Psychiatry Research 229 (2015) 517–523

Contents lists available at ScienceDirect

Psychiatry Research

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

The impact of revised *DSM-5* criteria on the relative distribution and inter-rater reliability of eating disorder diagnoses in a residential treatment setting

Jennifer J. Thomas [a,b,*], Kamryn T. Eddy [a,b], Helen B. Murray [a], Marilou D.P. Tromp [a], Andrea S. Hartmann [c], Melissa T. Stone [b], Philip G. Levendusky [b,d], Anne E. Becker [a,b,e]

[a] Eating Disorders Clinical and Research Program, Massachusetts General Hospital, 2 Longfellow Place, Suite 200, Boston, MA, USA
[b] Department of Psychiatry, Harvard Medical School, 25 Shattuck Street, Boston , MA, USA
[c] Institute of Psychology, University of Osnabrück, Neuer Graben, 49074 Osnabrück, Germany
[d] Klarman Eating Disorders Center, McLean Hospital, 115 Mill Street, Belmont, MA, USA
[e] Department of Global Health & Social Medicine, Harvard Medical School, 641 Huntington Avenue, Boston, MA, USA

ARTICLE INFO

Article history:
Received 1 October 2014
Received in revised form
5 June 2015
Accepted 25 June 2015
Available online 27 June 2015

Keywords:
DSM-5
Eating disorder
Diagnostic reliability
Other specified feeding or eating disorder (OSFED)
Eating disorder not otherwise specified (EDNOS)

ABSTRACT

This study evaluated the relative distribution and inter-rater reliability of revised *DSM-5* criteria for eating disorders in a residential treatment program. Consecutive adolescent and young adult females ($N=150$) admitted to a residential eating disorder treatment facility were assigned both *DSM-IV* and *DSM-5* diagnoses by a clinician ($n=14$) via routine clinical interview and a research assessor ($n=4$) via structured interview. We compared the frequency of diagnostic assignments under each taxonomy and by type of assessor. We evaluated concordance between clinician and researcher assignment through inter-rater reliability kappa and percent agreement. Significantly fewer patients received either clinician or researcher diagnoses of a residual eating disorder under *DSM-5* (clinician—12.0%; researcher—31.3%) versus *DSM-IV* (clinician—28.7%; researcher—59.3%), with the majority of reassigned *DSM-IV* residual cases reclassified as *DSM-5* anorexia nervosa. Researcher and clinician diagnoses showed moderate inter-rater reliability under *DSM-IV* ($\kappa=.48$) and *DSM-5* ($\kappa=.57$), though agreement for specific *DSM-5* other specified feeding or eating disorder (OSFED) presentations was poor ($\kappa=.05$). *DSM-5* revisions were associated with significantly less frequent residual eating disorder diagnoses, but not with reduced inter-rater reliability. Findings support specific dimensions of clinical utility for revised *DSM-5* criteria for eating disorders.

1. Introduction

The fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-5*) (American Psychiatric Association, 2013) substantially expanded the scope and descriptive classification of eating disorders. Whereas *DSM-IV* included just three eating disorder diagnoses—anorexia nervosa, bulimia nervosa, and a residual category called eating disorder not otherwise specified (EDNOS)—*DSM-5* describes several distinct feeding and eating disorders. In addition to broadening the criteria for anorexia nervosa and bulimia nervosa, *DSM-5* includes binge eating disorder as a formal diagnosis, introduces avoidant/restrictive food intake disorder as a broadened definition of the former *DSM-IV* feeding disorder of infancy or early childhood, and is reorganized to combine the eating and feeding disorders. Furthermore, the previous residual diagnostic category, EDNOS, was reformulated as other specified feeding or eating disorder (OSFED) and unspecified feeding or eating disorder. Specifically, OSFED encompasses five named examples (i.e., atypical anorexia nervosa, subthreshold bulimia nervosa, subthreshold binge eating disorder, purging disorder, and night eating syndrome) as well as additional clinically significant presentations that do not meet full diagnostic criteria for another feeding or eating disorder for a specified reason (i.e., OSFED-other). In contrast, unspecified feeding or eating disorder encompasses "situations in which the clinician chooses not to specify the reason that the criteria are not met for a specific feeding or eating disorder, and includes presentations in which there is insufficient information to make a more specific diagnosis" (American Psychiatric Association, 2013; p. 354). The primary purpose of these revisions was to enhance clinical utility, but critics have raised two primary concerns.

* Corresponding author at: Eating Disorders Clinical and Research Program, Massachusetts General Hospital, 2 Longfellow Place, Suite 200, Boston, MA 02114, USA.
E-mail address: jjthomas@mgh.harvard.edu (J.J. Thomas).

http://dx.doi.org/10.1016/j.psychres.2015.06.017
0165-1781/

First, Fairburn and Cooper (2011) predicted that *DSM-5* revisions would not appreciably reduce the preponderance of residual eating disorders. A core limitation of *DSM-IV* was the reliance on EDNOS to describe the heterogeneous presentations not meeting criteria for full-threshold diagnoses (Keel et al., 2011; Peebles et al., 2010; Thomas et al., 2009). Indeed, recent studies comparing the relative frequency of *DSM-IV* EDNOS to *DSM-5* OSFED have indicated that a substantial minority of individuals may still fall into the residual category. For example, under *DSM-IV*, EDNOS comprised from 45% (Nakai et al., 2013) to 77% (Machado et al., 2013) of eating disorder cases in clinical and community samples, respectively. Within these same samples, OSFED still comprised 26% (Nakai et al., 2013) to 51% (Machado et al., 2013) of eating disorder cases when *DSM-5* criteria were retrospectively applied. In a larger-scale epidemiological study, the lifetime prevalence of residual eating disorders decreased from 64% (*DSM-IV*) to just 16% (*DSM-5*) after retrospective application of *DSM-5* criteria (Smink et al., 2014). However, it remains unknown whether prospectively applying *DSM-5* criteria will yield similar reductions in the comparative frequency of residual presentations. Whereas revised diagnostic criteria for *DSM-5* AN allow clinicians to infer fear of weight gain from collateral history, longitudinal clinical course, or observed behavior, the assessment of cognitive eating disorder symptoms still relies primarily on self-report (Becker et al., 2009). For this reason, symptom ascertainment remains especially challenging in youth due to their less well-developed verbal and abstracting abilities, which may preclude them from endorsing complex phenomena such as overvaluation of shape and weight. Moreover, the correspondence between *DSM-5* OSFED example presentations and those actually observed in clinical practice remains unclear. One recent study of adults with overweight and obesity seeking weight-loss treatment found that approximately one-third of presentations classified as OSFED did not align with one of *DSM-5*'s five named examples (Thomas et al., 2014).

A second potential barrier to the clinical utility of *DSM-5* changes was raised by Frances and Widiger (2012), who expressed concern that the increased complexity of *DSM-5* diagnoses would result in poor inter-rater reliability in everyday clinical practice, where practitioners are unlikely to use structured interviews to confer diagnoses. Only two prior studies (Sysko et al., 2012; Thomas et al., 2014) have evaluated the inter-rater reliability of *DSM-5* eating disorder diagnoses, finding moderate to substantial agreement for the broad categories of anorexia nervosa, bulimia nervosa, binge eating disorder, and OSFED. These findings are consistent with researcher–clinician agreement under *DSM-IV* (Rettew et al., 2009; Thomas et al., 2010). However, studies of *DSM-5* eating disorder prevalence have typically operationalized criteria in much greater detail than is actually specified in *DSM-5*. For example, although *DSM-5* provides no strict cut-off for "significantly low weight" in Criterion A for anorexia nervosa, several studies (Birgegård et al., 2012; Brown et al., 2014; Nakai et al., 2013) have set specific body mass index (BMI) thresholds. Similarly, another study (Stice et al., 2013) operationalized anorexia nervosa Criterion B more stringently than presented in *DSM-5's* revised Criterion B, which now can be met either by fear of weight gain or persistent behavior that interferes with weight gain. Although creating operational definitions is sometimes necessary for research purposes (Brown et al., 2014), their increased specificity may result in higher reliability estimates than implementation of the published criteria would be expected to yield in routine practice. In keeping with this concern, a recent study that examined inter-rater reliability between clinicians' and researchers' diagnostic assessment of individual OSFED examples found poor agreement (Thomas et al., 2014).

Thus, the purpose of the present study was to evaluate two dimensions of clinical utility of *DSM-5* criteria (First et al., 2004): (1) the relative distribution of diagnoses, and (2) the ability or willingness of practicing clinicians to correctly apply highly complex criteria. We assessed the first dimension of clinical utility by comparing the proportion of residual category eating disorder diagnoses generated by clinicians when implementing both *DSM-5* and *DSM-IV* diagnoses in a naturalistic treatment setting. We evaluated the second dimension by examining inter-rater reliability of diagnostic assessments by clinician and research assessors for both *DSM-5* and *DSM-IV* criteria. This concurrent diagnostic assessment also allowed inclusion of questions covering diagnostic concepts not included in *DSM-IV*, such as symptoms specific to avoidant/restrictive food intake disorder. We hypothesized that (1) significantly fewer participants would receive a residual eating disorder diagnosis under *DSM-5* versus *DSM-IV*; and (2) inter-rater reliability between clinician and researcher diagnoses would remain similar under *DSM-IV* and *DSM-5*.

2. Methods

2.1. Participants

Participants included adolescent and young adult females (ages 13–27) presenting to a residential eating disorder treatment center in New England. We invited all 164 patients consecutively admitted to the program during the 16-month data collection period from July 2011 to October 2012 to take part in the study (excluding participants who were re-admitted during this timeframe),

Table 1
Demographic characteristics of 150 adolescent and young adult female study participants in a residential eating disorders program.

Mean age (SD)	18.1 (2.6) years
Adolescents < 18 years old, *n* (%)	65 (43.3)
Participants with BMI < 10th Percentile, *n* (%)	38 (58.5)
Mean BMI centile (SD)	15.9 (20.3)
Adults ≥ 18 years old, *n* (%)	85 (56.7)
Participants with BMI < 18.5, *n* (%)	34 (40.0)
Mean BMI (SD)	20.1 (4.4) kg/m^2
Mean duration of Illness (SD)	4.3 (3.5) years
Mean number of psychiatric hospitalizations (SD)[a]	3.5 (5.8)
Ethnicity, *n* (%)[b]	
Hispanic/Latino	6 (4.0)
Not Hispanic/Latino	140 (93.3)
Race, *n* (%)[b]	
American Indian/Alaska native	2 (1.3)
Black/African American	4 (2.7)
Asian	6 (4.0)
Native Hawaiian/Other Pacific Islander	0 (0)
White	141 (94.0)
Sexual orientation, *n* (%)[c]	
Heterosexual	132 (88.0)
Homosexual	7 (4.7)
Bisexual	8 (5.3)
Other	1 (0.7)
Education level[d], *n* (%)[d]	
Some high school	71 (47.3)
High school graduate without college education	19 (12.7)
Some college education	52 (34.7)
Degree from 4-year college or more	6 (4.0)

Note. BMI = body mass index.

[a] Three participants did not report number of prior psychiatric hospitalizations.

[b] The percentages do not add to 100% because 4 participants did not report ethnicity and 4 did not report race. Furthermore, some participants reported more than one race and/or ethnicity.

[c] Two participants did not report sexual orientation.

[d] Two participants did not report education level.