Received: 6 August 2023 | Revised: 21 December 2023 | Accepted: 22 December 2023

DOI: 10.1002/eat.24133

ORIGINAL ARTICLE

International Journal of EATING DISORDERS WILEY

Cutoff scores of the Eating Disorder Examination–Questionnaire for the German population

Adrian Meule PD Dr[1,2,3] | Anja Hilbert Prof. Dr[4] | Martina de Zwaan Prof. Dr[5] | Elmar Brähler Prof. Dr[4,6] | Stefan Koch Dr[2] | Ulrich Voderholzer Prof. Dr[1,2,7]

[1]Department of Psychiatry and Psychotherapy, LMU University Hospital, LMU Munich, Munich, Germany

[2]Schoen Clinic Roseneck, Prien am Chiemsee, Germany

[3]Institute of Medical Psychology, Faculty of Medicine, LMU Munich, Munich, Germany

[4]Integrated Research and Treatment Center Adiposity Diseases, Behavioral Medicine Research Unit, Department of Psychosomatic Medicine and Psychotherapy, University of Leipzig Medical Center, Leipzig, Germany

[5]Department of Psychosomatic Medicine and Psychotherapy, Hannover Medical School, Hannover, Germany

[6]Department of Psychosomatic Medicine and Psychotherapy, University Medical Center of the Johannes Gutenberg University of Mainz, Mainz, Germany

[7]Department of Psychiatry and Psychotherapy, University Hospital Freiburg, Freiburg, Germany

Correspondence
Adrian Meule, Schön Klinik Roseneck, Am Roseneck 6, 83209 Prien am Chiemsee, Germany.
Email: adrian.meule@med.uni-muenchen.de

Action Editor: Ruth Striegel Weissman

Abstract

Objective: The Eating Disorder Examination-Questionnaire (EDE-Q) is one of the most widely used self-report measures for the assessment of eating disorder (ED) symptomatology. However, proposed cutoff scores that may indicate the presence of an ED have been heterogeneous. Therefore, the current study derived cutoff scores from two large samples: one representative for the German population and one composed of persons with EDs at admission to inpatient treatment.

Method: Receiver operating characteristic analysis was used with the EDE-Q global score as independent variable and group (controls: $n = 2519$, patients: $n = 2038$) as dependent variable. These analyses were also conducted separately with the patient group divided into persons with anorexia nervosa (AN; $n = 1456$), bulimia nervosa (BN; $n = 370$), and other EDs ($n = 212$) and after matching groups for age and sex distribution.

Results: The EDE-Q global score discriminated well between controls and patients (AUC >91%, sensitivity >.84, specificity >.79). A score of 1.6 discriminated best between controls and patients in general and persons with AN in particular. Optimal thresholds for discriminating between controls and persons with BN and other EDs ranged between scores of 1.8 and 2.4.

Discussion: In the German population, cutoff scores between 1.6 and 2.4 may be used to screen for the presence or absence of an ED or evaluate treatment outcome, with slightly higher cutoff scores for persons with BN and other EDs than for persons with AN.

Public Significance: Questionnaire scores have little value when it is unclear which scores indicate the likely presence of an ED, as such scores can be used to estimate the prevalence of or screen for EDs in the general population and evaluate outcome at the end of ED treatment. The current study indicates a score around 2 on the EDE-Q as an optimal threshold for this.

KEYWORDS
anorexia nervosa, assessment, binge-eating disorder, bulimia nervosa, Eating Disorder Examination–Questionnaire, evaluation, screening, sensitivity, specificity

1 | INTRODUCTION

The Eating Disorder Examination-Questionnaire (EDE-Q; Fairburn & Beglin, 1994) is one of the most widely used self-report measures for assessing eating disorder symptomatology (Berg et al., 2012; Meule, 2023). It has 28 items, six of which assess binge and purge behaviors over the past 28 days. The other 22 items assess restrained eating, eating concern, weight concern, and shape concern over the past 28 days on a seven-point scale ranging from 0 (e.g., no days/not at all) to 6 (e.g., every day/extremely). Higher mean global scores of these 22 items (which can, thus, range between 0 and 6) indicate higher eating disorder psychopathology.

The EDE-Q is not a diagnostic measure, that is, global scores do not allow for a clear decision whether a person has an eating disorder or not. However, several cutoff scores have been proposed that indicate that a person might have an eating disorder, but these are quite heterogeneous (Meule, 2021). For example, studies have proposed cutoff scores between 1.1 and 2.4 for men (Rica et al., 2022; Schaefer et al., 2018), while other studies in female or mixed samples suggested scores between 2 and 3 (Melisse et al., 2022; Mond et al., 2004, 2008; Rø et al., 2015; Wade et al., 2021). Yet, other studies applied a score of 4 (e.g., Jennings & Phillips, 2017; Lavender et al., 2010; Luce et al., 2008) based on a study by Mond et al. (2006), in which 4 represented the 95th percentile in a sample of young adult women.

Heterogeneous cutoff scores for the EDE-Q may be explained by at least three factors. First, they have been partially derived in selective, homogeneous samples (e.g., Rica et al., 2022). Second, they have been derived from distribution percentiles but not from receiver operating characteristic (ROC) analysis (e.g., Mond et al., 2006). Third, studies that used ROC analysis may suffer from base rate fallacy (e.g., Rosenfeld et al., 2000). Specifically, the precision of a diagnostic test (positive predictive value) is not intrinsic to the test but also depends on the prevalence of the disorder. That is, a study may show that a certain EDE-Q score has high sensitivity and specificity but may still have low precision for classifying persons correctly when the analyses are based on a large number of persons without but a small number of persons with eating disorders. To our knowledge, no study that has derived cutoff scores for the EDE-Q based on approximately equally sized groups of persons with eating disorders and controls, that is, the control group has been substantially larger than the group of persons with eating disorders in previous studies (e.g., Melisse et al., 2022; Mond et al., 2004; Rica et al., 2022; Rø et al., 2015; Schaefer et al., 2018; Wade et al., 2021).

The current study aimed at deriving cutoff scores for the EDE-Q from a large community sample (nationally representative for the German population in age and sex distribution) and an almost equally large sample of inpatients with eating disorders. Such cutoff scores may be useful when the EDE-Q is used as a screening instrument, for example, when estimating the prevalences of eating disorders in epidemiological studies or for identifying persons with or without eating disorders for inclusion or exclusion in a study. However, samples of persons with eating disorders usually are younger and have a higher proportion of women than samples from the general population and, similarly, women tend to have higher EDE-Q scores than men while age negatively correlates with EDE-Q scores (Hilbert et al., 2012). Therefore, practitioners may also be interested in evaluating EDE-Q scores not only in reference to the general population but also in reference to persons that are more similar to persons with eating disorders regarding age and sex distribution. For example, the EDE-Q score of a young woman at the end of eating disorder treatment may still be high when compared to the general population but may be in the normal range of other young women who do not have an eating disorder. Thus, we also determined cutoff scores for the EDE-Q after matching groups for age and sex distribution.

2 | METHOD

2.1 | Samples

Data of a previously published study with 2520 participants (representative for the German population ≥14 years in terms of age and sex distribution) were used as control group (Hilbert et al., 2012). One participant had to be excluded because of missing EDE-Q data. Descriptive statistics for sex, age, body mass index, EDE-Q global scores, and EDE-Q items on binge and purge behaviors are displayed in Table 1. A more detailed description of the sample and the recruitment procedure can be found in Hilbert et al. (2012).

Data of 2038 persons with eating disorders, who were admitted to inpatient treatment at the Schoen Clinic Roseneck (Prien am Chiemsee, Germany) between 2020 and 2023 constituted the patient group. Descriptive statistics for sex, age, body mass index, EDE-Q scores at admission, comorbid mental disorders, global functioning, and illness duration are displayed in Table 1. Most patients ($n = 1456$) had anorexia nervosa (ICD-10 codes F50.0 [full syndrome, $n = 1241$] or F50.1 [atypical, $n = 215$]), $n = 370$ had bulimia nervosa (ICD-10 codes F50.2 [full syndrome, $n = 305$] or F50.3 [atypical, $n = 65$]), and $n = 212$ had other eating disorders (ICD-10 codes F50.4 [overeating associated with other psychological disturbance, $n = 5$], F50.8 [other, $n = 141$], or F50.9 [unspecified, $n = 66$]; Table 2).

At the hospital, data from the routine diagnostic assessments (e.g., age, sex, diagnoses, body mass index, and questionnaire scores) were automatically transferred to a database from which they could be exported without any identifying information (e.g., name, date of birth, and place of residence) by authorized employees. Thus, accessing individual patient charts was not necessary. According to the

TABLE 1 Sample characteristics as a function of groups.

Variables	Controls		Patients		Test statistics	*p*	Effect sizes
Sex (female)	$N = 2519$	$n = 1353$ (53.7%)	$N = 2038$	$n = 1951$ (95.7%)	$\chi^2_{(1)} = 996$	<.001	$\varphi = 0.47$
Age (years)	$N = 2519$	$M = 50.5$ ($SD = 18.6$)	$N = 2038$	$M = 24.0$ ($SD = 10.6$)	$W = 4{,}536{,}439$	<.001	$r_{rb} = 0.66$
Body mass index (kg/m^2)	$N = 2480$	$M = 25.3$ ($SD = 4.15$)	$N = 2011$	$M = 18.8$ ($SD = 6.92$)	$W = 4{,}322{,}145$	<.001	$r_{rb} = 0.63$
Eating Disorder Examination–Questionnaire (mean global score)	$N = 2519$	$M = 0.53$ ($SD = 0.78$)	$N = 2038$	$M = 3.83$ ($SD = 1.42$)	$W = 167{,}509$	<.001	$r_{rb} = 0.81$
Consumption of large amounts (number of times)	$N = 2515$	$M = 0.41$ ($SD = 2.06$)	$N = 1977$	$M = 7.60$ ($SD = 12.9$)	$W = 1{,}401{,}065$	<.001	$r_{rb} = 0.47$
Binge-eating episodes (number of times)	$N = 2515$	$M = 0.16$ ($SD = 1.24$)	$N = 1967$	$M = 6.15$ ($SD = 12.0$)	$W = 1{,}471{,}122$	<.001	$r_{rb} = 0.48$
Binge-eating days (number of days)	$N = 2515$	$M = 0.13$ ($SD = 1.17$)	$N = 1999$	$M = 5.70$ ($SD = 9.06$)	$W = 1{,}500{,}305$	<.001	$r_{rb} = 0.49$
Vomiting (number of times)	$N = 2517$	$M = 0.03$ ($SD = 0.67$)	$N = 1987$	$M = 6.41$ ($SD = 15.3$)	$W = 1{,}707{,}977$	<.001	$r_{rb} = 0.44$
Use of laxatives (number of times)	$N = 2515$	$M = 0.06$ ($SD = 0.71$)	$N = 1998$	$M = 1.13$ ($SD = 4.83$)	$W = 2{,}275{,}596$	<.001	$r_{rb} = 0.20$
Excessive exercise (number of times)	$N = 2513$	$M = 0.21$ ($SD = 1.59$)	$N = 1985$	$M = 8.65$ ($SD = 11.1$)	$W = 1{,}130{,}922$	<.001	$r_{rb} = 0.60$
Any comorbid mental disorder	–	–	$N = 2038$	$n = 1398$ (68.6%)	–	–	–
Comorbid affective disorders	–	–	$N = 2038$	$n = 1183$ (58.1%)	–	–	–
Comorbid anxiety disorders	–	–	$N = 2038$	$n = 473$ (23.2%)	–	–	–
Global Assessment of Functioning	–	–	$N = 1573$	$M = 44.4$ ($SD = 10.9$)	–	–	–
Illness duration (years)	–	–	$N = 1369$	$M = 6.69$ ($SD = 7.85$)	–	–	–

guidelines by the ethics committee of the LMU Munich, retrospective studies conducted on already available, anonymized data are exempt from ethics approval.

2.2 | Measures

The German version of the EDE-Q (Hilbert & Tuschen-Caffier, 2016) was used in the study by Hilbert et al. (2012) and is part of the routine diagnostic assessment at admission at the hospital. Internal reliability coefficients for the 22 items with which the mean global score was computed were $\omega = 0.96$ in the control group and $\omega = 0.95$ in the patient group. For the patient group, information on diagnoses, global functioning, and illness duration were taken from the clinical records. That is, diagnoses were not obtained by a structured clinical interview but where made by the patients' therapists in a non-standardized way. Therapists also rated the patients' global functioning at admission on the Global Assessment of Functioning scale (American Psychiatric Association, 1994), scores of which can range between 1 (*severely impaired*) and 100 (*extremely high functioning*), and provided an estimate of the patients' illness duration in years.

2.3 | Data analyses

Data were analyzed with R version 4.3.1 in RStudio version 2023.06.1 using the packages *foreign*, *summarytools*, *psych*, *rcompanion*, *stats*, *rstatix*, *rROC*, and *MatchIt*. Specifically, controls and patients were compared regarding sex with a χ^2-test and regarding age, body mass index, and EDE-Q scores with Wilcoxon rank sum tests. Eating disorder subgroups were compared regarding sex and comorbid mental disorders with χ^2-tests and regarding age, body mass index, EDE-Q scores, global functioning, and illness duration with Kruskal–Wallis rank sum tests. To determine optimal thresholds for EDE-Q global scores, ROC analyses were run for controls vs. patients (i.e., all eating disorder groups), controls versus patients with anorexia nervosa, controls vs. patients with bulimia nervosa, and controls versus patients with other eating disorders. Specifically, we determined sensitivity (true positive rate), specificity (true negative rate), positive predictive value (precision), and negative predictive value at different values of the EDE-Q global score for classifying persons as controls or patients. We also report area under curve (cf. Figure 1), which can range between 0.5 and 1 and higher values (i.e., closer to 1) represent better overall classification performance (here: performance of EDE-Q global scores for classifying persons as controls or patients). Choosing a threshold for the EDE-Q global score at which sensitivity is 100% would mean that all persons who have an eating disorder score above this threshold but specificity would be low (i.e., there would also be persons without an eating disorder who score higher than the threshold). In turn, choosing a threshold for the EDE-Q global score at which specificity is 100% would mean that everyone without an eating disorder scored below this threshold but sensitivity would be low (i.e., there would also be persons with an eating disorder who score lower than the threshold). Thus, a threshold should be chosen at which both sensitivity and specificity are relatively high, which can be determined with Youden's J statistic (sensitivity + specificity – 1; Youden, 1950). This statistic can range between 0 and 1: a value of 1 would indicate that there are no false positives or false negatives (i.e., there would be both a sensitivity and a specificity of 100%). Optimal thresholds for EDE-Q global scores were, therefore, determined by choosing the value with the highest J statistic (i.e., closest to 1).

TABLE 2 Sample characteristics as a function of eating disorder subgroups.

Variables	Anorexia nervosa			Bulimia nervosa			Other eating disorders		Test statistics	*p*	Effect sizes
Sex (female)	$N = 1456$	$n = 1401$ (96.2%)	=	$N = 370$	$n = 362$ (97.8%)	>	$N = 212$	$n = 188$ (88.7%)	$\chi^2_{(2)} = 30.7$	<.001	$\varphi_c = 0.12$
Age (years)	$N = 1456$	$M = 22.4$ ($SD = 9.75$)	<	$N = 370$	$M = 25.6$ ($SD = 9.93$)	<	$N = 212$	$M = 32.0$ ($SD = 13.3$)	$H_{(2)} = 176$	<.001	$\eta^2 = 0.09$
Body mass index (kg/m^2)	$N = 1438$	$M = 15.8$ ($SD = 2.50$)	<	$N = 363$	$M = 22.4$ ($SD = 4.50$)	<	$N = 210$	$M = 32.8$ ($SD = 10.1$)	$H_{(2)} = 1020$	<.001	$\eta^2 = 0.50$
Eating Disorder Examination–Questionnaire (mean global score)	$N = 1456$	$M = 3.77$ ($SD = 1.47$)	<	$N = 370$	$M = 4.10$ ($SD = 1.25$)	>	$N = 212$	$M = 3.72$ ($SD = 1.30$)	$H_{(2)} = 16.0$	<.001	$\eta^2 = 0.01$
Consumption of large amounts (number of times)	$N = 1407$	$M = 4.63$ ($SD = 10.3$)	<	$N = 366$	$M = 16.4$ ($SD = 16.9$)	>	$N = 204$	$M = 12.3$ ($SD = 12.1$)	$H_{(2)} = 461$	<.001	$\eta^2 = 0.23$
Binge-eating episodes (number of times)	$N = 1404$	$M = 3.17$ ($SD = 8.71$)	<	$N = 362$	$M = 15.3$ ($SD = 16.8$)	>	$N = 201$	$M = 10.6$ ($SD = 11.9$)	$H_{(2)} = 558$	<.001	$\eta^2 = 0.27$
Binge-eating days (number of days)	$N = 1424$	$M = 2.92$ ($SD = 6.89$)	<	$N = 367$	$M = 13.5$ ($SD = 9.98$)	>	$N = 208$	$M = 11.0$ ($SD = 10.0$)	$H_{(2)} = 585$	<.001	$\eta^2 = 0.29$
Vomiting (number of times)	$N = 1419$	$M = 4.37$ ($SD = 12.5$)	<	$N = 364$	$M = 17.3$ ($SD = 22.0$)	>	$N = 204$	$M = 1.24$ ($SD = 7.59$)	$H_{(2)} = 426$	<.001	$\eta^2 = 0.21$
Use of laxatives (number of times)	$N = 1427$	$M = 1.29$ ($SD = 5.27$)	=	$N = 364$	$M = 0.82$ ($SD = 3.48$)	>	$N = 207$	$M = 0.58$ ($SD = 3.46$)	$H_{(2)} = 7.88$	.020	$\eta^2 = 0.003$
Excessive exercise (number of times)	$N = 1413$	$M = 10.1$ ($SD = 11.7$)	>	$N = 364$	$M = 5.93$ ($SD = 8.74$)	>	$N = 208$	$M = 3.48$ ($SD = 7.12$)	$H_{(2)} = 94.3$	<.001	$\eta^2 = 0.05$
Any comorbid mental disorder	$N = 1456$	$n = 978$ (67.2%)	=	$N = 370$	$n = 259$ (70.0%)	=	$N = 212$	$n = 161$ (75.9%)	$\chi^2_{(2)} = 7.03$	.030	$\varphi_c = 0.06$
Comorbid affective disorders	$N = 1456$	$n = 805$ (55.3%)	<	$N = 370$	$n = 240$ (64.9%)	=	$N = 212$	$n = 138$ (65.1%)	$\chi^2_{(2)} = 15.9$	<.001	$\varphi_c = 0.09$
Comorbid anxiety disorders	$N = 1456$	$n = 364$ (25.0%)	>	$N = 370$	$n = 62$ (16.8%)	=	$N = 212$	$n = 47$ (22.2%)	$\chi^2_{(2)} = 11.4$	.003	$\varphi_c = 0.07$
Global Assessment of Functioning	$N = 1112$	$M = 42.8$ ($SD = 11.4$)	<	$N = 301$	$M = 48.1$ ($SD = 8.29$)	=	$N = 160$	$M = 49.2$ ($SD = 7.76$)	$H_{(2)} = 83.7$	<.001	$\eta^2 = 0.04$
Illness duration (years)	$N = 996$	$M = 5.92$ ($SD = 7.36$)	<	$N = 260$	$M = 8.37$ ($SD = 7.70$)	=	$N = 113$	$M = 9.69$ ($SD = 10.7$)	$H_{(2)} = 63.8$	<.001	$\eta^2 = 0.03$

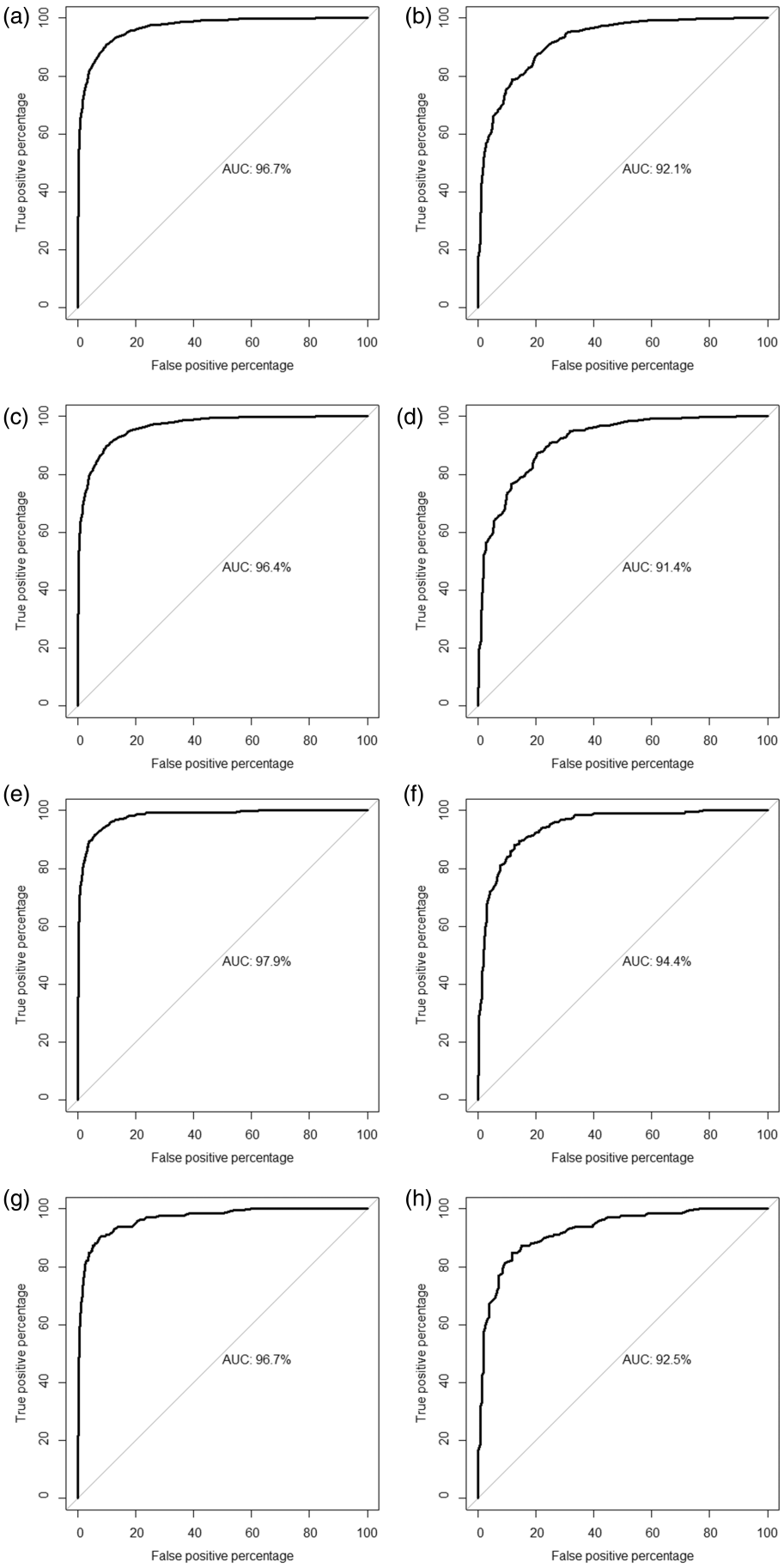

FIGURE 1 Receiver Operating Characteristic curves for (a) controls vs. patients (all eating disorders), (b) controls vs. patients (matched for age and sex), (c) controls vs. patients with anorexia nervosa, (d) controls vs. patients with anorexia nervosa (matched for age and sex), (e) controls vs. patients with bulimia nervosa, (f) controls vs. patients with bulimia nervosa (matched for age and sex), (g) controls vs. patients with other eating disorders, and (h) controls vs. patients with other eating disorders (matched for age and sex).

The ROC analyses were also re-run after matching each of the two respective groups for age and sex distribution using 1:1 nearest-neighbor propensity score matching without replacement. For this type of matching, a propensity score is calculated for each case and represents the probability of membership in a particular condition (here: controls vs. patients) given a set of measured variables (here: age and sex). Nearest-neighbor matching then assigns a control case that is nearest to each patient case in their respective propensity scores. Here, a maximum allowable distance in probabilities (or absolute difference in the logit of propensity scores) needs to be set (the so-called caliper width). A wider caliper affects the quality of the matching as controls that are further away from patients in their propensity scores increase the chance for bias to remain and a narrower caliper may reduce sample size if there are not enough similar cases that can be matched, thus leading to more variance. The choice of the caliper width is, therefore, a trade-off between bias (no well-matched groups) and variance (well-matched but smaller groups; Beal & Kupzyk, 2014). If the 1:1 nearest-neighbor matching did not lead to well-matched groups, we lowered caliper width until standardized mean differences in age and sex were below 0.1.

3 | RESULTS

3.1 | Optimal EDE-Q thresholds for discriminating controls versus patients

The patient group had a higher percentage of female persons, was younger, had a lower body mass index, and had higher EDE-Q scores than the control group (Table 1). The EDE-Q global score discriminated well between controls and patients (Figure 1a), and Youden's J statistic suggested either a score of 1.57 or 1.67 as optimal threshold (Table 3). Nearest-neighbor matching did not result in well-matched groups. With a caliper of 0.08, however, 657 controls could be matched to 657 patients with standardized mean differences in age and sex of <0.1. Although re-running the ROC analysis with these matched groups indicated slightly lower accuracy, the EDE-Q global score again discriminated well between controls and patients (Figure 1b) with the optimal threshold being a score of 1.57 (Table 3).

3.2 | Optimal EDE-Q thresholds for discriminating controls versus patients with anorexia nervosa

Like the analyses with the full sample, the EDE-Q global score discriminated well between controls and patients with anorexia nervosa (Figure 1c) with an optimal threshold of 1.57 (Table 3). Again, nearest-neighbor matching did not result in well-matched groups, but with a caliper of 0.08, 489 controls could be matched to 489 persons with anorexia nervosa with standardized mean differences in age and sex of <0.1. Although re-running the ROC analysis with these matched groups indicated slightly lower accuracy, the EDE-Q global score again discriminated well between controls and patients with anorexia nervosa (Figure 1d) with the optimal threshold being a score of 1.57 (Table 3).

3.3 | Optimal EDE-Q thresholds for discriminating controls versus patients with bulimia nervosa

The EDE-Q global score discriminated well between controls and patients with bulimia nervosa (Figure 1e) with an optimal threshold being 2.07 (Table 3). Nearest-neighbor matching did not result in well-matched groups but with a caliper of 0.25, 304 controls could be

TABLE 3 Results of the Receiver Operating Characteristic analysis.

Groups	Area under the curve	Threshold	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Youden's J
All eating disorders							
Controls vs. patients	96.7%	1.57/1.67	.907/.897	.901/.911	.881/.891	.923/.917	.808/.808
Controls vs. patients (matched for age and sex)	92.1%	1.57	.877	.795	.810	.866	.671
Anorexia nervosa							
Controls vs. anorexia nervosa	96.4%	1.57	.897	.901	.839	.938	.798
Controls vs. anorexia nervosa (matched for age and sex)	91.4%	1.57	.873	.793	.809	.862	.667
Bulimia nervosa							
Controls vs. bulimia nervosa	97.9%	2.07	.914	.942	.698	.987	.856
Controls vs. bulimia nervosa (matched for age and sex)	94.4%	2.34	.882	.875	.876	.881	.757
Other eating disorders							
Controls vs. other eating disorders	96.7%	1.80	.901	.922	.494	.991	.823
Controls vs. other eating disorders (matched for age and sex)	92.5%	2.39	.849	.882	.878	.854	.731

matched to 304 persons with bulimia nervosa with standardized mean differences in age and sex of <0.1. Re-running the ROC analysis with these matched groups resulted in slightly lower area under the curve (Figure 1f), sensitivity, specificity, and negative predicted value but higher positive predictive value, with an optimal threshold being a score of 2.34 (Table 3).

3.4 | Optimal EDE-Q thresholds for discriminating controls versus other eating disorders

The EDE-Q global score discriminated well between controls and patients with other eating disorders (Figure 1g) except that the positive predictive value was rather low (Table 3). The optimal threshold was a score of 1.80. Nearest-neighbor matching resulted in well-matched groups with standardized mean differences in age and sex of <0.1. That is, no specification of the caliper was necessary, and 212 controls could be matched to all 212 persons with other eating disorders. Re-running the ROC analysis with these matched groups resulted in slightly lower area under the curve (Figure 1h), sensitivity, specificity, and negative predicted value but higher positive predictive value, with an optimal threshold being a score of 2.39 (Table 3).

4 | DISCUSSION

The current study aimed at deriving cutoff scores for the EDE-Q based on a large community sample and a large sample of patients with eating disorders. To our knowledge, this is the first study that determined cutoff scores for the EDE-Q with almost equally sized groups (thus, reducing the risk of base rate fallacy) and the first study that determined cutoff scores for the German population. Results indicated that a global score of 1.6 (when rounded to one decimal) discriminated best between controls and persons with any eating disorder. Specifically, only 9% of patients ($n/N = 189/2038$) had a score below 1.57 and, thus, were misclassified as having no eating disorder. Similarly, only 10% of controls ($n/N = 250/2519$) had a score above 1.57 and were, thus, classified as having an eating disorder. Further analyses revealed, however, that a similar score optimally distinguished controls from persons with anorexia nervosa. The sample of patients with eating disorders was largely composed of persons with anorexia nervosa (>70%), which may have skewed the analyses of the overall sample to a lower cutoff score.

Indeed, optimal thresholds were slightly higher when examining patients with bulimia nervosa and other eating disorders. Interestingly, for discriminating between persons with bulimia nervosa and age- and sex-matched controls, a score of 2.3 was derived as an optimal threshold. This score equals the cutoff value in the study by Mond et al. (2004), in which a score of 2.3 had adequate sensitivity and specificity but a low positive predictive value of 0.30 for discriminating between 195 controls and 13 cases. Of note, the positive predictive value in Mond and colleagues' study (Mond et al., 2004) increased in conjunction with the occurrence of binge-eating episodes and excessive exercise, indicating that this score may be primarily appropriate for persons with bulimia nervosa.

A similar score of 2.4 was the optimal threshold for discriminating between persons with other eating disorders and age- and sex-matched controls. While binge-eating disorder is not included as an eating disorder in the ICD-10, it is usually coded as "overeating associated with other psychological disturbance," "other eating disorder," or "unspecified eating disorder," that is, the diagnoses that the group of other eating disorders was composed of in the current study. When examining Table 2, it is likely that most of the patients in this group had binge-eating disorder as this group reported regular binge-eating episodes but had a higher percentage of male persons, was older, had a higher body mass index, and reported less frequent purging behaviors than patients with anorexia nervosa or bulimia nervosa. That is, this group's characteristics matches well with the typical characteristics of persons with binge-eating disorder (Giel et al., 2022).

EDE-Q cutoff scores can be used for different purposes, such as estimating prevalence of eating disorders in epidemiological studies or evaluating treatment outcome. This study suggests that—at least in the German population—a score above 1.57 in patients with anorexia nervosa, a score above 2.34 in patients with bulimia nervosa, and a score above 2.39 in patients with binge-eating disorder might still represent elevated eating disorder psychopathology at the end of treatment, as these scores discriminated best between patients and age- and sex-matched controls. Researchers may also use these cutoff scores when deciding about inclusion or exclusion of study participants. For example, studies that aim to investigate a control group of persons without eating disorder symptomatology may only include persons that score below a certain cutoff score of the EDE-Q. As the lowest cutoff score in the current analyses was 1.6, this might be a threshold that researchers could use to ensure having a control group of persons without eating disorders. This score is somewhat lower than a score of 2.3 that researchers have used for this purpose in previous studies (e.g., Brockmeyer et al., 2013).

Regardless of the exact purpose that the cutoff scores might be used for, the current study shows that across different eating disorders and control groups, the optimal EDE-Q cutoff scores seem to lie somewhat around a global score of 2, which is in line with other studies (Melisse et al., 2022; Mond et al., 2004, 2008; Rica et al., 2022; Rø et al., 2015; Wade et al., 2021). Thus, using a score of 4, as has been done in some studies (e.g., Jennings & Phillips, 2017; Lavender et al., 2010; Luce et al., 2008), seems to be inappropriate.

4.1 | Limitations

Strengths of the current study include the large samples, which reduced the risk of base rate fallacy. However, findings of the current study are limited to Germany, that is, the cutoff scores found in this study may not translate to other regions of the world. Yet, the striking overlap with the findings by Mond et al. (2004), which was conducted in Australia, suggests that results might not only be restricted to a

German population but may apply to other Western or Westernized countries as well. Results are also somewhat in line with the study by Rø et al. (2015), which was conducted in Norway, in that optimal thresholds are higher for persons with bulimia nervosa and other eating disorders than for persons with anorexia nervosa. Moreover, although the optimal threshold for anorexia nervosa was slightly higher (score of 2.1) in the study by Rø et al. (2015) than in the current study, a score of 1.6 in that study discriminated best between non-underweight and underweight persons.

Another crucial aspect is that all persons with eating disorders in the current study were admitted to inpatient treatment and, thus, presented with a higher severity than other persons with eating disorders (e.g., those who receive outpatient treatment). This is illustrated, for example, in Tables 1 and 2, which demonstrate a high rate of comorbid mental disorders, low levels of global functioning, and long illness duration. Thus, it may be that examining optimal thresholds for the EDE-Q global score in other persons with eating disorders that do not receive inpatient treatment would yield different cutoff values.

The small subset of male patients with eating disorders precluded performing analyses separately for males and females. Thus, we could not determine whether different cutoff scores should be applied to men. Furthermore, despite matching groups for age and sex distribution, there might still have been differences between controls and patients such as socioeconomic status, possibly affecting results. Finally, we relied on clinical diagnoses (which may be less precise than diagnoses obtained with structured diagnostic interviews) and—as a nationally representative sample—the control group also included persons with eating disorders. Thus, it is possible that these diagnostic inaccuracies may also have somewhat biased results.

5 | CONCLUSION

In the German population, cutoff scores between 1.6 and 2.4 may be used to screen for the presence or absence of an eating disorder or evaluate treatment outcome, with slightly higher cutoff scores for persons with bulimia nervosa and other eating disorders than for persons with anorexia nervosa. The range of these cutoff scores is largely in line with those suggested in other countries, indicating that higher thresholds that have been applied in previous studies (e.g., a score of 4) should not be used.

AUTHOR CONTRIBUTIONS

Adrian Meule: Conceptualization; data curation; formal analysis; writing – original draft. **Anja Hilbert:** Investigation; project administration; resources; writing – review and editing. **Martina de Zwaan:** Investigation; project administration; resources; writing – review and editing. **Elmar Brähler:** Investigation; project administration; resources; writing – review and editing. **Stefan Koch:** Investigation; project administration; resources; writing – review and editing. **Ulrich Voderholzer:** Investigation; project administration; resources; writing – review and editing.

ACKNOWLEDGMENT

Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

No funding was received for this study.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Adrian Meule https://orcid.org/0000-0002-6639-8977
Anja Hilbert https://orcid.org/0000-0003-2775-1296
Martina de Zwaan https://orcid.org/0000-0002-7918-6957
Elmar Brähler https://orcid.org/0000-0002-2648-2728
Stefan Koch https://orcid.org/0009-0006-4731-4455
Ulrich Voderholzer https://orcid.org/0000-0003-0261-3145

REFERENCES

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). American Psychiatric Association.

Beal, S. J., & Kupzyk, K. A. (2014). An introduction to propensity scores: What, when, and how. *Journal of Early Adolescence*, *34*, 66–92. https://doi.org/10.1177/0272431613503215

Berg, K. C., Peterson, C. B., Frazier, P., & Crow, S. J. (2012). Psychometric evaluation of the eating disorder examination and eating disorder examination-questionnaire: A systematic review of the literature. *International Journal of Eating Disorders*, *45*, 428–438. https://doi.org/10.1002/eat.20931

Brockmeyer, T., Grosse Holtforth, M., Bents, H., Herzog, W., & Friederich, H.-C. (2013). Lower body weight is associated with less negative emotions in sad autobiographical memories of patients with anorexia nervosa. *Psychiatry Research*, *210*, 548–552. https://doi.org/10.1016/j.psychres.2013.06.024

Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *International Journal of Eating Disorders*, *16*, 363–370. https://doi.org/10.1002/1098-108X(199412)16:4<363::AID-EAT2260160405>3.0.CO;2-%23

Giel, K. E., Bulik, C. M., Fernandez-Aranda, F., Hay, P., Keski-Rahkonen, A., Schag, K., Schmidt, U., & Zipfel, S. (2022). Binge eating disorder. *Nature Reviews Disease Primers*, *8*(1), 16. https://doi.org/10.1038/s41572-022-00344-y

Hilbert, A., de Zwaan, M., & Braehler, E. (2012). How frequent are eating disturbances in the population? Norms of the Eating Disorder Examination-Questionnaire. *PLoS One*, *7*(1), e29125. https://doi.org/10.1371/journal.pone.0029125

Hilbert, A., & Tuschen-Caffier, B. (2016). *Eating Disorder Examination–Questionnaire*. dgvt Verlag. https://www.dgvt-verlag.de/e-books/2_Hilbert_Tuschen-Caffier_EDE-Q_2016.pdf

Jennings, K. M., & Phillips, K. E. (2017). Eating disorder examination-questionnaire (EDE-Q): Norms for clinical sample of female adolescents with anorexia nervosa. *Archives of Psychiatric Nursing*, *31*(6), 578–581. https://doi.org/10.1016/j.apnu.2017.08.002

Lavender, J. M., De Young, K. P., & Anderson, D. A. (2010). Eating Disorder Examination Questionnaire (EDE-Q): Norms for undergraduate men. *Eating Behaviors*, *11*, 119–121. https://doi.org/10.1016/j.eatbeh.2009.09.005

Luce, K. H., Crowther, J. H., & Pole, M. (2008). Eating Disorder Examination Questionnaire (EDE-Q): Norms for undergraduate women. *International Journal of Eating Disorders*, *41*, 273–276. https://doi.org/10.1002/eat.20504

Melisse, B., van Furth, E. F., & de Beurs, E. (2022). Eating disorder examination questionnaire (EDE-Q): Validity and norms for Saudi nationals. *Eating and Weight Disorders*, *27*, 139–150. https://doi.org/10.1007/s40519-021-01150-3

Meule, A. (Ed.). (2023). *Assessment of eating behavior*. Hogrefe. https://doi.org/10.1027/00616-000

Meule, A. (2021). Reconsidering the use of cut-off scores for the eating disorder examination–questionnaire. *Eating Disorders*, *29*, 480–484. https://doi.org/10.1080/10640266.2019.1678981

Mond, J. M., Hay, P. J., Rodgers, B., & Owen, C. (2006). Eating disorder examination questionnaire (EDE-Q): Norms for young adult women. *Behaviour Research and Therapy*, *44*, 53–62. https://doi.org/10.1016/j.brat.2004.12.003

Mond, J. M., Hay, P. J., Rodgers, B., Owen, C., & Beumont, P. J. V. (2004). Validity of the eating disorder examination questionnaire (EDE-Q) in screening for eating disorders in community samples. *Behaviour Research and Therapy*, *42*, 551–567. https://doi.org/10.1016/S0005-7967(03)00161-X

Mond, J. M., Myers, T. C., Crosby, R. D., Hay, P. J., Rodgers, B., Morgan, J. F., Hubert Lacey, J., & Mitchell, J. E. (2008). Screening for eating disorders in primary care: EDE-Q versus SCOFF. *Behaviour Research and Therapy*, *46*, 612–622. https://doi.org/10.1016/j.brat.2008.02.003

Rica, R., Solar, M., Compte, E. J., & Sepúlveda, A. R. (2022). Establishing the optimal male cut-off point: Confirmatory factor analysis of the eating disorder examination-questionnaire (EDE-Q) in a representative sample of Spanish university students. *Eating and Weight Disorders*, *27*, 969–977. https://doi.org/10.1007/s40519-021-01234-0

Rø, Ø., Reas, D. L., & Stedal, K. (2015). Eating disorder examination questionnaire (EDE-Q) in Norwegian adults: Discrimination between female controls and eating disorder patients. *European Eating Disorders Review*, *23*, 408–412. https://doi.org/10.1002/erv.2372

Rosenfeld, B., Sands, S. A., & Gorp, W. G. V. (2000). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology*, *15*, 349–359. https://doi.org/10.1093/arclin/15.4.349

Schaefer, L. M., Smith, K. E., Leonard, R., Wetterneck, C., Smith, B., Farrell, N., Riemann, B. C., Frederick, D. A., Schaumberg, K., Klump, K. L., Anderson, D. A., & Thompson, J. K. (2018). Identifying a male clinical cutoff on the eating disorder examination-questionnaire (EDE-Q). *International Journal of Eating Disorders*, *51*, 1357–1360. https://doi.org/10.1002/eat.22972

Wade, T., Pennesi, J.-L., & Zhou, Y. (2021). Ascertaining an efficient eligibility cut-off for extended Medicare items for eating disorders. *Australasian Psychiatry*, *29*, 519–522. https://doi.org/10.1177/10398562211028632

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*(1), 32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

How to cite this article: Meule, A., Hilbert, A., de Zwaan, M., Brähler, E., Koch, S., & Voderholzer, U. (2024). Cutoff scores of the Eating Disorder Examination–Questionnaire for the German population. *International Journal of Eating Disorders*, 1–9. https://doi.org/10.1002/eat.24133