

Instructions for Using the MInCir Scale to Assess Methodological Quality in Diagnostic Accuracy Studies

Instrucciones para Utilizar la Escala MInCir para Valorar Calidad Metodológica de Estudios de Diagnóstico o Pruebas Diagnósticas

Carlos Manterola^{*,**,***,****}; Ricardo Cartes-Velasquez^{***,*****} & Tamara Otzen^{***,****,*****}

MANTEROLA, C.; CARTES-VELASQUEZ, R. & OTZEN, T. Instructions for using the MInCir scale to assess methodological quality in diagnostic accuracy studies. *Int. J. Morphol.*, 34(1):78-84, 2016.

SUMMARY: Globally, the assessment of methodological quality (MQ) in biomedical research is an area of dynamic development over recent years. Understood as a complex and multidimensional construct, various groups have proposed tools for its assessment. Our team has designed and validated scales to assess MQ of therapy and prognosis studies. However, as with other instruments, it is necessary to specify in detail how it is applied, in such a way as to be able to standardize the measurements made with this instrument. A detailed description is presented of the 3 domains (type of research design [single item domain], size of population studied [single item domain], and methodology [7 items]) and 9 items comprising the scale, specifying for each item how to assess the characteristics and score articles of diagnostic accuracy studies. The application of this instrument requires critical evaluation of each of the items to assign the respective score, which sum to give a total score of between 9 and 45 points. The cut-off point to define the construct MQ (dichotomous) for this type of study is 20 points. This article provides an application guideline that may help to improve inter-observer and intra-observer reliability of the MInCir MQ scale for diagnostic accuracy studies. The aim of this article is to provide a guideline for the standardized application of the MInCir MQ scale for diagnostic accuracy studies.

KEY WORDS: Investigative Techniques; Methods/epidemiology; methodological studies; Diagnosis; Diagnostic tests; Reproducibility of Results; Evaluation Studies as Topic; Epidemiological Studies; Evidence-Based Medicine.

INTRODUCTION

Every day it is becoming more difficult to access and use the growing amount of scientific information available on the various biomedical databases, as the time available to achieve this objective remains the same or is reducing.

On the other hand, it is obvious that in this era of evidence based clinical practice (EBCP) it is essential to have tools to critically evaluate the literature in order to filter quality research in reasonable time, so as to optimize our time and provide care to patients based on the best available evidence (du Prel *et al.*, 2009). Nevertheless, there is still no consensus on the best way to evaluate the quality of biomedical research and a plethora of concepts, strategies and tools to address this deficiency has been created (Cascais da Silva *et al.*, 2013).

Therefore, assessing the methodological quality (MQ) of a scientific article is even more complex. The MQ construct can be understood as a multidimensional concept to evaluate multiple characteristics of an article, such as: type of design, sample size used, methodology and analysis used, quality of reporting, etc. (Manterola *et al.*, 2009).

Since 2000, we have initiated a research line aimed at measuring the MQ in different clinical scenarios: therapy or therapeutic procedures, prognosis, and diagnosis. The first was a scale, valid and reliable for assessing MQ in therapy or therapeutic procedures. This has allowed us to weigh the evidence and conduct systematic reviews (SR) and bibliometric studies (Manterola *et al.*, 2005a, 2005b, 2006,

* Department of Surgery, Universidad de La Frontera, Temuco, Chile.

** Center for Morphological and Surgical Studies (CEMyQ), Universidad de La Frontera, Temuco, Chile.

*** Ph.D. Program in Medical Sciences, Universidad de La Frontera, Temuco, Chile.

**** Center for Biomedical Research, Universidad Autónoma de Chile, Temuco, Chile.

***** Faculty of Dentistry, Universidad de Concepción, Concepción, Chile.

***** Fellow Research, Universidad Científica del Sur, Lima, Perú.

2009; Cartes-Velasquez *et al.*, 2014). Later scales were designed to assess MQ in prognosis and diagnostic accuracy studies (Burgos *et al.* 2011).

Why diagnostics? Because diagnosis is one of the most important areas of clinical work; because there is interest in evaluating the diagnostic tests in terms of the impact they can have on the outcome of the diagnostic and therapeutic process; because there are a number of difficulties in designing studies of this type (Dinnes *et al.*, 2005). Moreover, because the studies related to diagnostic tests have a number of peculiarities in terms of MQ; for example, the wide range of biases inherent in this type of study and corresponding strategies to minimize them (Whiting *et al.*, 2004a). Therefore, there are no MQ scales for diagnostic accuracy studies, because "... these criteria have not been uniform... have not been well described and quantitative techniques for assessing quality have not been incorporated" (Mulrow *et al.*, 1989).

The aim of this article is to provide a guideline for the standardized application of the MInCir MQ scale for diagnostic accuracy studies.

INSTRUCTIONS FOR USING THE MInCir MQ SCALE

The MInCir scale to assess MInCir MQ for diagnostic accuracy studies consists of 3 domains with 9 items. The first domain (first item, single item domain) evaluates the type of research design, domain 2 (second item, single item domain) assigns scores to the size of population studied, including a justification factor (if the study includes a sample size calculation) and domain 3, consists of 7 items that assess the study's methodology (Fig. 1).

The application of this instrument by the user requires critical evaluation of each item to assign the respective score, which sum to give a total score of between 9 and 45 points. The cut-off point to define the construct MQ (dichotomous) for this type of study is 20 points; allowing differentiation as adequate or high MQ when the evaluated article receives a score ≥ 20 points; and inadequate or low MQ if the score is ≤ 19 points (Burgos *et al.*).

Below is a detailed guide to each domain and item contained in the scale for a correct interpretation and subsequent implementation:

Domain 1: Study design

The type of study design must be mentioned in the "materials and methods" of the article and must be the same as indicated

in the summary. This is what should be valued according to the scores indicated by the scale. Thus, a validating studies test with good reference standard will be assigned 12 points. Moreover, at the lower end of the domain, a case-series will be assigned only 1 point.

However, on many occasions it is difficult to clearly identify the study design, a situation that usually occurs when the authors, for various reasons, do not disclose or rather "make-up" the design used, for example:

a) Articles in which authors do not make an appropriate indication and description of the design used in the "material and methods" of the study, nor in the summary (or just do not report it); so it is necessary to determine from a thorough reading of the article which design was used.

b) In other articles, the authors make a brief description of some design features, such as: a diagnostic test study, instead of stating whether it is a validating studies test with good or poor reference standards; or retrospective study, instead of stating retrospective case series.

c) On other occasions, the study is not consistent with the design reported by the authors. In such situations, what is seen most often is that a type of design with a greater level of evidence is reported than was actually performed. For example, to declare a case control study was carried out, when in fact it was a case series.

d) Lastly, some articles mentioned a design that does not exist: for example, "case-control diagnostic study."

In such situations, it is necessary to rate what corresponds to the design after carefully reading the article. When the type of design used is not mentioned, or refers to nonexistent designs, it is usually a case series, thus it is appropriate to score 1 point. If the authors report having developed a case control study in the article, the odds ratios and confidence intervals must appear as an outcome (it is very unlikely that a research group would conduct a study of these features and not properly report on them). It is equally acceptable to give a description of the characteristics of a design, for example: "prospective study with reference standards," rather than cohort study with reference standards, a situation that would warrant an allocation of 9 points.

Domain 2: Population studied and justification factor

In this domain a score is given according to the number of patients who actually participated in the study: that is, those actually included in the analysis of the reported results. Also it should be considered whether this number of

Fig. 1. MInCir scale to determine MQ for diagnostic accuracy studies.

Domains and items of the scale	Scoring (points)
Domain 1: Study design.	
Validating studies test with good reference standards *	12
Exploratory cohort study with reference standards **	9
Validating studies test with poor reference standards ***	6
Case control study poor or not-independent reference standard	4
Poor quality cohort studies ****	3
Case-series	1
Domain 2: Population studied x justification factor.	
≥ 201	6 or 12
151 - 200	5 or 10
101 - 150	4 or 8
61 - 100	3 or 6
31 - 60	2 or 4
≤ 30	1 or 2
Domain 3: Methodology.	
Item 1. Objectives	
Clear and concrete objectives	3
Vague objectives	2
No objective	1
Item 2. Design	
Clearly identified design	3
Unknown design	1
Item 3. Selection criteria	
Inclusion and exclusion criteria are described	3
Inclusion or exclusion is described	2
No selection criteria are described	1
Item 4. Characterization of the population under study	
The spectrum of the study subjects is representative of the population for which it is desired to extrapolate the results	3
The spectrum of the study subjects is partially representative of the population for which it is desired to extrapolate the results	1
Item 5. Characteristics of the reference standard applied	
The same reference standard is applied to all the patients independent of the result	3
The reference standard is applied partially	2
There is no report of the standard of reference used	1
Item 6. Characteristics of the diagnostic test under study	
The diagnostic test under study is described with sufficient detail to permit the replication of it	3
The diagnostic test under study is described only partially	2
The authors do not provide elements concerning the diagnostic test under study that allow the study to be replicated	1
Item 7. Sample size	
Sample size is justified	3
Sample size is not justified	1
Final score (item 1 + item 2 + item 3)	9 - 45

*Validating studies test the quality of a specific diagnostic test, based on prior evidence. Good reference standards are independent of the test, and applied blindly or applied objectively to all patients.

**Exploratory studies collect evidence and search the data to find which factors are important.

***Poor reference standards are arbitrarily applied, but independent of the test.

****Includes non-consecutive study without consistently applied reference standards.

patients was estimated with a suitable statistical procedure during the planning of the study (estimate or calculation of the minimum sample size to conduct the study). This will allow the "justification factor" to be applied that may double the score of the domain. For example, a study in which 300 patients were studied will have at least 6 points, but if the minimum sample size for conducting the study was estimated, then 12 points will be assigned. In the case of a study with 135 patients, with and without justification values, it will be awarded 4 and 8 points respectively. Thus, recognizing the "justification factor" is important, since in many cases the studies are conducted including only the available patients without consideration to ensure the statistical power of the results.

Domain 3: Methodology

This domain assesses seven aspects of the methodology: objective, design, selection criteria, characterization of the population in study, characteristics of the standard reference applied, characteristics of the diagnostic test under study, and sample size.

Item 1. Objectives: They are usually found at the bottom of the introduction. In some cases and depending on the format of the journal, they could be at the beginning of the methodology or in another section. There must be consistency between the objectives stated in the body of the text and the summary, but considering that the latter could be in a slightly abridged form, according to the journal's editorial criteria. In shorter articles, such as the cases series, the objective may only mentioned in the summary. Following this line of argument, objectives can be assessed at three levels:

1. Clear and concrete objectives: In this case, it is possible to clearly recognize what will be measured, by what means, by whom, where and the follow-up. The methodology is based on this objective. Three points are assigned.

2. Vague objectives: When any of the aforementioned elements is absent or improperly described, such as: "The aim of this study is to report our experience with...", "compared to conventional diagnosis", etc., in this type of scenario it is not possible to clearly recognize the purpose of the study. Two points are assigned.

3. No objective: If these are not visible in any of the previously mentioned places, one point is assigned.

Item 2. Design: This will be assessed in the methodology, sometimes entitled "Methods", "Materials and Methods" or "Patients and Methods." Similar to the objectives, the

design must also be mentioned in the article's summary, even in the title. For a case report, the design may be mentioned only in the title or in the abstract. Options for this item are as follows:

1. Clearly identified design: It is imperative that the reported design corresponds to that actually carried out. It is equally acceptable to have a description of the characteristics of a design, for example: "retrospective study," rather than poor quality cohort studies. Three points will be allocated.

2. Unknown design: It is not possible to identify the design or the design does not conform to what is actually carried out. One point is assigned. For example, the authors report that the design is a case control study, when in fact it is a case series.

Item 3. Selection criteria: These will be assessed in the methodology, sometimes entitled "Methods", "Materials and Methods" or "Patients and Methods." Options for this item are as follows:

1. Inclusion and exclusion criteria are described: These will be considered when both are clearly described by the authors. It is equally acceptable if the authors mention the inclusion criteria and declare the absence of exclusion criteria. Three points are assigned.

2. Inclusion or exclusion is described: Only one of the two criteria is described, regardless of the location in "Material and Methods." Two points are allocated.

3. No selection criteria are described: The authors declare no criteria. One point is assigned.

Item 4. Characterization of the population under study: This will be assessed in the methodology, sometimes-entitled "Methods," "Materials and Methods" or "Patients and Methods." Options for this item are as follows:

1. The spectrum of the study subjects is representative of the population for which it is desired to extrapolate the results: Clinical and bio-socio-demographic characteristics of the sample should be described and must be representative of the target population. Three points will be assigned.

2. The spectrum of the study subjects is partially representative of the population for which it is desired to extrapolate the results: Clinical and bio-socio-demographic characteristics of the sample should be described and are partially representative of the target population. One point is allocated.

Item 5. Characteristics of the reference standard applied:

These will be assessed in the methodology, sometimes-entitled "Methods," "Materials and Methods" or "Patients and Methods." Options for this item are as follows:

1. The same reference standard is applied to all the patients independent of the result: There is a clear description and justification of the reference standard. In all cases this standard is used, no matter the results of the diagnostic test under study. Three points will be allocated.
2. The reference standard is applied partially: There is a vague description and justification for it. In most cases this standard is used, depending on the results of the diagnostic test under study. Two points will be assigned.
3. There is no report of the standard of reference used. One point is allocated.

Item 6. Characteristics of the diagnostic test under study:

These will be assessed in the methodology, sometimes entitled "Methods," "Materials and Methods" or "Patients and Methods." Options for this item are as follows:

1. The diagnostic test under study is described with sufficient detail to permit the replication of it: The authors give enough information about the use of the test in the study, including technical specifications and the users' guide. Three points will be allocated.
2. The diagnostic test under study is described only partially: The authors give partial information about the use of the test under evaluation, including some technical specifications or the users' guide. Two points will be assigned.
3. The authors do not provide elements concerning the diagnostic test under study that allow the study to be replicated. One point is given.

Item 7. Sample size: The sample size may or may not be stated by the authors; therefore, the options for this item are as follows:

1. Sample size is justified: In this case the estimation or calculation of the sample size is reported. Another option is to justify the sample size, for example, when all subjects in the universe or target population were included. Three points are assigned.
2. Sample size is not justified: No reference to an estimate or calculation of the minimum sample size for conducting the study is mentioned. This is routine in the case series. One point is assigned.

DISCUSSION

Measurement is one of the core activities of a research process in all its disciplines. Measurement allows to define dimensions and to categorize, thus facilitating the description and communication of results; part of the daily activity of clinicians and researchers. This becomes critical when you want to measure the existence and severity of a disease or event of interest (Streiner & Geoffrey, 2003).

Today there are thousands of diagnostic tests, which are routinely used in patient care. In everyday clinical practice processes, these help both diagnostic and therapeutic decision-making. It is therefore vital that the properties of these tests are optimal, so that they can provide valid and reliable information. However, determining the accuracy of a diagnostic test requires investigation that meets minimum criteria MQ (Cook *et al.*, 2007).

Studies of diagnostic tests have usually been directed towards one of two principal objectives: The first one is to determine the diagnostic accuracy of the test, which is generally achieved by reporting sensitivity, specificity, likelihood ratios and the predictive value of the test. The second one is to estimate the impact of one or more diagnostic approaches (for example, combining diagnostic tests) on therapeutic decisions or patient outcomes. Both objectives have to be considered in the selected study methodology (Whiting *et al.*, 2004a).

In order to guide and evaluate research on diagnostic test accuracy, we have developed a number of tools, among which there are checklists for the preparation of studies and publication of results, such as STARD (Standards for Reporting of Diagnostic Accuracy), QUADAS (Quality Assessment of Diagnostic Accuracy Studies) and QAREL (Quality Appraisal of Reliability Studies). However, the process of validation of such systems has not been sufficient to determine MQ (Begg *et al.*, 1998; Bossuyt *et al.*, 2003, 2015; Whiting *et al.*, 2006; Cook *et al.*; Simel *et al.*, 2008; Kim *et al.*, 2009; Fontela *et al.*, 2009; Lucas *et al.*, 2010; MacPherson *et al.*, 2010).

Following this line of thought, we designed and developed this scale to measure MQ in diagnostic accuracy studies; obtaining a valid and reliable scale (Burgos *et al.*). Although the use of this scale may appear to be self-explanatory, using this and other scales developed by the group have shown that the valuation of the items is a more complicated task. Because the quality of reporting is low, there is poor writing and the use of methodological terminology is inadequate in a substantial proportion of the

studies, this increases the difficulty of establishing this construct (Manterola *et al.*, 2006).

It is important to remember that biases may be present not only in the measuring instrument itself but also in the evaluation using the instrument (Manterola & Otzen, 2015), so it is important that this has the greatest number of resources to make valid and reliable measurements; especially when the instrument used is dependent on the evaluator, such as these MQ scales. Moreover, there is evidence suggesting that the major sources of bias in these studies come from the prevalence of the demographic characteristics of the population and the event of interest under study (Whiting *et al.*, 2004b). Other sources of bias to be considered in studies of diagnostic tests are if the case-control design was used, the variability of the observer, the availability of clinical information; and prevalence and severity of the disease or event of interest under study (Whiting *et al.*, 2013).

We hope these instructions provide a useful contribution to researchers, readers, reviewers and editors of medical journals; assuming that they should be subject to periodic review and updating by users.

MANTEROLA, C.; CARTES-VELASQUEZ, R. & OTZEN, T. Instrucciones para utilizar la escala MInCir para valorar calidad metodológica de estudios de diagnóstico o pruebas diagnósticas. *Int. J. Morphol.*, 34(1):78-84, 2016.

RESUMEN: La evaluación de la calidad metodológica (CM) en la investigación biomédica es un área de desarrollo dinámico en los últimos años. Entendida, como un constructo multidimensional y complejo; diversos grupos han propuesto herramientas para su valoración. Nuestro equipo de trabajo ha diseñado y validado escalas para valorar CM de estudios de terapia y pronóstico. Sin embargo, como ocurre con otros instrumentos, es necesario especificar detalladamente la forma en que se aplica, de forma tal de poder estandarizar las mediciones. De este modo, se presenta una descripción pormenorizada de los 3 dominios (tipo de diseño, tamaño de la población en estudio y metodología) y 9 ítems que componen la escala, precisando para cada uno de ellos las características a evaluar y puntuar en los artículos primarios de diagnóstico o pruebas diagnósticas. La aplicación de este instrumento requiere de la evaluación crítica de cada ítem para asignar una puntuación final que en su sumatoria puede variar entre 9 y 45 puntos, con un punto de corte de 20 puntos que define el constructo CM para este tipo de estudios. Este artículo aporta una guía de aplicación que permite otorgar una adecuada confiabilidad intra e inter observador a las mediciones que se realicen aplicando la escala MInCir de CM para valorar estudios primarios de escenarios de diagnóstico o pruebas diagnósticas. El objetivo de este artículo es proporcionar una guía para la aplicación estandarizada de la escala MInCir de CM para valorar estudios primarios de escenarios de diagnóstico o pruebas diagnósticas.

PALABRAS CLAVE: Técnicas de Investigación; Metodología; Calidad metodológica; Diagnóstico; Pruebas diagnósticas; Estudios de evaluación; Estudios epidemiológicos; Medicina basada en evidencia.

REFERENCES

- Begg, C.; Cho, M.; Eastwood, S.; Horton, H.; Moher, D.; Olkin, I.; Pitkin, R.; Rennie, D.; Schulz, K. F.; Simel, D. & Stroup, D. F. Improving the quality of reports on randomized controlled trials. Recommendations of the CONSORT Study Group. *Rev. Esp. Salud Publica*, 72(1):5-11, 1998.
- Bossuyt, P. M.; Reitsma, J. B.; Bruns, D. E.; Gatsonis, C. A.; Glasziou, P. P.; Irwig, L. M.; Moher, D.; Rennie, D.; de Vet, H. C.; Lijmer, J. G. & Standards for Reporting of Diagnostic Accuracy. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann. Intern. Med.*, 138(1):W1-12, 2003.
- Bossuyt, P. M.; Reitsma, J. B.; Bruns, D. E.; Gatsonis, C. A.; Glasziou, P. P.; Irwig, L.; Lijmer, J. G.; Moher, D.; Rennie, D.; de Vet, H. C.; Kressel, H. Y.; Rifai, N.; Golub, R. M.; Altman, D. G.; Hooft, L.; Korevaar, D. A.; Cohen, J. F. & STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *B. M. J.*, 351:h5527, 2015.
- Burgos D., M. E.; Manterola D., C. & Sanhueza C., A. Diseño de una escala para evaluar calidad metodológica de estudios de pruebas diagnósticas. Estudio piloto. *Rev. Chil. Cir.*, 63(5):493-7, 2011.
- Cartes-Velásquez, R. A.; Manterola, C.; Aravena, P. & Moraga, J. Reliability and validity of MINCIR scale for methodological quality in dental therapy research. *Braz. Oral Res.*, pii: S1806-83242014000100232, 2014.
- Cascaes da Silva, F.; Valdivia Arancibia, B. A.; da Rosa Iop, R.; Barbosa Gutierrez Filho, P. J. & da Silva, R. Escalas y listas de evaluación de la calidad de estudios científicos. *Rev. Cuba. Inf. Cienc. Salud*, 24(3):295-312, 2013.
- Cook, C.; Cleland, J. & Huijbregts, P. Creation and Critique of Studies of Diagnostic Accuracy: Use of the STARD and QUADAS Methodological Quality Assessment Tools. *J. Man. Manip. Ther.*, 15(2):93-102, 2007.
- Dinnes, J.; Deeks, J.; Kirby, J. & Roderick, P. A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. *Health Technol. Assess.*, 9(12):1-113, 2005.
- du Prel, J. P.; Röhrig, B. & Blettner, M. Critical appraisal of scientific articles: part 1 of a series on evaluation of scientific publications. *Dtsch. Arztebl. Int.*, 106(7):100-5, 2009.

- Fontela, P. S.; Pant Pai, N.; Schiller, I.; Dendukuri, N.; Ramsay, A. & Pai, M. Quality and reporting of diagnostic accuracy studies in TB, HIV and malaria: evaluation using QUADAS and STARD standards. *PLoS One*, 4(11):e7753, 2009.
- Kim, D. W.; Park, A. W.; Lee, E. J.; Choo, H. J.; Kim, S. H.; Lee, S. H. & Eom, J. W. Ultrasound-guided fine-needle aspiration biopsy of thyroid nodules smaller than 5 mm in the maximum diameter: assessment of efficacy and pathological findings. *Korean J. Radiol.*, 10(5):435-40, 2009.
- Lucas, N. P.; Macaskill, P.; Irwig, L. & Bogduk, N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J. Clin. Epidemiol.*, 63(8):854-61, 2010.
- MacPherson, H.; Altman, D. G.; Hammerschlag, R.; Youping, L.; Taixiang, W.; White, A.; Moher, D. & STRICTA Revision Group. Revised Standards for Reporting Interventions in Clinical Trials of Acupuncture (STRICTA): extending the CONSORT statement. *PLoS Med.*, 7(6):e1000261, 2010.
- Manterola, C.; Pineda, V. & Vial, M. Open versus laparoscopic resection in non-complicated colon cancer. A systematic review. *Cir. Esp.*, 78(1):28-33, 2005a.
- Manterola, C.; Pineda, V.; Vial, M.; Losada, H. & Muñoz, S. Surgery for morbid obesity: selection of operation based on evidence from literature review. *Obes. Surg.*, 15(1):106-13, 2005b.
- Manterola, C.; Pineda, V.; Vial, M.; Losada, H. & MINCIR Group. What is the methodologic quality of human therapy studies in ISI surgical publications? *Ann. Surg.*, 244(5):827-32, 2006.
- Manterola, C.; Vial, M.; Pineda, V. & Sanhueza, A. Systematic review of literature with different types of designs. *Int. J. Morphol.*, 27(4):1179-86, 2009.
- Manterola, C. & Otzen, T. Los sesgos en investigación clínica. *Int. J. Morphol.*, 33(3):1156-64, 2015.
- Mulrow, C. D.; Linn, W. D.; Gaul, M. K.; Pugh, J. A. Assessing quality of a diagnostic test evaluation. *J. Gen. Intern. Med.*, 4(4):288-95, 1989.
- Simel, D. L.; Rennie, D. & Bossuyt, P. M. The STARD statement for reporting diagnostic accuracy studies: application to the history and physical examination. *J. Gen. Intern. Med.*, 23(6):768-74, 2008.
- Streiner, D. L. & Norman, G. R. *Health Measurement Scales: A Practical Guide to Their Development and Use*. 3rd ed. Oxford, Oxford University Press, 2003.
- Whiting, P.; Rutjes, A. W.; Dinnes, J.; Reitsma, J. B.; Bossuyt, P. M. & Kleijnen, J. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol. Assess.*, 8(25):iii, 1-234, 2004a.
- Whiting, P.; Rutjes, A. W.; Reitsma, J. B.; Glas, A. S.; Bossuyt, P. M. & Kleijnen, J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann. Intern. Med.*, 140(3):189-202, 2004b.
- Whiting, P. F.; Westwood, M. E.; Rutjes, A. W.; Reitsma, J. B.; Bossuyt, P. N. & Kleijnen, J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med. Res. Methodol.*, 6:9, 2006.
- Whiting, P. F.; Rutjes, A. W.; Westwood, M. E.; Mallett, S. & QUADAS-2 Steering Group. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J. Clin. Epidemiol.*, 66(10):1093-104, 2013.

Correspondence to:
Dr. Carlos Manterola
Department of Surgery and CEMyQ
Universidad de La Frontera.
Manuel Montt 112, Office 408
Temuco
CHILE

Phone: 56-45-2325760
Fax: 56-45-2325761

Email: carlos.manterola@ufrontera.cl

Received: 16-09-2015
Accepted: 22-12-2015