RESEARCH ARTICLE

Development and Validation of a Decision Tool for Early Identification of Adult Patients with Severe and Complex Eating Disorder Psychopathology in Need of Highly Specialized Care

Alexandra E. Dingemans[1]*, Maartje Goorden[2], Freek J.B. Lötters[2], Clazien Bouwmans[2], Unna N. Danner[3,4], Annemarie A. van Elburg[3,4], Eric F. van Furth[1,5] & Leona Hakkaart-van Roijen[2]

[1]Rivierduinen Eating Disorders Ursula, Leiden, The Netherlands
[2]Institute of Health Policy and Management (iBMG) and institute of Medical Technology Assessment (iMTA), Erasmus University Rotterdam, The Netherlands
[3]Altrecht Eating Disorders Rintveld, Zeist, The Netherlands
[4]Department of Clinical and Health Psychology, Utrecht University, Utrecht, The Netherlands
[5]Department of Psychiatry, Leiden University Medical Center, Leiden, The Netherlands

Summary

Patients with complex and severe eating disorders often receive a number of ineffective or/and insufficient treatments. Direct referral of these patients to highly specialized tertiary treatment facilities in an earlier stage of the disorder is likely to be more (cost)-effective. The aim of the study was to develop a decision tool that aids clinicians in early identification of these patients. After identification of criteria that were indicative of severity and complexity of eating disorder psychopathology by means of a systematic review of literature and consultation of a focus group, a Delphi method was applied to obtain consensus from experts on the list of relevant criteria. Finally, the decision tool was validated in clinical practice, and cut-off criteria were established. The tool demonstrated good feasibility and validity to identify patients for highly specialized tertiary care. The final decision tool consisted of five criteria that can easily be implemented in clinical practice. Copyright © 2017 John Wiley & Sons, Ltd and Eating Disorders Association.

Received 23 January 2017; Revised 20 April 2017; Accepted 15 May 2017

Keywords
eating disorder; tertiary care; decision tool; criteria; severe; complex; treatment

***Correspondence**
Alexandra E Dingemans, PhD, Rivierduinen Eating Disorders Ursula.
Email: a.dingemans@rivierduinen.nl

Published online in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/erv.2528

Introduction

Eating disorders are serious and complex psychiatric disorders that often co-occur with other psychiatric disorders (Hudson, Hiripi, Pope, & Kessler, 2007) and that have one of the highest mortality rates of all psychiatric disorders (Arcelus, Mitchell, Wales, & Nielsen, 2011). The course and outcome vary among the type, nature and severity of the eating disorder but can be considered to be unsatisfactory. About 50% recovers fully, 30% shows some improvement without full recovery and 20% of all patients display a chronic course (Steinhausen, 2002; Steinhausen & Weber, 2009; Keel & Brown, 2010). Relapse is also a common problem among individuals with an eating disorder, even after successful treatment (Keel, Dorer, Franko, Jackson, & Herzog, 2005; Field *et al.*, 1997) with most relapse rates falling between 35 and 50% (McFarlane, Olmsted, & Trottier, 2008). However, if treated early in the course of the illness, patients have a higher chance of recovery (Zipfel, Giel, Bulik, Hay, & Schmidt, 2015).

In the Netherlands, the majority of patients are referred to specialized secondary mental health care services by their general practitioner. In cases where secondary care appears to be insufficient, patients are referred to a higher level of specialty care (i.e. tertiary care). For the majority of patients, this stepped care approach seems to be effective and less expensive than immediate referral to high intensity treatment and/or highly specialized treatment programs (Crow *et al.*, 2013; Brown & Keel, 2012). For some patients, however, the stepped care approach seems to be less efficient as there is substantial attrition throughout the referral–assessment–treatment–discharge pathway in patients referred to specialist eating disorder services (Waller *et al.* 2009). Especially for patients with more severe and complex symptoms, it is expected that direct referral to highly specialized programs may be more beneficial compared to the stepped care approach. Early identification of these patients who cannot be managed by general psychiatric (secondary) services and require highly specialized tertiary care could enhance need-based patient stratification. Need-based patient stratification can also reduce costs as the patients with complex and severe disorders often receive a number of ineffective and/or insufficient treatments before they are referred to highly specialized tertiary care programs, where more tailored therapy can be provided. Direct referral of complex and severe cases to highly specialized tertiary

treatment facilities in an earlier stage of treatment is likely to be more (cost)-effective.

Knowledge of criteria to identify patients for highly specialized tertiary treatment is still limited. Hence, there is a need to identify discriminative criteria in patients with eating disorders in order to determine the appropriate treatment pathway in an early stage of the disorder. Although prognostic factors for recovery, relapse and drop-out are described in the literature, it is unknown to what extent these criteria can be effectively used in deciding which treatment pathway is most appropriate. The aim of this study was to develop a decision tool that would aid clinicians in identifying adult patients with severe and complex eating disorder psychopathology who may benefit from highly specialized tertiary care.

Materials and methods

The first step in the development of this decision tool was to carry out a review of the literature in order to identify criteria that are indicative of severity and complexity of eating disorder psychopathology. Subsequently, the criteria that were identified in the literature search were discussed among an independent focus group of experts (psychiatrists, psychologists and researchers) with ample experience in treatment of patients with eating disorders. These experts could add relevant criteria to complement the initial list. In the second step, a Delphi study (Uegaki *et al.*, 2007; Diamond *et al.*, 2014) was performed to develop a consensus-based decision tool for the identification of patients in need of highly specialized tertiary care based on these criteria. Finally, in the third step, the decision tool was validated in clinical practice in order to select the most predictive criteria and to establish cut-off criteria.

Literature review study and focus group

A systematic literature review was conducted in the following electronic databases: Pubmed, PsychINFO and Embase. Studies were selected if they met the following criteria: (i) Language in English or in Dutch; (ii) published between 2002 and 2014; (iii) (systematic) review or meta-analysis; (iv) published in peer reviewed journals and (v) an abstract was available. The following search terms were applied: (eating disorder [MeSH Terms] OR anorexia OR bulimia OR 'binge eating') AND (severity OR classification OR criteria OR 'clinical practice' OR treatment OR outcome OR recovery OR prognosis OR dropout OR co-morbidity OR staging OR 'stages of change') AND ('dutch'[Language] OR 'english'[Language]) AND ('2002'[Date—Publication]: '2014'[Date—Publication]) AND (review OR meta-analysis). The literature search resulted in a list of potential criteria. All relevant criteria were discussed in an independent focus group with six experts (psychiatrists, psychologists and researchers) with ample experience in assessment and treatment of patients with eating disorders. The purpose of presenting the results of the literature search to the focus group was to establish the feasibility of the criteria and their cut-off points for clinical practice. In this focus group, the criteria, and their concomitant cut-off points, were discussed and selected for inclusion in the Delphi study.

Delphi method

In the second step, a Delphi method was applied to obtain consensus from experts in the eating disorder field on the list of relevant criteria. The experts consisted of members of the Dutch Academy for Eating Disorders who had at least 3 years of clinical experience in treating patients with an eating disorder. The method of data collection and data analyses were based on a study performed by Uegaki *et al.* (2007). The Delphi method results in a list of relevant criteria that are quantified according to the level of agreement between raters (Diamond *et al.*, 2014). An online survey was used for data collection. During the first round, the experts responded to questions regarding the relevance of the criteria in identifying patients for highly specialized tertiary care. The main question per criterion was: '*[criterion] can predict the course of the eating disorder over time. Can you indicate the relevance of this criterion in determining whether highly specialized (tertiary) treatment is necessary on a 6 point Likert scale form 0 (completely irrelevant) to 5 (extremely relevant)?*' A criterion was considered relevant if the score given was ≥3. In this case, respondents then indicated for which eating disorder subtype(s) (Anorexia Nervosa (AN), Bulimia nervosa (BN), Eating Disorders NOS (EDNOS) or Binge Eating Disorder (BED)) the criterion was applicable and whether it was relevant for children, adolescents and/or adults (18 years or older). Finally, the respondents provided an appropriate cut-off point. At the end of the first round, experts were given the opportunity to add new criteria. The following four threshold values were used to achieve consensus and to determine the criteria set for the second round in the Delphi study: (i) if less than 50% of the experts rated the criterion as relevant, the criterion was excluded from further consideration; (ii) if 70% or more of the experts rated a criterion as relevant, consensus was reached, the criterion was labelled as relevant and did not need to be re-considered in the second round; and (iii) if 50 to 69% of the experts considered a criterion to be relevant, this criterion was eligible for the second round. Consensus over a cut-off point was reached if 70% of the respondents rated the presented cut-off as being relevant. In the second Delphi round, a summary of the first-round findings was provided to the experts which included a histogram of the first round ratings, the percentage of experts who rated the item as relevant, and the mean and standard deviation of the rating. The experts were then asked to re-rate the relevance (0–5) of this reduced set of criteria.

Demographic characteristics of the respondents were assessed, i.e. age, profession, number of years working in their current profession and number of years of experience in treating patients with an eating disorder. In both rounds, the experts were requested to submit their answers within a two-week period. After one week, a reminder was sent out to those who had not responded.

Validation

In the pilot study, the decision tool was tested among a small group of patients with an eating disorder in three separate eating disorder centres. This was followed by a larger validation study in six eating disorder centres in order to select the best predictive criteria for the identification of patients in need of highly specialized tertiary care. The pilot study aimed to explore the use of the decision tool, before undertaking a large-scale study. Based on the pilot data, minor changes were made in the wording of some of the items, resulting in the final version of the decision

tool. For the pilot and the validation study, the decision tool was filled out by a clinician during the intake phase. Additional information was requested about the clinician (i.e. profession, number of years working in the eating disorder field). In the absence of a gold standard test to stratify ED patients to highly specialized care, the experts' clinical judgement constituted the reference standard in assessing the criterion validity. This clinical judgement was made during the clinical staff meeting. Subsequently, each treatment centre checked whether personal information was removed from each decision tool form. The anonymous forms were sent to an independent researcher who was not involved in the data collection. Approval by a Medical Ethics Committee was not required. The study conforms to the provisions of the Declaration of Helsinki.

Data analysis

For the pilot study and the validation study, databases were built using Excel 2010. In the pilot study, similarity between the criteria and the clinical decision was examined by calculating the percentage of criteria that were checked positive on the decision tools where clinicians indicated that a patient should be referred to a highly specialized treatment. In the validation study, a multilevel model was applied, using R, that employed the criteria on the decision tool as independent variables (present/not present) and clinical judgement (tertiary care yes/no) as a dependent variable. A binomial family of function was used with the logit link function. The correlation structure was 'exchangeable'. In addition, it is important that a model discriminates between observations with low and a high risk. This is described in more detail below (criterion validity: sensitivity and specificity). The validity and reliability of the decision tool were investigated. The internal validity was investigated by checking whether the criteria listed on the decision tool were representative of what the instrument aimed to measure. To assess the criterion validity, a receiver-operating characteristic (ROC) curve was constructed by using the multilevel model. Experts' clinical judgement constituted the criterion standard. From the ROC analysis, sensitivity and specificity were generated and area under the curve (AUC) was calculated. Sensitivity is the ability of the instrument to identify patients that belong to highly specialized tertiary care. Specificity is the ability to identify those patients that do not belong to highly specialized tertiary care. A threshold was derived from the multilevel model to determine the optimal number of criteria as a cut-off.

Results

Literature review study and focus group

Applying the basic search MesH term 'Eating disorders' and the above mentioned filters resulted in 1594 hits. Combining these search terms with MesH terms yielded 524 hits. After reviewing the title and abstract, 61 potential reviews were further screened for their eligibility on the following terms: Prognostic factors and risk factors for drop-out or relapse. Sixteen systematic reviews appeared to be useful. Additional criteria were added by the focus group. This resulted in 25 criteria and cut-off points which could be indicative of complexity and/or severity of psychopathology in individuals with an eating disorder according to the literature and focus group (see Supporting Information Table 1). The criteria could be classified in: (i) behavioural symptoms; (ii) co-morbid psychiatric disorders as defined in the DSM-IV; (iii) severe health risks due to eating disorder psychopathology; and (iv) social participation and social functioning aspects due to eating disorder psychopathology.

Delphi study

A total of 45 experts were invited to participate in the online Delphi Study, of which 35 (78%) responded to the invitation. Twenty-six respondents (58%) completed the first online survey, of which 20 (77%) participated in the second round. The mean age of the respondents was 47.6 years (SD = 9.0) with a mean of 12.6 years (SD = 8.1) of experience in treating individuals with eating disorders. The total group existed of nurse specialists ($n = 3$; 12%), psychiatrists ($n = 9$; 35%), psychologists/psychotherapists/clinical psychologists ($n = 11$, 42%) and others ($n = 3$; 12%).

After the first round, consensus was achieved on 14 of the initial 25 criteria (see Supporting Information Table 2), of which seven were rated as relevant and 7 as not relevant. In addition to the remaining 11 criteria, three new criteria were added in the second round (e.g. (a) previous treatment(s) for eating disorder, (b) two or more comorbid DSM-IV axis-I or axis-II disorders, and (c) stagnation of growth). Of these 14 criteria included in the second round, 11 criteria were considered to be relevant by 70% or more of the respondents. Two criteria (i.e. stagnation of growth and a poor parent–child relationship) were only relevant for children and adolescents. The literature review revealed no clear distinction between adolescents (13–18 years) and adults (18+) with regard to prognostic factors and risk factors for drop-out or relapse. Also, very few studies included children (age 8 to 12). Therefore, we decided to focus the criteria for the decision tool on adults. Thus, the criteria 'stagnation of growth' and 'a poor parent-child relationship' were removed from the criteria list as they are only applicable to children and adolescents. Finally, the criterion 'two or more comorbid axis-I or axis-II disorders ' was rated as more relevant (i.e. a higher mean score) than the separate criteria for depressive disorder, obsessive compulsive personality disorder and borderline personality disorder and was therefore included instead of the separate disorders in the final criteria list. Overall, the ratings after re-assessment in round two were higher than the ratings in round one (see Supporting Information Table 2).

In Table 1, the proposed cut-off points of these 13 criteria are also presented. Consensus over a cut-off point was reached if 70% of the respondents rated the presented cut-off as being relevant. A consensus was not achieved on the cut-off points for all of the criteria (i.e. frequency of self-induced vomiting, frequency of laxative abuse, readiness to change, and frequency of binge eating episodes). For these criteria, we used the cut-off points that were rated most frequently by the respondents (i.e. 7 times a week for frequency of vomiting (60%), the use of 49 pills for laxative abuse per week (69%) and 7 times a week for the frequency of binge eating episodes (54%)). To sum up, after two Delphi rounds, a total of 13 criteria were considered to be relevant for distinguishing patients with an eating disorder with severe and complex psychopathology from those with milder and less complex psychopathology (see Table 1).

Table 1 Mean score (of 6-point Likert scale) of those criteria that achieved consensus is achieved, including scores on the relevance of the cut-off points for each criteria. And the eating disorder and age category for which the criterion is applicable

	Criteria	Score* (mean)	Cut-off points	% relevance for cut-off†	AN	BN	BED	EDNOS
1	Earlier treatment(s) without remission	4.30	1 or more inefficient treatment trajectories	n/a‡	X	X	X	X
2	Extremely low BMI	4.26	≤15 kg/m†	70.0%	X			
3	Severe health risks	4.14	Pulse < 45 beats/min	92.9%		X		X
			Electrolyte disorder	92.9%				
			Hypertension	50.0%				
			High level of cholesterol	46.4%				
			Several issues like: heartbeat disorder, hypotension or disturbed blood pattern§	57.1%				
4	Duration of the eating disorder	3.77	>12 months 2–5 years§	36.7%	X	X	X	X
				76.7%				
5	Frequent vomiting	3.71	7 times per week	60.0%	X	X		X
			>10 times a week§	46.7%				
6	Two or more comorbid Axis-I or Axis-II disorders	3.50	≥2 As-I of As-II disorders	n/a‡	X	X	X	X
7	Somatic co-morbidity	3.46	Diabetes	96.0%	X	X	X	X
			Inflammatory bowel diseases	80.0%				
			Heart and vascular diseases	76.0%				
			Endocrine anomaly	72.0%				
			Food allergies	60.0%				
			Severe invalidating somatic disorder	52.0%				
8	Disordered eating pattern	3.26	Restrictive	96.6%	X	X	X	X
			Excessive	69.0%				
			Less than half of recommended	58.6%				
9	Misuse of laxatives	3.16	49 pills per week 20–100 pills§	69.0%	X	X		X
				44.8%				
10	Little to no readiness to change	3.07	Pre-contemplation phase combined	58.6%	X	X	X	X
			with severe disorder/health risk§	58.6%				
11	Complex use of psychiatric medications	2.82	2 psychiatric medications	73.9%	X	X	X	X
			>3 psychiatric medications§	43.5%				
12	Frequent objective binge eating episodes	2.77	7 times a week 14 to 21 times a week§	53.6%		X	X	X
				50.0%				
13	Extremely high BMI	2.65	≥40 kg/m†	81.5%			X	X

AN = anorexia nervosa; BN = bulimia nervosa; BED = binge eating disorder; EDNOS = eating disorder not otherwise specified: BMI = body mass index; n/a = not applicable.

*Mean score over all respondents from the Delphi study (see Supporting Information Table 2); the higher the score the more relevant.

†Consensus about the cut-off is achieved when 70% of the experts rated it as relevant; answers on the open questions are also taken into account.

‡Submitted additionally in the second Delphi round. These items already contained a cut-off point in their description.

§Cut-offs that were mentioned additionally by the experts.

Validation study

Participants

The results of the pilot study to test the decision tool were discussed in a meeting in which errors and inconsistencies were identified and the tool was subsequently adjusted. Next, a larger validation study was performed in which six centres participated (four specialized (secondary care) and two highly specialized centres (tertiary care) for eating disorders). For the highly specialized centres, mean age and total years of experience of the clinicians was 36.1 (SD = 7.8) and 9.3 (SD = 5.1) years, respectively and for the specialized centres 43.3 (SD = 12.6) and 17.3 (SD = 12.7) years, respectively. Therapists differed significantly on mean years of experience between the two types of centres ($t(30.18) = -2.75$, $p < 0.01$). In total, 576 patients (246 (43%) in highly specialized and 330 (57%) in specialized centres) were enrolled in this study. Patients were diagnosed by experienced psychiatrists and clinical psychologists at intake, which revealed that 31 patients did not have an eating disorder (according to DSM criteria) and were subsequently removed from the analyses. Of the final sample ($n = 532$), 165 (29%) fulfilled the criteria for AN, 87 (15%) for BN, 97 (17%) for BED and 183 (33%) for EDNOS. The majority was female with only 53 men (9.2%). Patients in the specialized secondary centres were significantly older than those in the highly specialized tertiary centres ($M = 31.3$ (SD = 11.5) and $M = 29.2$ (SD = 10.2), respectively; $t(527.28) = -2.37$, $p = 0.02$).

Internal validity

The internal validity of the tool was warranted given the use of a systematic method, the Delphi method, to organize the input of experts in the eating disorder field regarding the inclusion and exclusion of criteria.

Criterion validity

Multilevel model. A multilevel model revealed five criteria (i.e. previous treatment for eating disorder ($\beta = 0.73$ (SE = 0.31), $p = 0.02$), extremely low (<15) or high BMI (>40) ($\beta = 0.14$ (SE = 0.20), $p = 0.04$), duration of illness ($\beta = 1.10$ (SE = 0.25), $p < 0.001$), two or more comorbid Axis-I or Axis-II disorders ($\beta = 1.00$ (SE = 0.26), $p < 0.001$), restrictive or chaotic eating pattern ($\beta = 1.27$ (SE = 0.15), $p < 0.001$) which significantly predicted whether, according to the clinicians, patients were in need of highly specialized treatment. Extremely low (<15) and extremely high BMI (>40) were combined into one variable. Severe health risks ($\beta = 0.61$ (SE = 0.45), $p = 0.18$), frequency of vomiting ($\beta = -0.36$ (SE = 0.22), $p = 0.10$), somatic comorbidity ($\beta = -0.36$ (SE = 0.27), $p = 0.18$), use of psychiatric medications (>2) ($\beta = 0.80$ (SE = 0.59), $p = 0.17$), misuse of laxatives ($\beta = 0.25$ (SE = 0.50), $p = 0.62$), little to no readiness to change ($\beta = 0.43$ (SE = 0.61), $p = 0.48$) and frequency of objective binge eating episodes ($\beta = -0.05$ (SE = 0.22), $p = 0.84$) appeared not to be significant predictors.

Receiver-operating characteristic curve. Using the multilevel model, two ROC curves were plotted: one for all of the variables and one for only the five significant variables (see Figure 1). The area under the curve (AUC) was almost equal for both models; for all predictors, it was 0.80 (95% CI: 0.75–0.85) and for only the significant predictors: 0.78 (95% CI: 0.72–0.83). The best specificity and sensitivity for the predictions were 0.80 and 0.71 (respectively) for the total dataset and 0.81 and 0.63 (respectively) for the five predictors only. For reasons of implementation, a simple rating system was chosen, by only adding the five significant criteria. The final version of the decision tool consists of five criteria which are (i) previous treatment for eating disorder; (ii) extremely low (<15) or high BMI (>40); (iii) duration of illness ≥2 years; (iv) two or more comorbid Axis-I or Axis-II disorders; and (v) highly restrictive or chaotic eating pattern. Different thresholds were used for the number of criteria that had to be present for a patient to be considered in need of highly specialized treatment. For every threshold, both the sensitivity and specificity were determined. Fulfilling three of the five criteria appeared to have to best sensitivity and specificity (see Figure 2).

Discussion

The aim of this study was to develop a decision tool for early identification of adult patients with severe and complex eating

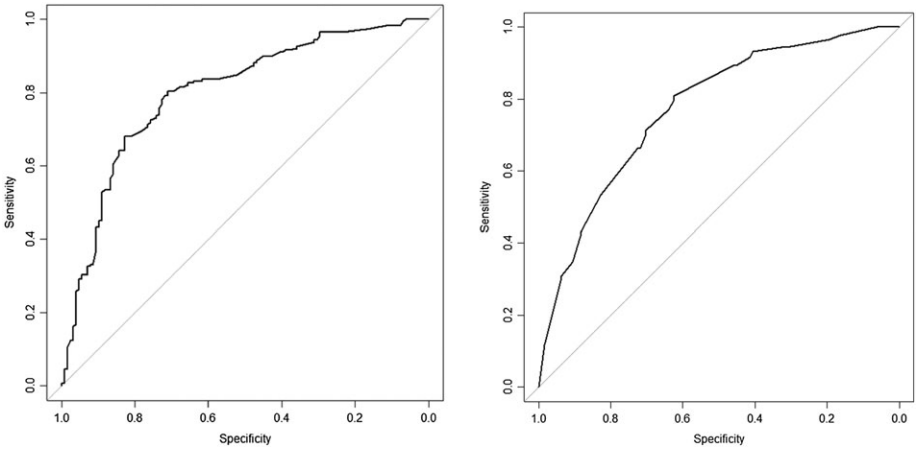

Figure 1. ROC curve multilevel model total (left); ROC curve only with significant predictors

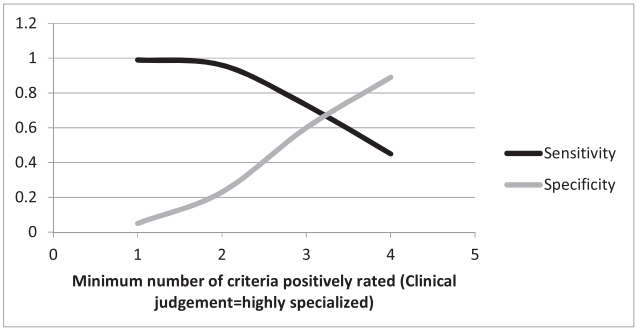

Figure 2. Sensitivity and specificity in relationship to cut off score

disorders in order to direct them to highly specialized tertiary treatment programs. The final version of the decision tool consists of five criteria which are (i) previous treatment for eating disorder; (ii) extremely low (<15) or high BMI (>40); (iii) duration of illness >2 years; (iv) two or more comorbid Axis-I or Axis-II disorders; and (v) highly restrictive or chaotic eating pattern. These criteria can be applied transdiagnostically to all adults with an eating disorder (Fairburn, Cooper, & Shafran, 2003). Fulfilling three of the five criteria appeared to have the optimal trade-off between sensitivity and specificity.

This decision tool may aid in the early identification of patients in need of a highly specialized tertiary treatment facility and thus may aid clinicians in tailoring treatment in an early phase. Specific tailoring of treatment in an early phase of the disorder could improve outcome and increase cost-effectiveness (Zerwas *et al.*, 2013) while decreasing the number of individuals receiving suboptimal or inappropriate therapy (Keel & Brown, 2010; Zerwas *et al.*, 2013). One of the strengths of this study is that the development of the decision tool was evidence based and employed a systematic approach integrating results drawn from the literature with expert opinions assessed *via* a Delphi method and using a validation study with a large number of experts to test the validity of the tool. Another strength of the study is that we recruited the participants from six different eating disorder centres in the Netherlands, thereby increasing the generalizability of the results.

Although this study has many strengths, there are also some limitations. The first limitation is that each criterion was rated separately by the members of the expert panel. However, several experts noted that specific combinations of criteria were of importance to indicate complexity and treatment selection, such as the combination of psychiatric co-morbidity, low BMI and health risks. This is in line with findings of Bruce and Steiger (2005) who stated that specific combinations of psychopathology should be taken into account when treating patients with an eating disorder. They argued that personality disorder (Axis-II) co-morbidities and the related personality factors such as perfectionism, impulsivity and obsessiveness should not be ignored in eating disorder treatment as these factors reinforce each other. Also, degree of readiness to change combined with the severity of eating disorder psychopathology and the concomitant health risks should be taken into consideration (Touyz, Thornton, Rieger, George, & Beumont, 2003). In future studies, the impact of specific combinations of symptoms on the course of the illness should be investigated.

Also, no clear consensus was achieved on the cut-off points of the four criteria (i.e. frequency of self-induced vomiting, frequency of laxative misuse, the readiness to change and frequency of binge eating episodes). For these criteria, we used the cut-off points that were rated most frequently by the experts. Therefore, it is important for future studies to provide more applicable cut-off points for these criteria.

Although the Delphi method is often used to achieve consensus in a structured way, it is not without its limitations. First, it has been stated that this method might be subject to bias because the investigator limits the scope of the items evaluated by the expert panellists (Graham, Regehr, & Wright, 2003). In the current study this limitation was minimized by providing criteria that were derived from a systematic literature review and by allowing the experts to add new criteria, if deemed essential, in the first Delphi round. This resulted in three additional items that were rated during the second round. Second, there were no face-to-face encounters between the experts in a Delphi study. Dynamic group processes between experts may lead to different outcomes with regard to consensus when they interact which each other during a face-to-face discussion (Graham *et al.*, 2003). However, persuasive pleas by well known and highly respected experts in the field can steer the discussion and thereby influence the direction of consensus. Due to the anonymous nature of this Delphi study, all participating experts in the Delphi study had an equal weight in reaching consensus.

Another limitation of the study was the use of a subjective clinical judgement as the criterion standard against which the decision tool was validated. However, in the absence of a standardized clinical tool by which the need for highly specialized care can be assessed, the experts' clinical judgement was considered to be the most accurate and clinically meaningful indicator. Additionally, these clinicians were highly experienced. In the future, a longitudinal validation study should evaluate the predictive validity of the decision tool in clinical practice to investigate whether patients with severe and complex eating disorder psychopathology (according to the decision tool) benefit more from highly specialized tertiary care than from usual specialized secondary care with regard to (eating disorder) psychopathology, cost-effectiveness and quality of life. Furthermore, in a next study the present decision tool would benefit from assessment on convergent validity next to criterion validity. It should be tested whether our decision tool is correlated with other questionnaires designed to measure theoretically similar concepts. Finally, the decision tool is sensitive; however, the specificity is low. The ability of the tool to identify patients that belong to highly specialized tertiary care (sensitivity) was valued higher by clinicians than the ability to identify those patients that do not belong to highly specialized tertiary care (specificity). The tertiary care centres also provide secondary mental health care services in their own region of the country.

This study is the first systematic study of a decision tool aid in the early identification of adult patients in need of highly specialized tertiary care. It resulted in the development of a short decision tool consisting of five criteria that can easily be implemented in clinical practice. This decision tool may aid direct referral of complex and severe cases to highly specialized treatment facilities in an earlier stage of the treatment which is likely to be more (cost)-effective. Future research should

investigate if patients that are in need of highly specialized treatment based on the decision tool and are indeed referred to highly specialized treatment are treated more (cost) effectively compared to the same patients that are not referred to highly specialized treatment. Specific tailoring of treatment in an early phase could improve outcome and be more cost-effective (Zerwas *et al.*, 2013) while decreasing the number of individuals receiving suboptimal or inappropriate therapy.

Acknowledgements

This study was commissioned by the Foundation for Top Clinical Mental Health Care (Stichting TopGGZ) with financial support from the Dutch Innovation Fund of Health Care Insurers (number 2435). We would like to thank Melissa Parks for the English correction. All authors declare no conflict of interest.

REFERENCES

Arcelus, J., Mitchell, A. J., Wales, J., & Nielsen, S. (2011). Mortality rates in patients with anorexia nervosa and other eating disorders. A meta-analysis of 36 studies. *Archives of General Psychiatry*, *68*, 724–731. https://doi.org/10.1001/archgenpsychiatry.2011.74.

Brown, T. A., & Keel, P. K. (2012). Current and emerging directions in the treatment of eating disorders. *Substance Abuse: Research and Treatment*, *6*, 33–61. https://doi.org/10.4137/SART.S7864.

Bruce, K. R., & Steiger, H. (2005). Treatment implications of Axis-II comorbidity in eating disorders. *Eating Disorders: The Journal of Treatment and Prevention*, *13*, 93–108. https://doi.org/10.1080/10640260590893700.

Crow, S. J., Agras, W. S., Halmi, K. A., Fairburn, C. G., Mitchell, J. E., & Nyman, J. A. (2013). A cost effectiveness analysis of stepped care treatment for bulimia nervosa. *International Journal of Eating Disorders*, *46*, 302–307. https://doi.org/10.1002/eat.22087.

Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., et al. (2014). Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, *67*, 401–409. https://doi.org/10.1016/j.jclinepi.2013.12.002.

Fairburn, C. G., Cooper, Z., & Shafran, R. (2003). Cognitive behaviour therapy for eating disorders: A 'transdiagnostic' theory and treatment. *Behaviour Research and Therapy*, *41*, 509–528. https://doi.org/10.1016/S0005-7967(02)00088-8.

Field, A. E., Herzog, D. B., Keller, M. B., West, J., Nussbaum, K., & Colditz, G. A. (1997). Distinguishing recovery from remission in a cohort of bulimic women: How should asymptomatic periods be described? *Journal of Clinical Epidemiology*, *50*, 1339–1345. https://doi.org/10.1016/S0895-4356(97)00220-5.

Graham, B., Regehr, G., & Wright, J. G. (2003). Delphi as a method to establish consensus for diagnostic criteria. *Journal of Clinical Epidemiology*, *56*, 1150–1156. https://doi.org/10.1016/S0895-4356(03)00211-7.

Hudson, J. I., Hiripi, E., Pope, H. G., & Kessler, R. C. (2007). The prevalence and correlates of eating disorders in the national comorbidity survey replication. *Biological Psychiatry*, *61*, 348–358. https://doi.org/10.1016/j.biopsych.2006.03.040.

Keel, P. K., & Brown, T. A. (2010). Update on course and outcome in eating disorders. *International Journal of Eating Disorders*, *43*, 195–204. https://doi.org/10.1002/eat.20810.

Keel, P. K., Dorer, D. J., Franko, D. L., Jackson, S. C., & Herzog, D. B. (2005). Postremission predictors of relapse in women with eating disorders. *American Journal of Psychiatry*, *162*, 2263–2268. https://doi.org/10.1176/appi.ajp.162.12.2263.

McFarlane, T., Olmsted, M. P., & Trottier, K. (2008). Timing and prediction of relapse in a transdiagnostic eating disorder sample. *International Journal of Eating Disorders*, *41*, 587–593. https://doi.org/10.1002/eat.20550.

Steinhausen, H. C. (2002). The outcome of anorexia nervosa in the 20th century. *American Journal of Psychiatry*, *159*, 1284–1293. https://doi.org/10.1176/appi.ajp.159.8.1284.

Steinhausen, H. C., & Weber, S. (2009). The outcome of bulimia nervosa: Findings from one-quarter century of research. *American Journal of Psychiatry*, *166*, 1331–1341. https://doi.org/10.1176/appi.ajp.2009.09040582.

Touyz, S., Thornton, C., Rieger, E., George, L., & Beumont, P. (2003). The incorporation of the stage of change model in the day hospital treatment of patients with anorexia nervosa. *European Child & Adolescent Psychiatry*, *12*(Suppl 1), I65–I71. https://doi.org/10.1007/s00787-003-1109-5.

Uegaki, K., de Bruijne, M. C., Anema, J. R., van der Beek, A. J., van Tulder, M. W., & van Mechelen, W. (2007). Consensus-based findings and recommendations for estimating the costs of health-related productivity loss from a company's perspective. *Scand J Work Environ.Health*, *33*, 122–130. https://doi.org/10.5271/sjweh.1115.

Waller, G., Schmidt, U., Treasure, J., Murray, C., Aleyna, J., Emanuelli, F., et al. (2009). Problems across care pathways in specialist adult eating disorder services. *Psychiatric Bulletin*, *33*, 26–29. https://doi.org/10.1192/pb.bp.107.018325.

Zerwas, S., Lund, B. C., Von, H. A., Thornton, L. M., Berrettini, W. H., Brandt, H., et al. (2013). Factors associated with recovery from anorexia nervosa. *Journal of Psychiatric Research*, *47*, 972–979. https://doi.org/10.1016/j.jpsychires.2013.02.011.

Zipfel, S., Giel, K. E., Bulik, C. M., Hay, P., & Schmidt, U. (2015). Anorexia nervosa: Aetiology, assessment, and treatment. *Lancet Psychiatry*, *2*, 1099–1111. https://doi.org/10.1016/S2215-0366(15)00356-9.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article.