# Comparison of Single- and Double-Stage Designs in the Prevalence Estimation of Eating Disorders in Community Samples

María Angeles Peláez-Fernández[1], Francisco Javier Labrador[2], and Rosa María Raich[3]
[1]University of Toronto at Mississauga (Canada)
[2]Universidad Complutense (Spain)
[3]Universidad Autónoma de Barcelona (Spain)

The aim of this research was to compare two different case-identification designs: (a) a one-stage anonymous design using the Eating Disorders Examination-Questionnaire (EDE-Q; Fairburn & Beglin, 1994) as diagnostic instrument and (b) a two-stage-non-anonymous design using the Eating Attitudes Test (EAT; Garner & Garfinkel, 1979) and the EDE-Q as screening instruments and the clinical interview Eating Disorders Examination (EDE; Fairburn & Cooper, 1993) as diagnostic instrument, in the estimation of eating disorders prevalence in community samples. Both epidemiological designs were compared in: eating disorders prevalence, population at risk, and weekly frequency of associated symptomatology (binge eating episodes, self-vomiting) within a sample of 559 scholars (14 to 18 year-old males and females) studying in the region of Madrid. Eating disorders prevalence estimation using single-stage design was 6.2%, and 3% using the two-stage design; however, these differences were not significant (p = .067). No significant differences between the two procedures were found either in population at risk or in weekly frequency of reported self-vomiting. Reported binge eating episodes were higher in the one-stage design. The use of a two-stage procedure with clinical interview (vs. questionnaire) leads to a better understanding of the items (specially the most ambiguous ones) and thus, to a more accurate prevalence estimation.
*Keywords: eating disorders, prevalence, case-identification design, one-stage design, two-stage design*

El objetivo era comparar la adecuación de dos protocolos en la estimación de prevalencia de trastornos de la conducta alimentaria (TCA): Protocolo de una fase, anónimo, usando como instrumento diagnóstico "Eating Disorders Examination-Questionnaire" (EDE-Q; Fairburn & Beglin, 1994); y Protocolo de doble fase, no anónimo, usando como instrumentos de "cribado" el "Eating Attitudes Test" (EAT; Garner & Garfinkel, 1979) y el EDE-Q; y como instrumento diagnóstico el "Eating Disorders Examination" (EDE; Fairburn & Cooper, 1993). Ambos protocolos fueron comparados en estimación de prevalencia de TCA, población en riesgo y frecuencia semanal de sintomatología asociada (episodios de sobreingesta y vómitos autoinducidos), en 559 adolescentes (ambos sexos) 14 - 18 años escolarizados en la Comunidad de Madrid. La estimación de prevalencia TCA con el protocolo de una fase fue 6,2%; y con el protocolo de dos fases, 3%, aunque las diferencias no fueron significativas (p = 0,067). No hubo diferencias significativas en cuanto a población en riesgo ni en frecuencia semanal de vómitos autoinducidos obtenida por ambos protocolos. La frecuencia semanal de atracones fue superior en el grupo de una fase. El protocolo de dos fases permite una mejor comprensión de los ítems y, por tanto, es más aconsejable para una estimación más precisa de la prevalencia de TCA.
*Palabras clave: trastornos de la conducta alimentaria, prevalencia, diseño de identificación de caso, diseño de una fase, diseño de doble fase*

Eating disorders (ED) have recently become an important concern in our society. Some cases reported in detail by the mass media, the age at which these symptoms usually appear, and the difficulty to understand such behaviors—which in some cases result in death—have caused significant social alarm. But how much truth is there in all this? The reality is that the numbers reported in epidemiological studies sometimes present important discrepancies, which does not help to clarify the situation. These differences are probably due to the method or the protocol followed to establish these diagnoses.

It is generally accepted the use of the diagnostic clinical interview performed by expert professionals is the method of choice to diagnose ED (Crowther & Sherwood, 1997; Fairburn & Beglin, 1994; Garner, 2002; Wilson & Smith, 1989). But, in view of the low estimated prevalence of these disorders (about 1% in the general population, American Psychiatric Association [APA], 2000), this procedure is very costly and, moreover, not a very good method to perform epidemiological studies, or simply to study an entire population in order to identify these rare cases of ED.

One of the responses has been to use ED identification protocols based on a questionnaire or test that can be easily applied to an entire population and thus detect the presence of these ED. But, among other problems, this strategy presents that of the quality or precision of a clinical diagnosis thus obtained.

As an alternative, two-stage ED identification protocols have been developed. In the first stage, a screening instrument (normally, a questionnaire) is administered, and the clinical diagnostic interview is only applied to the percentage of people identified as possible cases in this screening, along with a similar number of randomly selected possible no-cases. For some authors, this procedure is more adequate because it combines the possibility of administration to an entire population and precision in the clinical diagnosis (Dunn, Pickles, Tansella, & Vazquez-Barquero, 1999; Peláez Fernández, Labrador, & Raich, 2005, 2006, 2007). But these two-stage protocols also present some problems. The first is the difficulty of guaranteeing subjects' anonymity, because they are asked for some identification (name or code) to be able to locate them at the second stage, which may provoke refusal to participate in some participants or centers and/or a greater tendency to withhold data. The second is the possible loss of subjects from the first to the second stage. Other biases are due to the screening instrument used (Peláez Fernández et al., 2005). Although some of these problems have been reduced, others, such as the problem of the loss of anonymity with the two-stage protocols, have not yet been studied.

Regarding the instruments to detect ED, although the Eating Disorders Inventory (EDI; Garner, Olmsted, & Polivy, 1983) has been used frequently in epidemiological studies, it is not very economic as a screening instrument and it has not been validated for this purpose (Garner, 1991). The most frequently employed screening instrument, both in Spain (Peláez Fernández, Labrador & Raich, 2004) and in other

countries (Mintz & O'Halloran, 2000), is the Eating Attitudes Test (EAT; Garner & Garfinkel, 1979), although its validity as a case-detection instrument, especially of partial and incomplete cases, has not yet been demonstrated (Hay, Marley, & Lemar, 1998; Mond, Hay, Rodgers, Owen, & Beumont, 2004; Patton & Szmukler, 1995).

With regard to the diagnosis of ED, the Eating Disorder Examination (EDE; Fairburn & Cooper, 1993) interview is considered the instrument of choice (Garner, 2002) and should, therefore, be selected to diagnose ED in the second stage of the two-stage protocol. In the one-stage protocol, the most appropriate instrument to estimate the ED prevalence seems to be the Eating Disorder Examination-Questionnaire (EDE-Q; Fairburn & Beglin, 1994), a questionnaire derived from the EDE interview. Validity studies of the EDE-Q have shown a positive, but moderate, correlation between results obtained with the EDE and the EDE-Q in the assessment of psychopathological ED in the general population (Fairburn & Beglin, 1994), a clinical sample of women with a substance-abuse disorder (Black & Wilson, 1996), and a sample of patients diagnosed with bulimia nervosa (BN) and binge eating according to the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 1994) criteria (Wilfley, Schwartz, Spurrell, & Fairburn, 1997).

The principal aim of this investigation is to determine whether it is more appropriate to use a one-stage protocol with a diagnostic questionnaire or a two-stage protocol with an initial screening questionnaire and a subsequent diagnostic interview of possible detected cases to estimate the ED prevalence in community samples. A specific goal is to determine whether there are significant differences in the ED prevalence detected in both protocols (one- and two-stage). Complimentarily, we wish to determine whether there are differences in the results obtained (i.e., number of ED cases, monthly frequency of binge episodes and self-induced vomiting) by means of the EDE and the EDE-Q.

## Method

### Sample Design and Type of Sampling

A representative probabilistic sample of the school population in the Region of Madrid was selected. The sample was selected by means of random cluster sampling, first randomly selecting the School Centers and then the classrooms.

### Participants

The sample comprised 559 adolescent students, males and females between 14 and 18 years, registered during the 2000-2001 academic year in 3rd and 4th grade of Compulsory Secondary Education, or 1st and 2nd grade of high school,

in seven schools and institutes of the Autonomous Community of Madrid. Three of them were concerted and secular and four, private and religious.

The optimum sample size was calculated by means of the GPower program of Bruchner, Erdfelder, and Faul (1997), considering a 3% ED prevalence among the Spanish adolescent population (Pérez-Gaspar et al., 2000; Ruiz et al., 1998), a 91% sensitivity and 69.2% specificity of the EAT-40 (when used with a cut-off point of 20) for a 90% confidence level (CI; $\alpha = 0.1$) and a Type II error of .1 (power = .9). Under these assumptions, the sample size required was 466 students.

### Procedure and Measures

We established previous contact with the directors and guidance counselors of the educational centers to explain to them the purpose, goals, and requirements of the study. They all agreed to participate.

Once the personal acceptance of the center personnel had been obtained, we asked the students' parents for their written consent for students under 18 years of age.

*Administration of the one-stage and two-stage procedures.* To prevent reducing the collaboration of the participants with ED because of their tendency to deny or to conceal their symptomatology, we disguised the purpose of the investigation in the protocols. Participants were informed that the purpose of the study was to "know their opinion and habits in topics such as image or eating," indicating that we would administer a general questionnaire of eating and nutrition habits.

In each center, half of the classrooms were randomly assigned to each protocol. If the number of classrooms was odd, in the unpaired classroom, each one of the protocols was administered to one half of the participants. However, in one of the centers, the counselors decided to administer the one-stage protocol to a larger number of classrooms than the two-stage protocol, in order to reduce the number of subsequent interviews.

*One-stage protocol.* This was carried out in a single half-hour session; it was anonymous, and consisted of the administration of the following questionnaires: (a) a questionnaire of general sociodemographic data, in which were included items about age, sex, mean grade obtained in the past academic course, profession of both parents, etc.; (b) the Spanish version of the Eating Attitudes Test-40 (EAT-40; Garner & Garfinkel, 1979, translated and adapted by Castro, Toro, Salamero, & Guimerá, 1991), with a cut-off point of 20; (c) the Spanish translation of the EDE-Q (Eating Disorders Examination Questionnaire), carried out by the authors of the present investigation, which follows the diagnostic criteria of the DSM-IV-TR (APA, 2000).

All the participants were weighed and measured without heavy clothing but with their shoes on to calculate their Body Mass Index (BMI; kg/[height in meters]$^2$). Weight was corrected by subtracting 1.5 kg from the value obtained. Height was corrected by measuring the heel of the shoe and subtracting it from the value obtained.

*Two-stage protocol.* The first stage was identical to the one-stage protocol, except that it was not anonymous.

The selection criteria to go on to the second stage were: (a) scoring 20 or higher in the EAT-40; (b) having a BMI equal to or lower than 17.5. Participants who met at least one of these criteria ($n = 47$) were selected to be subsequently interviewed with the 12th version of the EDE (Fairburn & Cooper, 1993), in its Spanish adaptation (Raich, Mora, Sánchez Carracedo, & Torras, 2000), which follows the diagnostic criteria of the DSM-IV (APA, 2000). Likewise, an equivalent number of participants who scored lower than 20 in the EAT-40 ($n = 45$) were randomly selected to be interviewed, in order to estimate the rate of false negatives in this instrument. To estimate the prevalence in the two-stage group, we used the formula of Villaverde, Gracia, de la Fuente, González de Rivera, and Rodríguez-Pulido (1993):

$$\text{Prevalence} = (1 / N) \times \text{Summatory i } [(ei \times ni) / mi], \quad (1)$$

where $N$ = number of participants; $ei$ = number of detected cases; $ni$ = number of participants pre-selected as probable cases who were interviewed; and $mi$ = number of participants pre-selected as probable cases and randomly chosen to be interviewed.

Most of the interviews took place one week after administration of the first stage. As the complete administration of the EDE interview takes about 60 minutes per person, and in view of the center directors' refusal to allow an individual assessment that would take so long, we decided—like Hay (1998) and Colton, Woodside, and Kaplan (1999)—to only include the EDE items that refer to the diagnostic criteria of the DSM-IV-TR (APA, 2000). With this modification, the duration of the interviews was reduced to 15 minutes per person.

*Control of confounding variables.* To control the systematic influence of sociodemographic and personal variables such as age, family socioeconomic level, residence district, and type of center—private versus public—in each center, we applied the one-stage protocol to one half of the students, and the two-stage protocol to the other half. We assumed that, given the size of the sample, these variables would be distributed similarly in both groups. To verify whether the two groups (one-stage and two-stage) were equal with regard to the variables sex, mean grade obtained in the past academic course, and parents' profession, we performed chi square tests. No significant differences were found in any of them. In the case of BMI, we performed a Student's $t$ test, finding no significant differences between the two groups. The variable interviewer was controlled for constancy, as the same person always administered all the

diagnostic interviews. In order to control possible bias associated with the order in which the participants of the one-stage protocol completed the EAT-40 and EDE-Q questionnaires, we administered a notebook in which the EAT-40 appeared first and the EDE-Q afterwards to one half of the subjects of each classroom, and to the other half of the classroom, a notebook in which the assessment instruments appeared in reverse order.

## Results

The total number of students who participated in the study was 563. Of them, 289 (51.6%) participated in the one-stage protocol and 270 (48.2%) in the two-stage protocol. The results of 4 participants, 3 males and 1 female, were discarded from the one-stage group: one for responding "*always*" to all the EAT-40 questions, one for responding "*almost always*", and the third one (male) for responding to the items about menstruation; and the fourth did not respond correctly to the EAT-40 questions, marking several responses for each item. The only female whose questionnaire was discarded met the DSM-IV-TR (APA, 2000) criteria for BN in the EDE-Q. None of the subjects from the two-stage protocol were discarded. After subtracting the discarded questionnaires, a total of 559 students participated in the study. Table 1 displays the demographic characteristics of the sample.

Table 1
*Sample Sociodemographic Characteristics (n = 559)*

| Variables | | Frequency | | Percentage | |
|---|---|---|---|---|---|
| | | One-stage | Two-stage | One-stage | Two-stage |
| IAge (years) | 14 | 85 | 89 | 15.2 | 15.9 |
| | 15 | 98 | 82 | 17.5 | 14.7 |
| | 16 | 68 | 56 | 12.2 | 10.0 |
| | 17 | 32 | 34 | 5.7 | 6.1 |
| | 18 | 4 | 9 | 0.7 | 1.6 |
| | N/C | 2 | 0 | 0.4 | 0.0 |
| | Total | 289 | 270 | 51.7 | 48.3 |
| Sex | Male | 124 | 116 | 22.2 | 20.8 |
| | Female | 165 | 154 | 29.5 | 27.5 |
| | Total | 289 | 270 | 51.7 | 48.3 |
| BMI | *M* | 21.68 | 22.02 | | |
| | *SD* | 3.48 | 3.74 | | |
| Mean grade in previous course | Flunked | 46 | 48 | 8.2 | 8.6 |
| | Passed | 124 | 107 | 22.2 | 19.1 |
| | Notable | 78 | 93 | 14.0 | 16.6 |
| | Outstanding | 21 | 21 | 3.8 | 3.8 |
| | Didn't reply | 20 | 1 | 3.6 | 0.2 |
| | Total | 289 | 270 | 51.7 | 48.3 |
| Father's profession | Unemployed/ Retired | 6 | 7 | 1.1 | 1.3 |
| | Worker* | 139 | 107 | 24.9 | 19.1 |
| | Administrative-civil servant | 50 | 59 | 8.9 | 10.6 |
| | Liberal profession-Businessman | 72 | 74 | 12.9 | 13.2 |
| | Didn't reply | 22 | 23 | 3.9 | 4.1 |
| | Total | 289 | 270 | 51.7 | 48.3 |
| Mother's profession | Unemployed/ Retired | 6 | 0 | 1.1 | 0.0 |
| | Housewife | 114 | 107 | 20.4 | 19.1 |
| | Worker* | 63 | 60 | 11.3 | 10.7 |
| | Administrative-civil servant | 50 | 38 | 8.9 | 6.8 |
| | Liberal profession-Businesswoman | 38 | 46 | 6.8 | 8.2 |
| | Didn't reply | 18 | 19 | 3.2 | 3.4 |
| | Total | 289 | 270 | 51.7 | 48.3 |

* Specialized or nonspecialized worker

## ED Prevalence Estimation in both Groups

Prevalence estimation in the one-stage group was as follows: number of cases / N = 18 / 289 = 6.2%. In the two-stage group, and applying the formula of Villaverde et al. (1993), we obtained a prevalence estimation of N = 1 / 270) × [8 + (0 × 47 / 45)] = 3.0%. In effect, eight of the 47 students pre-selected as probable cases met the DSM-IV-TR (APA, 2000) criteria in the interview. None of the probable controls ($n$ = 45) randomly selected for the interview was diagnosed as an ED case. We used a chi square test to determine whether the differences between the prevalence estimations were significant, obtaining a value of $\chi^2$ = 3.356 ($p$ = .067), which indicates that, despite the fact that the ED prevalence estimation in the one-stage group was twice that of the two-stage group, the differences did not reach statistical significance. The results are presented in Table 2. Table 3 displays the distribution of subjects who met the diagnostic criteria of ED: anorexia nervosa (AN), BN, and eating disorders not otherwise specified (EDNOS) as a function of the diagnostic instrument (i.e., EDE-Q in the one-stage group and the EDE interview in the two-stage group).

To verify whether there were differences in the ED prevalence estimation as a function of the variable *sex*, we conducted a chi square test. The result obtained was: $\chi^2$ = 8.45 ($p$ < .05), which indicates that a higher percentage of

Table 2
*Frequencies and Percentages of Participants with Diagnostic Criteria for ED*

|  | One-stage group | Two-stage group |
|---|---|---|
| *Met ED criteria* | 18 (6.2%) | 8 (3%) |
| No ED criteria | 271 | 262 |
| Total | 289 | 270 |

Table 3
*Distribution of Participants with ED Diagnostic Criteria*

|  | One-stage group (EDE-Q) | Two-stage group (EDE) |
|---|---|---|
| Restrictive AN | 0 | 0 |
| Purgative AN | 0 | 0 |
| Purgative BN | 5 | 4 |
| Non-purgative BN | 1 | 0 |
| EDNOS Type 1 | 0 | 1 |
| EDNOS Type 2 | 1 | 0 |
| EDNOS Type 3 | 5 | 2 |
| EDNOS Type 4 | 5 | 0 |
| EDNOS Type 6 | 1 | 1 |
| TOTAL | 8 | 18 |

*Note.* AN = Anorexia nervosa, BN = Bulimia nervosa, EDNOS = Eating disorders not otherwise especified.

Table 4
*Distribution of Participants with ED Diagnostic Criteria as a Function of Sex*

|  | One-stage group (EDE-Q) | Two-stage group (EDE) |
|---|---|---|
| Males | 3 | 1 |
| Females | 15 | 7 |
| Total | 18 | 8 |

Table 5
*Means and Standard Deviations of Global EAT-40 Scores*

| Group | M | SD |
|---|---|---|
| One-stage | 13.36 | 10.34 |
| Two-stage | 12.91 | 8.96 |
| ED Case | 33.73 | 14.43 |
| No-ED Case | 12.14 | 8.18 |

women than men presented ED criteria. The frequencies and percentages are displayed in Table 4.

## At Risk Population in both Groups

No significant differences were found between the means of the global scores obtained by both groups in the EAT-40. A complementary analysis carried out to determine whether there were differences between the global EAT-40 scores between participants who were diagnosed as ED *cases,* by means of the EDE or the EDE-Q, and the *no-cases* revealed that the global EAT-40 score among the cases was significantly higher, $t$ = –12.559, $p$ < .001.

Likewise, the number of one-stage and two-stage participants who scored 20 or higher in the EAT-40 was compared. The differences between the two groups were not significant.

## Frequency of Associated Symptomatology in both Protocols

We compared the weekly frequency of binge eating and self-induced vomiting episodes reported in both protocols by means of a *t*-test. The weekly binge eating frequency in the one-stage protocol was higher ($t$ = 2.851, $p$ = .005) than the one obtained in the two-stage protocol (using the EDE interview). No differences in the weekly frequency of self-induced vomiting were reported in either protocol (see Table 6).

## Power of the EAT-40 Screening Questionnaire

In Table 7 are displayed the sensitivity, specificity, and positive predictive value of the EAT-40 for a cut-off point of 20 obtained in both groups.

Table 6

*Weekly Frequency of Binge and Vomiting Episodes According to the EDE-Q (One-Stage Protocol) and the EDE (Two-Stage Protocol)*

|  | Group | *N* | *M* | *SD* |
|---|---|---|---|---|
| Binge frequency | One-stage (EDE-Q) | 275 | 0.94 | 2.46 |
|  | Two-stage (EDE) | 94 | 0.20 | 0.87 |
| Vomiting frequency | One-stage (EDE-Q) | 275 | 0.15 | 0.81 |
|  | Two-stage (EDE) | 94 | 0.14 | 0.54 |

Table 7

*Sensitivity, Specificity, and Predictive Value of the EAT-40 (Cut-off Point = 20)*

|  | One-stage group | Two-stage group |
|---|---|---|
| Sensitivity | 83.33 | 87.50 |
| Specificity | 86.34 | 84.73 |
| Positive Predictive Power | 28.85 | 14.89 |

## Discussion

As we attempted to lend particular relevance to the fact of anonymity/no-anonymity in the ED identification protocol, it was necessary to ensure that the one-stage protocol was completely anonymous, compared with the two-stage protocol, which was not. The non-anonymous nature of the two-stage protocol probably increased concealing the symptoms among participants with ED. The diagnostic instrument used in the two-stage protocol, the EDE interview, allows defining more precisely the meaning of ED symptomatology than the instrument used in the one-stage protocol (the EDE-Q). Therefore, we expected the number of ED cases detected (prevalence estimation), the number of probable ED cases, and the monthly frequency of associated symptomatology identified by the two-stage protocol to be significantly lower than those of the one-stage protocol.

### Prevalence Estimations in the One-Stage versus the Two-Stage Procedure

The ED prevalence estimation found in the one-stage group was twice that found in the two-stage group; although these differences did not reach statistical significance, the values were quite near ($p = .06$). In fact, if we had counted as an ED case the adolescent girl from the one-stage group who was excluded from the study for incorrectly completing the EAT-40 but who, according to the EDE-Q, met the DSM-IV-TR (APA, 2000) criteria for BN, the prevalence estimation in the one-stage group would have increased to 6.7% (9.7% in females and 2.4% in males), and the difference in the prevalence estimation between both groups would have been significant.

These notable differences between the prevalence estimations of both groups could be due to the following:

1. The condition of *anonymity* in the one-stage group, versus the no-anonymity in the two-stage group, which could minimize the tendency to conceal or deny symptomatology among the participants with ED in the one-stage group;

2. The type of instrument used to assess the diagnostic criteria of ED (EDE vs. EDE-Q). It is difficult to correctly interpret of the questionnaire items, especially in the case of the concept of "binge eating," or the reliability of individuals' recall of the frequency of binge eating and inadequate weekly compensatory behaviors, which is easier to achieve in the interview. Participants may tend to overestimate these symptoms with the EDE-Q.

3. The sensitivity of the screening instrument used. We call attention to the fact that the sensitivity of the EAT-40 for a cut-off point of 20 is 87.5% (see Table 7); that is, the ED prevalence estimation has been underestimated by 13.5% in the two-stage group; in other words, we can assume that one or two participants who were not interviewed were, in fact, cases. To control this problem, we interviewed a random number of participants from the two-stage group who scored lower than 20 in the EAT-40. Subsequently, the responses to the EAT-40 of the remaining participants were reviewed, finding that one participant acknowledged vomiting "almost always after eating" and another admitted vomiting "sometimes." Upon interviewing both participants, the first met the DSM-IV-TR (APA, 2000) criteria for EDNOS Type 3. However, this participant was not computed as an ED case in the two-stage group, as selection was not random. This finding is isolated but it reveals that more cases may have "slipped through" among the 171 participants of the two-stage group who were not interviewed. If we had interviewed all of them, the prevalence estimation found would probably have been higher. Moreover, the EAT-40 establishes *criteria of a quantitative nature* and, to a lesser extent, qualitative criteria. That is, items such as "I wake up early in the morning" or "I cut my food into small pieces" are computed in the same way as the responses to "I have gone on eating binges where I feel that I may not be able to stop" or "I vomit after I have eaten." Therefore, a higher global EAT-40 score does not necessarily imply a higher probability of meeting the DSM-IV-TR (APA, 2000) criteria for ED. One of the implications of the quantitative scoring

of the EAT-40 is that some participants who scored below the cut-off point established for this instrument might be cases whereas a number of those who scored above the cut-off point might not. This is verified in the low positive predictive power of the EAT-40;

4. A combination of all three factors.

If the ED prevalence estimations found in this study are compared to those found in previous works, it is observed that, except for the study of Rojo et al. (2003), the prevalence estimation found in the two-stage protocol—3% (4.5% in females and 0.89% in males) is the one that best matches with those presented by the adolescents from Madrid (3.7%; Boletín Epidemiológico CAM, 2002), Barcelona (2.31%; Muro-Sans & Amador-Campos, 2007), La Mancha (4.05%; Rodríguez-Cano, Beato-Fernández, & Belmonte-Llario, 2005), Navarre (4.1%; Pérez-Gaspar et al., 2000), and Zaragoza (4.52%; Ruiz-Lázaro et al., 1998). However, in the last two cited works, a cut-off point of *30* for the EAT-40 was used and a random sample of participants who scored below 30 was *not* interviewed to estimate the rate of false negatives. Taking into account that, in our study, only 5 of the 12 participants diagnosed with ED in the two-stage group scored 30 or above in the EAT-40 and that all 5 participants were female, if we had used 30 as the cut-off point and, moreover, if we had not interviewed a random sample of the probable controls (for this cut-off point), the final prevalence estimation for this study would have been 1.85% (1.85% for females and 0% for males).

We also emphasize the fact that the investigations of Pérez-Gaspar et al. (2000) and Ruiz et al., (1998) were carried out only with women, which explains why they obtained a prevalence of 4.1% and 4.52%, respectively, using a cut-off point of 30 in the EAT-40 and only interviewing participants who scored 30 or over.

Regarding the specific type of ED and its frequency, in the one-stage group, 6 out of the 18 participants (2% of the total) met the diagnostic criteria for purgative and no-purgative BN, and 4% for subclinical forms or EDNOS. None of the participants met the criteria for AN. In the two-stage group, 4 out of the 8 participants (1.48%) met the diagnostic criteria for purgative BN, and 2.22% for subclinical forms (EDNOS, Type 1 to 6).

These percentages concur with those reflected in other studies: The BN prevalence estimation among the adolescents of Navarre (*n* = 2,862; Pérez-Gaspar et al., 2000) was 0.8%; subclinical forms had a prevalence of 3.1%, being the EDNOS Type 3 the most frequent (1.4% of the total sample), and the least frequent were EDNOS Type 1 (0.2% of the sample) and EDNOS Type 5 (i.e., "repeatedly chewing and spitting out, but not swallowing, large amounts of food"), which was 0%. The AN prevalence was 0.3%.

In the study of Ruiz et al. (1998), 0.55% of the entire sample (*n* = 4,048) was diagnosed with BN; 3.83% with EDNOS; and 0.14% with AN. The fact of not having found any participant in our sample (*n* = 559) who met the AN criteria may be due to the small size of the sample and the low prevalence of the disorder. Moreover, the works of Pérez-Gaspar et al. (2000) and Ruiz et al. (1998) were carried out only with women, for whom AN prevalence is higher than for men.

## At Risk Population in the One-Stage versus the Two-Stage Procedure

The EAT-40 score differences between the one-stage versus the two-stage group were nonsignificant: both when comparing the groups' global EAT-40 scores and when considering the percentage of participants in these groups who scored over the cut-off point established for this instrument.

As the only difference in the administration of the EAT-40 in the two groups was the anonymous nature (i.e., anonymous in the one-stage group but not in the two-stage group), the differences in the ED prevalence estimations of the two groups cannot be attributed to this anonymity condition. If the different prevalence estimations for the two groups were due to the variable anonymity, then the differences in the proportion of probable cases and the global EAT-40 scores of the two groups would also be maintained, and this did not occur. Therefore, anonymity does not seem to be a decisive variable when selecting the identification protocol of EA.

## Frequency of Associated Symptomatology in both Protocols

The weekly frequency of binge eating reported in the one-stage protocol (using the self-report EDE-Q) was higher than that reported in the two-stage protocol (with the EDE interview). This finding confirms the results of Wilfley et al. (1997), who found a higher frequency of binge episodes (although the differences were nonsignificant) when the participants were assessed with the EDE-Q than when they were interviewed with the EDE.

The higher frequency of the binge episodes reported through the EDE-Q versus the EDE, as postulated by Fairburn and Beglin (1994), Black and Wilson (1996), and Wilfley et al. (1997), could be attributed to the fact that it is not possible to ensure the correct interpretation of ambiguous concepts such as binge eating in the questionnaire, and that the participants may respond intuitively according to their own concept of the term. In people who are more sensitive to these problems, episodes of lower intensity may be considered more intense. To this interpretation bias is added the difficulty of precisely recalling the binge episodes, which, by definition, occurs with a feeling of loss of control. In contrast, in the interview, the interviewer defines the meaning of the term, thus significantly reducing the frequency of the reported behavior.

Although one could logically assume that such a lack of definition of ambiguous terms like binge eating in the self-report questionnaire could lead to either infra- or overestimation, in practice, it has been noted that people tend to consider the term *binge* loosely, attributing it to any episode in which they had a subjective feeling of overeating. Our experience indicates that, in most cases, precision is low, that is, there is a tendency to underrate the importance of some episodes. In the interview, the precise criteria that constitute an episode of binge eating according to the DSM-IV-TR (APA, 2000) are specified: (a) eating, in a discrete period of time (e.g., within any 2-hour period), an amount of food that is definitely larger than most people would eat during a similar period of time and under similar circumstances; and (b) a sense of lack of control over eating during the episode (e.g., a feeling that one cannot stop eating or control what or how much one is eating); and therefore, a high percentage of people who indicate in the self-report questionnaire that they have binged, later admit that they misinterpreted the term and deny the presence of this symptomatology.

The weekly frequency of self-induced vomiting episodes reported in both protocols was similar, which confirms the results of other investigations (Black & Wilson, 1996; Fairburn & Beglin, 1994; Wilfley et al., 1997), which also found no significant differences in self-induced vomiting reported via the EDE-Q and the EDE.

## Conclusions

In ED epidemiological studies in community samples, it is more appropriate to use a two-stage protocol with an oral interview (vs. a one-stage protocol with a written diagnostic interview), as the oral diagnostic interview (vs. a questionnaire) guarantees the person's correct interpretation of the items, especially in the case of the more ambiguous ones, and therefore, it leads to greater precision in prevalence estimations.

The use of a two-stage identification protocol of cases with an oral diagnostic interview is also presented as an option to one-stage ED identification protocols with an oral interview, in which the entire sample is interviewed, because of the lower associated cost and the important reduction in time.

Along with the advantages of using a two-stage procedure with an oral interview, the following costs associated with this kind of design, which were present in this study, should be taken into account: (a) it means more trips to the centers and students' and teachers' refusal to participate in the procedure, because it involves face-to-face diagnostic interviews and non-anonymity (although there were differences in the precision of the information), in contrast to the one-stage procedures in which all the tests were administered in a single session; (b) participants are exposed to the power of the screening instrument used in the first stage; and (c) in the first stage, when ED prevalence

studies are made with scholastic adolescent populations, we recommend using the EAT-40 with a cut-off point of 20 instead of 30 because of the need to reduce the higher number of false negatives in the first stage without excessively reducing the number of false positives.

To palliate this cost of interviewing a large number of subjects, we propose the option of including the EDE-Q in the first stage along with the EAT-40, given the acceptable correlation of the EDE-Q with the EDE interview.

## References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.), revision. DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (4th ed.), text revision*. DC: Author.

Black, C. M. D., & Wilson, G. T. (1996). Assessment of eating disorders: Interview versus questionnaire. *International Journal of Eating Disorders, 20*, 43-50.

Boletín Epidemiológico de la Comunidad de Madrid. (2000). Trastornos del comportamiento alimentario: prevalencia de casos en mujeres adolescentes de la Comunidad de Madrid. *Author, Vol. 8*, 12-32.

Bruchner, A., Erdfelder, E., & Faul, F. (1997). How to use G*Power. URL.http://www.psychologie.unitrier.de:8000/projects/gpower/how_to_use_gpower.html.

Castro, J., Toro, J., Salamero, M., & Guimerá, E. (1991). The Eating Attitudes Test: Validation of the Spanish versión. *Evaluación Psicológica, 7*, 175-189.

Colton, P., Woodside, D. B., & Kaplan, A. S. (1999). Laxative withdrawal in eating disorders: Treatment protocol and 3 to 20 month follow-up. *International Journal of Eating disorders, 25*, 311-317.

Crowther, J. H., & Sherwood, N. E. (1997). Assessment. In D.M. Garner & P.E. Garfinkel (Eds.), *Handbook of treatment for eating disorders* (2nd ed., pp. 34-49). New York: Guilford Press.

Dunn, G., Pickles, A., Tansella, M., & Vazquez-Barquero, J. L. (1999). Two-phase epidemiological surveys in psychiatric research. *British Journal of Psychiatry, 174*, 95-100.

Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *International Journal of Eating Disorders, 16*, 363-370.

Fairburn, C. G., & Cooper, Z. (1993). The Eating Disorder Examination (12th ed.). In C. G. Fairburn & G. T. Wilson (Eds.), *Binge eating: Nature, assessment and treatment* (pp. 317-360). New York: Guilford Press.

Garner, D. M. (1991). *Eating Disorder Inventory-2: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Garner, D. (2002). Measurement of eating disorder psychopathology. In C. G. Fairburn & K. D. Brownell (Eds.), Eating disorders and obesity: A comprehensive handbook (2nd ed.) (pp. 141-146). New York: Guilford Press.

Garner, D. M., & Garfinkel, P. E. (1979). The Eating Attitudes Test: An index of the symptoms of anorexia nervosa. *Psychological Medicine, 9*, 273-279.

Garner, D. M., Olmsted, M. A., & Polivy, J. (1983). Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia. *International Journal of Eating Disorders*, *2*, 15-34.

Hay, P. (1998). The epidemiology of eating disorder behaviors: An Australian community-based survey. *International Journal of Eating Disorders, 23*, 371-382.

Hay, P. J., Marley, J., & Lemar, S. (1998). Covert eating disorders: The prevalence, characteristics and help-seeking of those with bulimic eating disorders in general practice. *Primary Care Psychiatry*, *4*, 95-99.

Mintz, L. B., & O'Halloran, M. S. (2000). The Eating Attitudes Test: Validation with DSM-IV eating disorder criteria. *Journal of Personality Assessment*, *74*, 489-503.

Mond, J. M., Hay, P. J., Rodgers, B., Owen, C., & Beumont, P. J. V. (2004). Validity of the Eating Disorders Examination Questionnaire (EDE-Q) in screening for eating disorders in community samples. *Behaviour Research and Therapy*, *42*, 551-567.

Muro-Sans, P., & Amador-Campos, J.A. (2007). Prevalence of eating disorders in a Spanish community adolescent sample. *Eating and Weight Disorders*, *12*, 1-6.

Patton, G. C., & Szmukler, G. I. (1995). Epidemiology of eating disorders. In A. Jablensky (Ed.), *Epidemiological psychiatry* (pp. 307-328). London: Balliere Tindall.

Peláez Fernández, M. A., Labrador, F. J., & Raich, R. M. (2004). Epidemiología de los trastornos de la conducta alimentaria en España. Revisión y estado de la cuestión. *Cuadernos de Medicina Psicosomática y Psiquiatría de Enlace, 71-72*, 33-43.

Peláez Fernández, M. A., Labrador, F. J., & Raich, R. M. (2005). Prevalence of eating disorders: Methodological considerations. *International Journal of Psychology and Psychological Therapy*, 5, 135-148.

Peláez Fernández, M. A., Labrador, F. J., & Raich, R. M. (2006). Epidemiología de los trastornos de la conducta alimentaria en España. In J. M. Mancilla Díaz & G. Gómez-Peresmitré (Eds.), *Trastornos alimentarios en Hispanoamérica* (pp. 89-122). Mexico: El Manual Moderno.

Peláez Fernández, M. A., Labrador, F. J., & Raich, R. M. (2007). Prevalence of eating disorders among adolescent and young

adult scholastic population in the region of Madrid (Spain). *Journal of Psychosomatic Research*, *62*, 681-690.

Pérez-Gaspar, M., Gual, P., De Irala-Estévez, J., Martínez González, M. A., Lahortiga, F., & Cervera, S. (2000). Prevalencia de trastornos de la conducta alimentaria (TCA) en adolescentes navarras. *Medicina Clínica, 114*, 481-486.

Raich, R. M., Mora, M., Sánchez Carracedo, D., & Torras, J. (2000). Adaptación y calidad psicométrica de la entrevista EDE (Eating Disorder Examination) para la evaluación de los trastornos alimentarios en población universitaria. *Revista AETCA, 1*, 3-5.

Rodríguez-Cano, T., Beato-Fernández, L., & Belmonte-Llario, A. (2005). New contributions to the prevalence of eating disorders in Spanish adolescents: Detection of false negatives. *European Psychiatry*, *20*, 173-8.

Rojo, L., Livianos, L., Conesa, L., García, A., Domínguez, A., Rodrigo, G., Sanjuán, L., & Vila, M. (2003). Epidemiology and risk factors of eating disorders: A two-stage epidemiologic study in a Spanish population aged 12-18 years. *Internacional Journal of Eating Disorders*, *34*, 281-9.

Ruiz-Lázaro, P. M., Alonso, J. P., Velilla, J. M., Lobo, A., Martín, A., Paumard, C., & Calvo, A. I. (1998). Estudio de prevalencia de trastornos de la conducta alimentaria en adolescentes de Zaragoza. *Revista de Psiquiatría Infanto-Juvenil, 3*, 148-162.

Villaverde, M. L., Gracia, R., de la Fuente, J., González de Rivera, J. L., & Rodríguez-Pulido, F. (1993). Estudio comunitario de salud mental en población urbana de Tenerife. In J. L. González de Rivera, F. Rodríguez Pulido & A. Sierra López (Eds.), *El método epidemiológico en salud mental* (pp. 187-200). Barcelona: Masson-Salvat.

Wilfley, D. E., Schwartz, M. B., Spurrell, E. B., & Fairburn, C. G. (1997). Assessing the specific psychopathology of binge eating disorder patients: Interview or self-report? *Behaviour Research and Therapy, 3*, 1151–1159.

Wilson, G.T., & Smith, D. (1989). Assessment of bulimia nervosa: An evaluation of the Eating Disorder Examination. *International Journal of Eating Disorders, 8*, 173-179.