2nd ed.). Washington, DC: American Psychological Association.

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's *d*, and *r*. *Law and Human Behavior, 29*(5), 615–620. https://doi.org/10.1007/s10979-005-6832-7.

Royal College of Psychiatrists (2013). What is HoNOS? Geraadpleegd via http://www.rcpsych.ac.uk/crtu/healthofthenation/whatishonos.aspx.

Scott, P. (1977). Assessing dangerousness in criminals. *British Journal of Psychiatry, 131*, 127–142.

Seynnaeve, K., Goyens, M., & Dheedene, J. (2018). Internering in een veranderend zorglandschap: wat zijn de vaststellingen na één jaar nieuwe wet op de internering? *Panopticon, 39*(3), 214–250.

Shaw, J., Davies, J., & Morey, H. (2001). An assessment of security, dependency and treatment needs of all patients in secure services in a UK health region. *The Journal of Forensic Psychiatry, 12*(3), 610–637.

Shinkfield, G., & Ogloff, J. (2015). Use and interpretation of routine outcome measures in forensic mental health. *International Journal of Mental Health Nursing, 24*(1), 11–18. https://doi.org/10.1111/inm.12092.

Stredny, R. V., Parker, A. L., & Dibble, A. E. (2012). Evaluator agreement in placement recommendations for insanity acquittees. *Behavioral Sciences & the Law, 30*(3), 297–307. https://doi.org/10.1002/bsl.1995.

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing risk for violence (Version 2).* Burnaby, BC: Simon Fraser University.

Wilson, C. M., Crocker, A. G., Nicholls, T. L., Charette, Y., & Seto, M. C. (2015). The use of risk and need factors in forensic mental health decision-making and the role of gender and index offense severity. *Behavioral Sciences & the Law, 33*(1), 19–38. https://doi.org/10.1002/bsl.2162.