

A Meta-Analysis of the Diagnostic Accuracy of the SCOFF

Juan Botella, Ana Rosa Sepúlveda, Huiling Huang and Hilda Gambara

Universidad Autónoma de Madrid (Spain)

Abstract. A meta-analysis was conducted to reach a pooled estimate of the diagnostic accuracy of the SCOFF. The 15 selected studies represented a total of 882 cases and 4350 controls. The main criterion for inclusion was that the primary study had provided diagnostic classification with both a diagnostic reference and with the SCOFF (with five items and a cut-off point of two). The pooled estimates were .80 (sensitivity) and .93 (specificity). The moderator variables gender and type of measure for the diagnostic reference (interview versus psychometric tests) account for part of the observed variability. For diagnostic references based on interviews the estimate of the efficacy improves significantly. For the studies that match this criterion the sensitivity is .882 and the specificity .925 (diagnostic odds ratio, 92.19). The main conclusion was that the five questions of the SCOFF constitute a very useful screening tool, in several languages; it is highly recommended for screening purposes.

Received 3 February 2012; Revised 7 June 2012; Accepted 13 September 2012

Keywords: SCOFF test, eating disorders, sensitivity and specificity, ROC curve.

Epidemiological studies have provided growing evidence that eating disorders (ED) are relevant illnesses in young females and adolescents (Becker, Thomas, Franko, & Herzog, 2005). The prevalence rate in western countries is around 4.5% (Cotrufo, Barretta, & Monteleone, 1998; Favaro, Ferrara, & Santonastasio, 2003; Morandé, Celada, & Casas, 1999; Pérez-Gaspar et al., 2000; Rojo et al., 2003; Ruíz-Lazaro et al., 1998; Steinhausen, Winkler, & Meier, 1997). The rate may be even greater, however, if all subclinical cases, the most difficult to detect, are included (increasing the rate to a possible 8%).

Although anorexia nervosa (AN) and bulimia nervosa (BN) are well-known disorders, in community studies they are not so prevalent. Most cases belong to the category of Eating Disorders Not Otherwise Specified (EDNOS) (Eddy et al., 2010) of the DSM-IV-R (APA, 2002). Although EDNOS can be even more severe and enduring, often they are not diagnosed until the disorder is well established, delaying the diagnosis and appropriate treatment. Above all, because the symptoms are presented first to non-specialists such as family doctors, they are not immediately recognized as related to an ED (Ogg, Millar, Puzstai, & Thom, 1997). Thus, it is crucial to identify the subpopulation at risk in order to encourage a higher awareness of the

availability of treatments (Becker et al., 2005; Herpertz-Dahlmann, Wille, Höling, Vloet, & Ravens-Sieberer, 2008). There is a great interest in the public health sector in improving effective screening and the strategies of secondary prevention for ED in adolescents. High-quality tools are required for effective screening in epidemiological studies or in primary care, given that because ED has a relatively low prevalence in the general population, the community studies that assess the diagnostic efficacy of screening tests need large samples to ensure adequate power and accuracy.

One of the tests most often recommended is the SCOFF questionnaire, which was developed as a quick and reliable tool for the screening of ED. Morgan, Reid, and Lacey (1999) designed the SCOFF as a test that could be administered by non-specialists able to detect symptoms of ED. According to the authors the proposed cut-off allows for a clinically appropriate balance between false positives and false negatives. This cut-off is a cue for suspicion that an ED exists and it should be followed by additional questioning about the patient's weight and their eating attitudes and behaviors (Hill, Reid, Morgan, & Lacey, 2010).

Several studies have examined the psychometric characteristics of the SCOFF, both in the original version and in its translation into several other languages (see Table 1). In those studies acceptable trade-offs between sensitivity and specificity have been found, along with high levels of reliability. Today there are enough assessment studies available to conduct a quantitative synthesis. A meta-analysis provides for integration of the results of several studies that assess the diagnostic efficacy of a tool (Botella & Huang, 2012;

Correspondence concerning this article should be addressed to Juan Botella. Facultad de Psicología. Departamento de Psicología Social y Metodología. Universidad Autónoma de Madrid. Cantoblanco s/n. 28049. Madrid (Spain).

E-mail: juan.botella@uam.es

This research was supported by the project PSI2009-12071, from the Spanish Ministerio de Educación y Ciencia.

Table 1. Identification and main characteristics of the 15 studies included in the meta-analysis

Study	Country	Age	Gender	Diagnostic Reference	Sample*	TP	FP	FN	TN
Berger et al, 2011	Germany ^a	12 (12–12)	52.7	EAT-26 \geq 20 ^I		64	191	17	535
Caamaño et al, 2002	Spain ^b	12** (11–13)	?	EAT-40 \geq 30 ^I		25	32	14	218
Cotton, Ball & Robison, 2003	UK ^c	29 (18–64)	77	Q-EDD ^I		21	23	6	173
Garcia et al., 2010	France ^d	20.9 (18–35)	100	MINI & interview DSM-IV ^{II}	8/29/0	35	19	2	363
Garcia et al, 2011	France ^d	22.1 (18–35)	100	MINI & interview DSM-IV ^{II}	67/45/0	106	6	6	108
Garcia-Campayo et al, 2005	Spain ^b	29.2 (15–53)	100	interview SCAN ^{II}	29/47/55	128	4	3	68
Lähteenmäki et al, 2009	Finland ^e	? (18–29)	57.7	interview SCID ^{II}		7	66	2	466
Luck et al, 2002	UK ^c	? (18–50)	100	interview DSM-IV ^{II}	1/3/9	11	34	2	294
Mond et al, 2008	USA ^c	27.6 (18–40)	100	EDE telephone & interview; EDEQ ^{II}	0/5/20	18	33	7	89
Morgan, Reid, & Lacey, 1999	UK ^c	? (18–40)	100	interview DSM-IV ^{II}	68/48/0	116	12	0	84
Muro-Sans, Amador-Campos, & Morgan, 2008	Spain ^f	13.5 (11–17)	49.8	EDI-2 (3 subscales) ^I		57	195	21	681
Pannocchia et al., 2011	Italy ^g	?	100	EDI-3 ^I		29	8	1	55
Parker, Lyons, & Bonner, 2005	USA ^c	? (20–51)	72.4	EDEQ & DSM-IV criteria ^I		32	16	28	220
Rueda, Diaz, Campo et al, 2005	Colombia ^b	21.2 (17–35)	100	interview CIDI ^{II}	0/8/43	40	46	11	144
Rueda, Díaz, Ortiz et al, 2005	Colombia ^b	14 (10–19)	100	interview CIDI ^{II}	4/14/54	59	36	13	133

^a German; ^b Spanish; ^c English; ^d French; ^e Finish; ^f Catalan; ^g Italian.

^I Psychometric test; ^{II} interview.

* The three values are the frequencies of cases diagnosed with the diagnostic reference being AN/BN/EDNOS.

** The mean has been estimated as the average between the higher and lower limits, given that the interval is very narrow.

Devillé et al., 2002; Gatsonis & Paliwal, 2006; Zweig & Campbell, 1993), which can also be done for other psychometric characteristics (Botella, Suero, & Gambará, 2010). The results are expressed mainly as the values of sensitivity and specificity of the pooled estimates that combine the estimates provided by primary studies. Our own meta-analysis also attempted to partially account for the variability observed, analyzing the role of several moderator variables.

In short, our aim was to perform a systematic review and meta-analysis to summarize the literature on assessments of the diagnostic efficacy of the SCOFF for detecting ED.

Method

Procedure

The SCOFF

The acronym refers to the SCOFF's questions, which are related to five key features in ED (Sick, Control, One, Fat and Food; see appendix A). Individuals must answer Yes/No to each question, scoring one point for each positive answer (Morgan et al., 1999). The authors proposed a cut-off of two or more points for recommending the clinician to follow up with a deeper and more rigorous assessment.

Literature search

We performed a systematic search of the literature using five international databases: *Medline*, *PsycINFO*, *EMBASE*, *Web of Science (Science Citation Index Expanded and Social Sciences Citation Index)* and *Cochrane Library*. Two of the authors (ARS and HG) searched for all the papers written in English, German or Spanish, published in peer-reviewed journals between 1999 and August 2011. The list of key words for the search included: *SCOFF questionnaire* combined with each of the terms 'eating disorder', 'anorexia nervosa', 'bulimia nervosa', 'screening', 'primary care', 'validation', 'psychometrics' and 'prevalence'. Each of those combinations was also combined with the term 'screening'. More than 50 combinations of words were searched in each of those data bases. We also did manual searches of the references cited in the selected papers.

Once we had read the abstracts we recovered copies of the relevant papers. Most of the retrieved studies reported the use of the SCOFF in a sample of participants in intervention studies or prevention programmes, but they did not provide relevant information related to diagnostic efficacy.

Inclusion and exclusion criteria

After the process described in the previous section we still had 30 papers for possible inclusion in the

meta-analysis. Studies were finally selected if they reported the use of the SCOFF in a sample of ED and other comparison sample, besides any measurement with a diagnostic reference. We excluded studies as follows: (a) those written in languages other than English, German or Spanish; (b) those not published in peer-review journals; (c) those which reported results for the SCOFF as a recommended tool for screening, including psychometric studies; (d) those which compared two forms of administration (oral versus written); (e) those in which although sensitivity and specificity were reported, the size of both samples was not reported, so that proper statistical treatment was not allowed; (f) those which related to a sample voluntarily participating in intervention programmes, as the data were likely to be biased.

No restrictions were placed on the gender and age of the participants, the type of sample, or the type of reference employed as a diagnostic reference. A total of 15 studies met all the inclusion criteria and were finally selected. They were performed in eight different countries and in seven different languages. Most of them are assessments of new translations to a different language. This gives us an opportunity to assess generalizability. The advantage of a tool with high generalizability is that it allows for comparisons and synthesis of the results from studies performed in different languages and countries.

The decision criteria for exclusion were independently applied by two of the authors (JB and ARS), with a high inter-rater agreement (coefficients equal to 1 for most variables, and all above .85).

Diagnostic references for the Gold Standard

Studies that assess the SCOFF's efficacy for screening have employed a large variety of tools as the diagnostic reference (gold standard). In 60% of cases individual diagnostic interviews are used which employ several tools (CIDI, DSM-IV, MINI, SCAN). Although they are not 100% reliable, their performance is good enough for them to be accepted as a gold standard.

Some studies, however, have chosen as diagnostic references specific cut-offs in psychometric tools, such as EAT-40, EAT-26 and EDI-2. Given that these procedures are less appropriate for the role of a gold standard, lower levels of diagnostic efficacy are to be expected compared with studies based on interviews.

Extraction of basic data

The data from each paper were systematically and independently extracted by two of the authors using a structured database; they reached perfect agreement for the four raw frequencies in the studies. The data related to the tool were transformed, when necessary,

to reach the four frequencies: true positives (TP), false positives (FP), false negatives (FN) and true negatives (TN). When the frequencies were reported separately according to the disorder (AN, BN, EDNOS) they were summed to reach the totals. In most cases the papers provided sensitivity and specificity, plus the size of each group, so that the frequencies were obtained from those quantities.

The following fields were coded: year of publication, country, language of the version employed, size of the samples, mean age, gender (percentage of women) and the tool employed for the diagnostic reference. We abandoned some initial fields because the number of studies providing the relevant information was very low (e.g. duration of the illness, mean body mass index).

Table 1 identifies each of the studies, with the values of the most relevant variables analyzed in the present research. The sources are marked with an asterisk in the references.

Statistical analyses

We have adjusted hierarchical models of the summary ROC curve (Gatsonis & Paliwal, 2006; Macaskill, 2004) employing the NLMIXED procedure of SAS (2008). These models provide combined estimates of the parameters that reflect diagnostic efficacy, the criterion for classification and the scale parameter. The parameters allow combined estimates of the performance indexes of the test to be derived, within the framework of Random Effects models. Random Effects models assume that the individual data sets are drawn from a distribution of populations. This is more credible than the assumption that all them come from a single population, as in Fixed Effect models (Borenstein, Hedges, Higgins, & Rothstein, 2010).

Besides the basic model we have explored several models that include as covariates the moderators for which we had enough information (mean age, gender, type of reference). These are Mixed models, as they involve a main Random Effect component, whereas the moderators are included as a Fixed Effect.

We have also obtained descriptive summaries and several types of figures from Review Manager (2008) and METADISC (Zamora, Abraira, Muriel, Kahn, & Coomarasamy, 2006).

Results

Characteristics of the studies

The set of 15 studies aggregated 882 cases with an ED and 4350 healthy controls, according to the diagnostic references employed. The main characteristics of the selected studies are described in Table 2.

Women predominate in the composition of the samples (between 50% and 100%). The average age is reported in only two-thirds of the studies (although all report the range).

Very varied procedures were employed for classification as the diagnostic reference. In 60% they are interviews, often structured by several well-known tools (CIDI, DSM-IV, MINI, SCAN). In the other 40% the classification was done by means of a cut-off in another test (EAT, EDI, EDEQ).

Diagnostic efficacy of the SCOFF

First of all we checked for any threshold effect. This effect is produced by shifts of the criteria for classification between different studies. It is reflected in a negative correlation between sensitivity and specificity. Although this effect is typical in tests with an implicit criterion for classification, it is also achievable with an explicit criterion that reflects a latent variable (such as the SCOFF). The Spearman's correlation between the sensitivities and the specificities in our 15 studies was not statistically significant ($r_{se} = .377$; $p = .166$), so we discarded this effect in our set of studies.

We fitted hierarchical models of the summary ROC curve (Gatsonis & Paliwal, 2006; Macaskill, 2004) by means of the NLMIXED procedure (SAS, 2008). Assuming logistic distributions, this model estimates several parameters, although there are really just two that characterize the performance of a binary classification tool. The first one, *alpha*, reflects the ability of the test for discrimination. The second one, *theta*, reflects how conservative or liberal the threshold for classification is. The model estimates the parameters jointly, so that it takes into account the joint variation inherent in any potential threshold effect (even if it is small and non-significant). Then the estimates of the parameters are converted into pooled estimates of the sensitivity and specificity.

Table 2. Primary studies' characteristics

Size of the samples; mean (range)	Targets, 58.8 (9 – 131) Controls, 290 (63 – 876)
Gender ($k = 14$); mean of the % of women (range)	86.4 (49.8 – 100)
Age ($k = 10$); mean (range)	20.1 (12 – 29.2)
Type of diagnostic reference	Based on interview 9 (60%) Test or questionnaire 6 (40%)

The basic hierarchical model (without covariates) fitted with the 15 studies provides combined estimates of the parameters that involve a sensitivity of .801 and a specificity of .934. The diagnostic odds ratio for these estimates is 56.96. The scale parameter, *beta*, is statistically different from zero ($p = .046$). This means that the summary curve is asymmetric (alpha and theta are not independent). The values for sensitivities show higher heterogeneity than those for specificities. This is a consequence of the fact that the size of the samples of targets is generally lower than the size of the samples of controls.

The homogeneity test for the values of sensitivity and specificity in the 15 studies shows that in both indexes the observed variability is larger than expected [sensitivity, $Q(14) = 139.98, p < .001$; specificity, $Q(14) = 202.55, p < .001$]. Figure 1 shows the combined forest plot of both indexes.

We fitted HSROC models that incorporate the moderator variables *mean age*, *gender* and *type of diagnostic reference* as covariates associated with the parameters alpha and/or theta. Whereas the models with the mean age did not show a significant association with alpha or theta, those including the gender and the type of diagnostic reference did show marginally significant associations with the parameter alpha, but not with theta. As regards gender, we did not have any a priori hypothesis for the direction of any eventual association. The observed association is positive. That is, the ability of the SCOFF to discriminate between cases and

normals is greater the larger the percentage of women in the sample. The efficacy of the test is marginally higher for women than for men ($p = .052$).

As regards the type of tool employed as the diagnostic reference, we expected a greater ability for discrimination when tools based on individual interviews were employed as compared with psychometric tools. Actually, the psychometric tools are not a real gold standard, given that their results are not beyond question. They have an imperfect reliability that underestimates the efficacy for classification of the tool for screening, as they classify some individuals (cases and normals) in the wrong group. This covariate has been included in the model as a *dummy* variable. As expected, it is associated with the test's capability for discrimination. The diagnostic efficacy is marginally greater when the diagnostic reference is based on an interview than when it is based on a psychometric test ($p = .057$).

Given that the diagnostic efficacy is significantly different according to the type of diagnostic reference and that the measurements based on interviews are clearly superior, the best estimates we can afford of the performance of the test are the estimates based on this last type of measurement. Fitting a new HSROC model only to the nine studies that employed diagnostic references based on interviews gives pooled estimates of .882 for sensitivity and .925 for specificity. The corresponding odds ratio for those estimates is 92.19.

Some of the primary studies have highlighted that perhaps the sensitivity of the SCOFF could be different

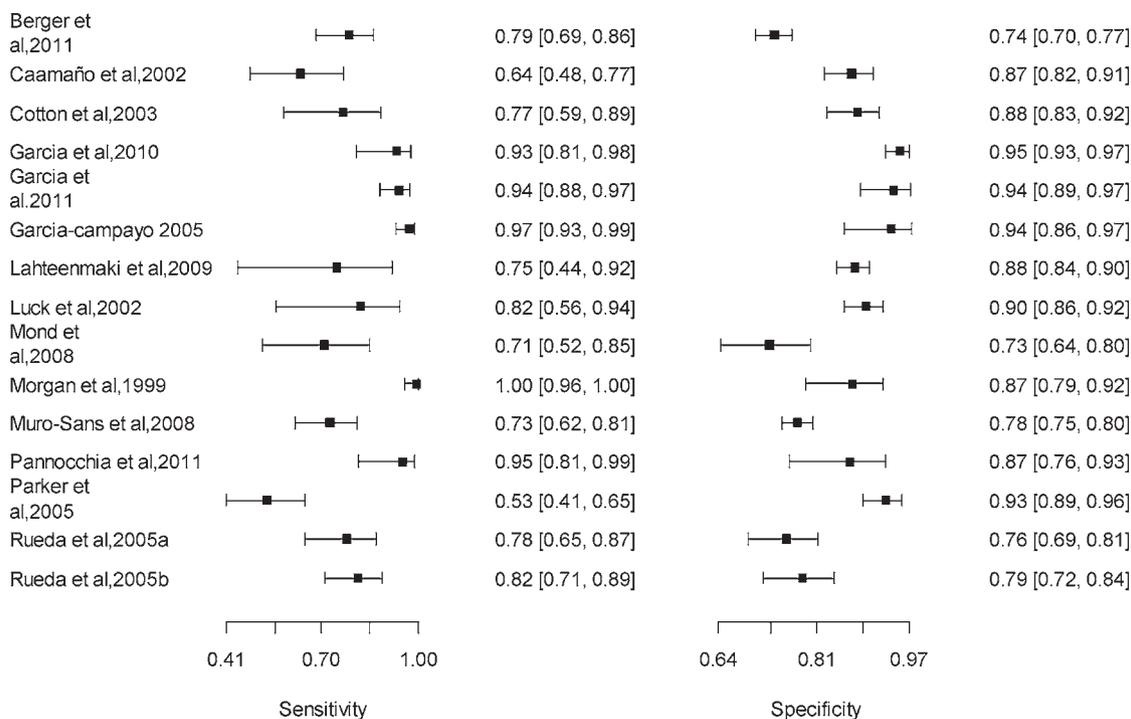


Figure 1. Forest plot of the 15 studies included in the meta-analysis.

for AN, BN and EDNOS (García-Campayo et al., 2005). In order to study this issue we selected the studies that provide relevant data. Of course, those are only the studies in which the measure of the diagnostic reference is based on any form of interview, given that the psychometric tests do not discriminate between the three diagnostics. Eight from the nine studies that employ some form of interview provide frequencies of positive classification by the SCOFF, disaggregated according to the diagnostic (see Table 1).

After testing as potential covariates the frequency of AN and BN we have concluded that they are not significantly different. When the proportion of EDNOS in the sample is included, however, a significant association with diagnostic efficacy is apparent. Specifically, the diagnostic efficacy is significantly better for EDNOS than for AN/BN ($p = .011$). This difference can be associated with a greater tendency of the AN/BN cases to deny the illness, hiding the symptoms when answering the SCOFF. It can also be related to a lower sensitivity in individuals with previous low weight. Specifically, question 3 of the SCOFF makes reference to a recent extreme weight loss. It is possible that patients with AN or BN consider that it has not happened recently, given that these patients reduce the Body Mass Index (BMI) step by step over a long period. Furthermore, the SCOFF was developed as a tool for detecting patients with ED in an early stage in the general population, not for identifying whether the adolescent has a BMI of 17.5 kg/m^2 , a key criterion for diagnosing a AN.

Discussion

The joint analysis of 15 studies that assess the diagnostic efficacy of the SCOFF shows that the test is highly effective as a quick screening tool for ED. Its brevity and simplicity have allowed easy translation into multiple languages that have shown comparable values of performance. The pooled estimates from the present 15 studies are .801 for sensitivity and .934 for specificity.

The analysis of the moderator variables shows that the diagnostic efficacy is associated with the gender and the type of diagnostic reference. Specifically, the efficacy increases as the percentage of women in the sample increases and if the diagnostic reference is based on an interview. The best estimates from these data are those combining the nine studies based on interviews; they provide values of .882 for sensitivity and .925 for specificity. The comparison of these values with those provided by the combined estimates of the 15 studies reflects another interesting characteristic. When interviews instead of psychometric tests are used the specificity remains practically unchanged; however, the sensitivity rises from .801 to .882. This increase reveals the ability of a professional (even a

non-specialist) to detect symptoms, without any associated increase in the false positives. That is, the effect is not based on a shift of the criteria, but on the genuine advantage of the interview over the psychometric tests in terms of discriminating between the cases and the controls.

Nevertheless, the effects associated to gender and the type of diagnostic reference have been detected with statistical tests only marginally significant ($p < .06$). Convergent evidence is needed to reinforce the role of both moderators.

In short, the results from the present meta-analysis reinforce the idea that the set of five questions of the SCOFF constitute a highly efficient tool for the detection of ED, even by a non-specialist, in several languages. Its use as a screening tool is highly recommended.

References

- American Psychiatric Association** (2002). *Diagnostic and statistical manual of mental disorders* (DSM-IV-TR; 4th Ed). Washington, DC: American Psychiatric Association.
- Becker A. E., Thomas J. J., Franko D. L., & Herzog D. B.** (2005). Disclosure patterns of eating and weight concerns to clinicians, educational professionals, family, and peers. *International Journal of Eating Disorders*, 38, 18–23. <http://dx.doi.org/10.1002/eat.20141>
- Berger U., Wick K., Hölling H., Schlack R., Bormann B., Brix C., ..., & Strau B.** (2011). Screening riskanten Essverhaltens bei 12-jährigen Mädchen und Jungen: Psychometrischer Vergleich der deutschsprachigen Versionen von SCOFF und EAT-26. [Screening of Disordered Eating in 12-Year-Old Girls and Boys: Psychometric Analysis of the German Versions of SCOFF and EAT-26]. *Psychotherapie, Psychosomatik & Medizinische Psychologie*, 61, 311–318. <http://dx.doi.org/10.1055/s-0031-1271786>
- Borenstein M., Hedges L. V., Higgins J. P. T., & Rothstein H. R.** (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1, 97–111. <http://dx.doi.org/10.1002/jrsm.12>
- Botella J., & Huang H.** (2012). Procedimientos de meta-análisis de la precisión de clasificaciones binarias. [Procedures for performing meta-analyses of the accuracy of tools for binary classification]. *Psicothema*, 24, 133–141.
- Botella J., Suero M., & Gambará H.** (2010). Psychometric inferences from a meta-analysis of reliability and internal consistency coefficients. *Psychological Methods*, 15, 386–397. <http://dx.doi.org/10.1037/a0019626>
- Caamaño F., Aguiar M., López-Otero M. J., & Takkouche B.** (2002). Validation of SCOFF questionnaire among preteenagers. *British Medical Journal*, (rapid response). Retrieved from http://www.bmj.com/search/Validation%20of%20SCOFF%20questionnaire%20among%20preteenagers?filters=type:bmj_rapid_response
- Cotrufo P., Barretta V., & Monteleone P.** (1998). Un estudio epidemiológico sobre los trastornos de la conducta alimentaria en dos institutos de segunda enseñanza en Nápoles. [An epidemiological study on eating disorders in

- two high schools in Naples] *European Psychiatry*, 12, 342–344. [http://dx.doi.org/10.1016/S0924-9338\(97\)80003-X](http://dx.doi.org/10.1016/S0924-9338(97)80003-X)
- Cotton M. A., Ball C., & Robinson P.** (2003). Four simple questions can help screen for eating disorders. *Journal of General Internal Medicine*, 18, 53–56. <http://dx.doi.org/10.1046/j.1525-1497.2003.20374.x>
- Devillé W. L., Buntinx F., Bouter L. M., Montori V. M., de Vet H. C., van der Windt D. A., & Bezemer P. D.** (2002). Conducting systematic reviews of diagnostic studies: Didactic guidelines. *BMC Medical Research Methodology*, 2, 2–9. <http://dx.doi.org/10.1186/1471-2288-2-9>
- Eddy K. T., Swanson S. A., Crosby R. D., Franko D. L., Engel S., & Herzog D. B.** (2010). How should DSM-V classify eating disorder not otherwise specified (EDNOS) presentations in women with lifetime anorexia or bulimia nervosa? *Psychological Medicine*, 40, 1735–1744. <http://dx.doi.org/10.1017/S0033291709992200>
- Favaro A., Ferrara S., & Santonastasio P.** (2003). The spectrum of eating disorders in young women: A prevalence study in a general population sample. *Psychosomatic Medicine*, 65, 701–708. <http://dx.doi.org/10.1097/01.PSY.0000073871.67679.D8>
- Garcia F. D., Grigioni S., Allais E., Houy-Durand E., Thibaut F., & Déchelotte P.** (2011). Detection of eating disorders in patients: Validity and reliability of the French version of the SCOFF questionnaire. *Clinical Nutrition*, 30, 178–181. <http://dx.doi.org/10.1016/j.clnu.2010.09.007>
- Garcia F. D., Grigioni S., Chelali S., Meyrignac G., Thibaut F., & Dechelotte P.** (2010). Validation of the French version of SCOFF questionnaire for screening of eating disorders among adults. *World Journal of Biological Psychiatry*, 11, 888–893. <http://dx.doi.org/10.3109/15622975.2010.483251>
- García-Campayo J., Sanz-Carrillo C., Ibañez J. A., Lou S., Solano S., & Alda M.** (2005). Validation of the Spanish version of the SCOFF questionnaire for the screening of eating disorders in primary care. *Journal of Psychosomatic Research*, 59, 51–55. <http://dx.doi.org/10.1016/j.jpsychores.2004.06.005>
- Gatsonis C., & Paliwal P.** (2006). Meta-analysis of diagnostic and screening test accuracy evaluations: Methodological primer. *American Journal of Roentgenology*, 187, 271–281. <http://dx.doi.org/10.2214/AJR.06.0226>
- Herpertz-Dahlmann B., Wille N., Hölling H., Vloet T. D., & Ravens-Sieberer U.** (2008). Disordered eating behavior and attitudes, associated psychopathology and health-related quality of life: Results of the BELLA study. *European Child & Adolescent Psychiatry*, 17, 82–91. <http://dx.doi.org/10.1007/s00787-008-1009-9>
- Hill L. S., Reid F., Morgan J. F., & Lacey J. H.** (2010). SCOFF, the development of an eating disorder screening questionnaire. *International Journal of Eating Disorders*, 43, 344–351. <http://dx.doi.org/10.1002/eat.20679>
- Lähteenmäki S., Aalto-Setälä T., Suokas J. T., Saarni S. E., Perälä J., Saarni S. I., ..., & Suvisaari J. M.** (2009). Validation of the Finnish version of the SCOFF questionnaire among young adults aged 20 to 35 years. *BMC Psychiatry*, 9. <http://dx.doi.org/10.1186/1471-244X-9-5>
- Luck A. J., Morgan J. F., Reid F., O'Brien A., Brunton J., Price C., ..., & Lacey J. H.** (2002). The SCOFF Questionnaire and clinical interview for eating disorders in general practice: Comparative study. *British Medical Journal*, 325, 755–756. <http://dx.doi.org/10.1136/bmj.325.7367.755>
- Macaskill P.** (2004). Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. *Journal of Clinical Epidemiology*, 57, 925–932. <http://dx.doi.org/10.1016/j.jclinepi.2003.12.019>
- Mond J. M., Myers T. C., Crosby R. D., Hay P. J., Rodgers B., Morgan J. F., ..., & Mitchell J. E.** (2008). Screening for eating disorders in primary care: EDE-Q versus SCOFF. *Behaviour Research and Therapy*, 46, 612–622. <http://dx.doi.org/10.1016/j.brat.2008.02.003>
- Morandé G., Celada J., & Casas J. J.** (1999). Prevalence of eating disorders in a Spanish school age population. *Journal of Adolescent Health*, 24, 212–219. [http://dx.doi.org/10.1016/S1054-139X\(98\)00025-1](http://dx.doi.org/10.1016/S1054-139X(98)00025-1)
- Morgan J. F., Reid F., & Lacey J. H.** (1999). The SCOFF questionnaire: Assessment of a new screening tool for eating disorders. *British Medicine Journal*, 319, 1467–1468. <http://dx.doi.org/10.1136/bmj.319.7223.1467>
- Muro-Sans P., Amador-Campos J. A., & Morgan J. F.** (2008). The SCOFF-C: Psychometric properties of the Catalan version in a Spanish adolescent sample. *Journal of Psychosomatic Research*, 64, 81–86. <http://dx.doi.org/10.1016/j.jpsychores.2007.06.011>
- Ogg E. C., Millar H. R., Pusztai E. E., & Thom A. S.** (1997). General practice consultation patterns preceding diagnosis of eating disorders. *International Journal of Eating Disorders*, 22, 89–93. [http://dx.doi.org/10.1002/\(SICI\)1098-108X\(199707\)22:1<89::AID-EAT12>3.0.CO;2-D](http://dx.doi.org/10.1002/(SICI)1098-108X(199707)22:1<89::AID-EAT12>3.0.CO;2-D)
- Pannocchia L., Di Fiorino M., Giannini M., & Vanderlinden J. A.** (2011). Psychometric exploration of an Italian translation of the SCOFF questionnaire. *European Eating Disorders Review*, 19, 371–373. <http://dx.doi.org/10.1002/erv.1105>
- Parker S. C., Lyons J., & Bonner J.** (2005). Eating disorders in graduate students: Exploring the SCOFF questionnaire as a simple screening tool. *Journal of American College Health*, 54, 103–107. <http://dx.doi.org/10.3200/JACH.54.2.103-107>
- 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 en las adolescentes Navarras. [Prevalence of eating disorders in a representative sample of female adolescents from Navarra-Spain] *Medicina Clínica-Barcelona*, 114, 481–486. [http://dx.doi.org/10.1016/S0025-7753\(00\)71340-6](http://dx.doi.org/10.1016/S0025-7753(00)71340-6)
- Review Manager** (2008). Version 5.0. Copenhagen, Denmark: The Nordic Cochrane Centre, The Cochrane Collaboration.
- Rojo L., Livianos L., Conesa L., Garcia A., Domínguez A., Rodrigo G., ..., & Vila M.** (2003). Epidemiology and risk factors of eating disorders: A two-stage epidemiologic study in a Spanish population aged 12–18 years. *International Journal of Eating Disorders*, 34, 281–291. <http://dx.doi.org/10.1002/eat.10179>
- Rueda G. E., Díaz L. A., Campo A., Barros J. A., Avila G. C., Oróstegui L. T., ..., & Cadena L. P.** (2005). Validation of the

SCOFF questionnaire for screening of eating disorders in university women. *Biomédica*, 25, 196–202.

Rueda G. E., Díaz L. A., Ortiz D. P., Pinzón C., & Rodrigues J. (2005). Validación del cuestionario SCOFF para el cribado de los trastornos del comportamiento alimentario en adolescentes escolarizadas. [Validation of the SCOFF questionnaire for screening the eating behavior disorders of adolescents in school] *Atención Primaria*, 35, 89–94.

Ruiz-Lazaro 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. [Study of the prevalence of eating disorders in adolescents in Zaragoza] *Revista de Psiquiatría Infanto-Juvenil*, 3, 148–162

SAS Institute Inc. (2008). The SAS System for Windows. Version 9.2. Cary, NC: SAS Institute Inc.

Steinhausen H. C., Winkler C., & Meier M. (1997). Eating disorders in adolescence in a Swiss epidemiological study. *International Journal of Eating Disorders*, 22, 147–151. [http://dx.doi.org/10.1002/\(SICI\)1098-108X\(199709\)22:2<147::AID-EAT5>3.3.CO;2-I](http://dx.doi.org/10.1002/(SICI)1098-108X(199709)22:2<147::AID-EAT5>3.3.CO;2-I)

Zamora J., Abaira V., Muriel A., Khan K., & Coomarasamy A. (2006). Meta-Disc: A software for meta-analysis of test accuracy data. *BMC Medical Research Methodology*, 6. <http://dx.doi.org/10.1186/1471-2288-6-31>

Zweig M. H., & Campbell G. (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39, 561–577.

Appendix A

The five questions of the SCOFF:

- S Do you make yourself Sick because you feel uncomfortably full?
- C Do you worry you have lost Control over how much you eat?
- O Have you recently lost more than One stone in a three-month period?
- F Do you believe yourself to be Fat when others say you are too thin?
- F Would you say that Food dominates your life?