

Cosa è l'iniziativa Stard?

(*STARD Initiative* proposta dal Gruppo *Standards for Reporting Diagnostic Accuracy*)

Un evento di particolare rilievo durante l'ultimo convegno nazionale Sibioc tenutosi a Padova è stata la lettura di Patrick M. Bossuyt, figura di riferimento per avere promosso l'iniziativa STARD, dal titolo "The potential for bias and the need for clear reporting in studies of diagnostic accuracy".

Perché è importante l'iniziativa STARD e perché questa sigla sarà sempre più citata?

Quando si cerca di valutare un test diagnostico Evidence Based, una delle difficoltà maggiori è di trovare pubblicazioni scientifiche primarie metodologicamente corrette, in altre parole non vi è nel disegno sperimentale dei lavori scientifici che si propongono di valutare un test diagnostico una metodologia forte simile allo studio in doppio cieco randomizzato utilizzato nei trial farmacologici. Questo rende difficile e di conseguenza "debole" una valutazione EBM del momento diagnostico. E' stato fatto il tentativo di introdurre una precisa checklist per valutare i lavori scientifici in rapporto alla accuratezza diagnostica, questa checklist già proposta nel 1997 è stata recentemente rivista. Nell'ultima revisione sono stati raccolti i suggerimenti inviati da diversi specialisti, sia di medicina di laboratorio che di epidemiologia clinica, statistica, da editori di riviste e ricercatori.

Risultato finale di questa attività in corso da anni è la *STARD Initiative* proposta dal Gruppo *Standards for Reporting Diagnostic Accuracy*, lo scopo del documento è migliorare la qualità strutturale ed espositiva degli studi concernenti l'accuratezza di test diagnostici, mediante l'impiego di una *checklist* e di un diagramma di flusso. La descrizione completa ed accurata dell'architettura di uno studio permette al lettore di scoprirne i potenziali vizi strutturali (*bias*) e di valutare se i suoi risultati sono generalizzabili e applicabili nella pratica clinica.

Nel 2003 lo **STARD** è stata pubblicato da Clinical Chemistry e dal British Medical con un editoriale dedicato e da Annals of Internal Medicine e Radiology. In particolare è da sottolineare come la cultura basata sull'Evidence Based Medicine trovi grande attenzione non solo sul BMJ, da sempre la rivista più "EBM", ma divenga il punto di riferimento metodologico per Clinical Chemistry che richiede come i contributi scientifici si adeguino alle richieste presenti nello STARD con un sostanziale approccio orientato all'EBLM (*BMJ 2003;326:3-4, BMJ 2003; 326:41-4, Clin Chem 2003;49:1-6, Clin Chem 2003;49:7-18, Ann Intern Med 2003;138:40-4*).

Lo **STARD** è composto da una check-list di 25 punti, che dal titolo alla discussione definisce i requisiti essenziali per la presentazione degli studi di accuratezza diagnostica. L'accettazione e l'applicazione da parte delle riviste mediche di questo standard dovrebbe garantire sia la completezza e l'accuratezza della presentazione dei risultati, sia la comparabilità ed il trasferimento delle informazioni alla pratica clinica.

La *checklist* è il cuore del documento e serve a controllare punto per punto la stesura di un articolo nelle sue varie sezioni. Ha la stessa struttura per tutte le attività diagnostiche, nell'assunto che il termine test diagnostico definisca una qualsiasi indagine idonea ad ottenere informazioni aggiuntive sullo stato di salute di una persona, sia essa di natura ematochimica, microbiologica, anatomo-patologica, funzionale o di *imaging*.

L'obiettivo dello Stard è di avere a disposizione studi primari sul test diagnostico di buona qualità a permettere metanalisi o revisioni sistematiche metodologicamente corrette che possano essere alla base per la definizione di linee guida cliniche supportate da evidenze robuste e rilevanti.

In questo numero di EBLM newsletter si riporta sia l'originale in inglese che una traduzione italiana ad opera di Pino Giocoli.

(*Tommaso Trenti, Dipartimento di Patologia Clinica, Ausl di Modena, t.trenti@ausl.mo.it*)

TOWARDS COMPLETE AND ACCURATE REPORTING OF STUDIES ON DIAGNOSTIC ACCURACY: THE STARD INITIATIVE

The STARD group
Test version, November 2001

This is a test version of the STARD checklist, the STARD flow diagram and the STARD statement. For evaluation purposes only.

The STARD group welcomes all comments, whether content or form, to improve the test version. Your comments and suggestions will be discussed among the members of STARD group during the final stage of preparing the first official version of the STARD checklist, flow diagram and statement. The official versions will be submitted for publication.

Addresses for comments and suggestions:

E-mail: stard@amc.uva.nl
Phone: +31-20-5666694
Fax: +31-20-6912683

Hans Reitsma MD PhD
Department of Clinical Epidemiology and Biostatistics
Academic Medical Center – University of Amsterdam
PO Box 22700, 1100 DE Amsterdam, The Netherlands

ABSTRACT

Objective - The objective of the Standards for Reporting of Diagnostic Accuracy (STARD) initiative is to improve the quality of reporting of studies on diagnostic accuracy by using a checklist and flow diagram. Complete and accurate reporting allows the reader to detect the potential for bias in the study and to evaluate the generalisability of the results.

Methods - The STARD steering group started an extensive search of the literature to identify publications on the conduct and reporting of diagnostic studies. Potential items were extracted into a long list. This long list was shortened, where appropriate, during a two-day consensus meeting attended by researchers, editors, and members of professional organisations. Special attention was given to the development of a generic flow diagram for studies of diagnostic accuracy.

Results - The search on published guidelines regarding diagnostic research yielded 33 lists, which resulted in a long list of 75 potential items. During a consensus meeting, the long list was reduced to a 25-item checklist. A prototypical flow diagram was developed, which provides information about the method of patient recruitment, the order of test execution and the numbers of patients undergoing the test under evaluation, the reference standard or both.

Conclusions - Evaluation of research depends on transparent reporting. The STARD group anticipates that the use of the checklist in combination with the flow diagram will enhance the quality of reporting of studies on diagnostic accuracy.

INTRODUCTION

The world of diagnostic tests is highly dynamic. New tests are developed at a fast rate and the technology behind existing tests is continuously improved. Exaggerated and biased results from poorly designed and reported diagnostic studies can trigger premature dissemination and mislead physicians to incorrect treatment decisions. A rigorous evaluation process of diagnostic tests before introduction into clinical practice could not only reduce the number of unwanted clinical consequences related to false test results but also limit health care costs by preventing unnecessary testing. Studies to determine the diagnostic accuracy of a test are a vital part in this evaluation process.¹⁻³

In studies on diagnostic accuracy, the information from one or more tests under evaluation is compared with information from the reference standard as measured in the same series of subjects suspected of the condition of interest. The word test refers to any method for obtaining additional information on a patient's health status. It includes information from laboratory tests, imaging tests, function tests, pathology, history and physical examination. The condition of interest or target condition can refer to a particular disease or to any other identifiable condition within a patient that may prompt clinical action, like the initiation, modification or termination of treatment. In this framework, the reference standard is considered to be the best available method for establishing the presence or absence of the condition of interest. The reference standard can be one or even a combination of methods to establish the presence of the target condition, including laboratory tests, imaging tests, pathology, but also clinical follow-up of subjects. The term accuracy refers to the amount of agreement between the information from the test under evaluation and the reference standard. Diagnostic accuracy can be expressed in many ways, including sensitivity and specificity, likelihood ratios, diagnostic odds ratio, and the area under a ROC curve.⁴⁻⁶

There are several potential threats to the internal and external validity of a study on diagnostic accuracy. A survey of studies on diagnostic accuracy published in four major medical journals between 1978 and 1993 revealed that the methodological quality was mediocre at best.⁷ Furthermore, it showed that information on key elements of design, conduct and analysis of diagnostic studies was often not reported.⁷ The absence of critical information about the design and conduct of diagnostic studies has been confirmed by authors of meta-analyses.^{8,9} Like any other

type of research, flaws in study design can lead to biased results. A report has shown that diagnostic studies with specific design features are associated with biased, optimistic, estimates of diagnostic accuracy compared to studies without such deficiencies.¹⁰

At the 1999 Cochrane Colloquium meeting in Rome, the Cochrane Diagnostic and Screening Test Methods Working Group discussed the low methodological quality and sub-standard reporting of diagnostic test evaluations. Urged by the findings that studies of poor quality overestimated diagnostic accuracy, the Group felt that the first step to correct these problems was to improve the quality of reporting of diagnostic studies. Following the successful CONSORT initiative¹¹⁻¹³, the Group aimed at the development of a comparable checklist of items that should be included in a report of study on diagnostic accuracy.

The objective of the Standards for Reporting of Diagnostic Accuracy (STARD) initiative is to improve the quality of reporting of studies on diagnostic accuracy. Complete and accurate reporting allows the reader to detect the potential for bias in the study (internal validity) and to judge the generalisability and applicability of the results (external validity).

METHODS

The STARD steering committee (see appendix for membership and details) started with an extensive search to identify publications on the conduct and reporting of diagnostic studies. This search included the MEDLINE, EMBASE, and the Cochrane Research and Methods database. In addition, the steering committee members examined reference lists of retrieved articles, searched personal files, and contacted other experts in field of diagnostic research. All relevant publications were reviewed and extracted into a long list of potential checklist items.

Subsequently, the STARD steering committee prepared a two-day consensus meeting for invited experts from the following interest groups: researchers, editors, methodologists and professional organisations. The aim of the conference was to reduce the long list of potential items, where appropriate, and to discuss the optimal format and phrasing of the checklist.

The meeting format consisted of a mixture of small group sessions and plenary sessions. Each small group focused on a group of related items of the long list. The suggestions of the small groups were then discussed in plenary sessions. Overnight a first draft of the STARD checklist was assembled based on the suggestions from the small group and the additional remarks from the plenary sessions. This version was then discussed the next day among all attendees of the consensus meeting, and additional changes were made. There was a final round of comments from the members of the STARD working group by electronic mail.

The conference version was then field-tested by potential users and additional comments were collected. This version was made available on the Internet at the CONSORT Website (www.consort-statement.org), together with a call for comments. All comments were discussed among the STARD group to arrive at the checklist presented in this paper.

RESULTS

The search on published guidelines for diagnostic research yielded 33 lists. Based on these published guidelines and from input of steering and working group members, a long list of 75 items was assembled. During the consensus meeting on September 16 and 17, 2000, the long list was reduced into a 25-item checklist. Major revisions were made with respect to the phrasing and format of the checklist during the conference. The STARD group received several valuable comments and remarks during the various stages of evaluation after the conference. This resulted in the version of the STARD checklist as presented in table 1.

Table 1. STARD checklist of items to improve the reporting of studies on diagnostic accuracy. *Test version, November 2001. For evaluation purposes only.*

Section and Topic	Item #	Describe	Reported on page #
TITLE/ABSTRACT/KEYWORDS	1	The article as a study on diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity')	
INTRODUCTION	2	The research question(s) such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups	
METHODS			
<i>Participants</i>	3	The study population: the inclusion and exclusion criteria, setting(s) and location(s) where the data were collected	
	4	Participant recruitment: was this based on presenting symptoms, results from previous tests, or the fact that the participants had received the index test(s) or the reference standard?	
	5	Participant sampling: was this a consecutive series of patients defined by selection criteria in (3) and (4)? If not specify how patients were further selected.	
	6	Data collection: were the participants identified and data collected before the index test(s) and reference standards were performed (prospective study) or after (retrospective study)?	
<i>Reference standard</i>	7	The reference standard and its rationale	
<i>Test methods</i>	8	Technical specification of material and methods involved including how and when measurements were taken, and/or cite references for index test(s) and reference standard	
	9	Definition and rationale for the units, cutoffs and/or categories of the results of the index test(s) and the reference standard	
	10	The number, training and expertise of the persons (a) executing and (b) reading the index test(s) and the reference standard	
	11	Whether or not the reader(s) of the index test(s) and reference standard were blind (masked) to the results of the other test(s) and describe any information available to them	
<i>Statistical methods</i>	12	Methods for calculating measures of diagnostic accuracy or making comparisons, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals)	
	13	Methods for calculating test reproducibility, if done	
RESULTS			
<i>Participants</i>	14	When study was done, including beginning and ending dates of recruitment	
	15	Clinical and demographic characteristics (e.g. age, sex, spectrum of presenting symptom(s), comorbidity, current treatment(s), recruitment center)	
	16	How many participants satisfying the criteria for inclusion did or did not undergo the index test and/or the reference standard; describe why participants failed to receive either test (a flow diagram is strongly recommended)	
<i>Reference standard</i>	17	Time interval and any treatment administered between index and reference standard	
	18	Distribution of severity of disease (define criteria) in those with the target condition; describe other diagnoses in participants without the target condition	
<i>Test results</i>	19	A cross tabulation of the results of the index test(s) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard	
	20	Indeterminate results, missing responses and outliers of index test(s) stratified by reference standard result and how they were handled	
	21	Adverse events of index test(s) and reference standard	
<i>Estimation</i>	22	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals)	
	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done	
	24	Measures of test reproducibility, if done	
DISCUSSION	25	The clinical applicability of the study findings	

The flow diagram provides information about the method of patient recruitment (e.g., based on a consecutive series of patients with specific symptoms, case-control), the order of test execution, and the number of patients undergoing the test under evaluation (index test) and the reference test (see figure 1). We provide one prototypical flowchart that reflects the most commonly employed design in diagnostic research.

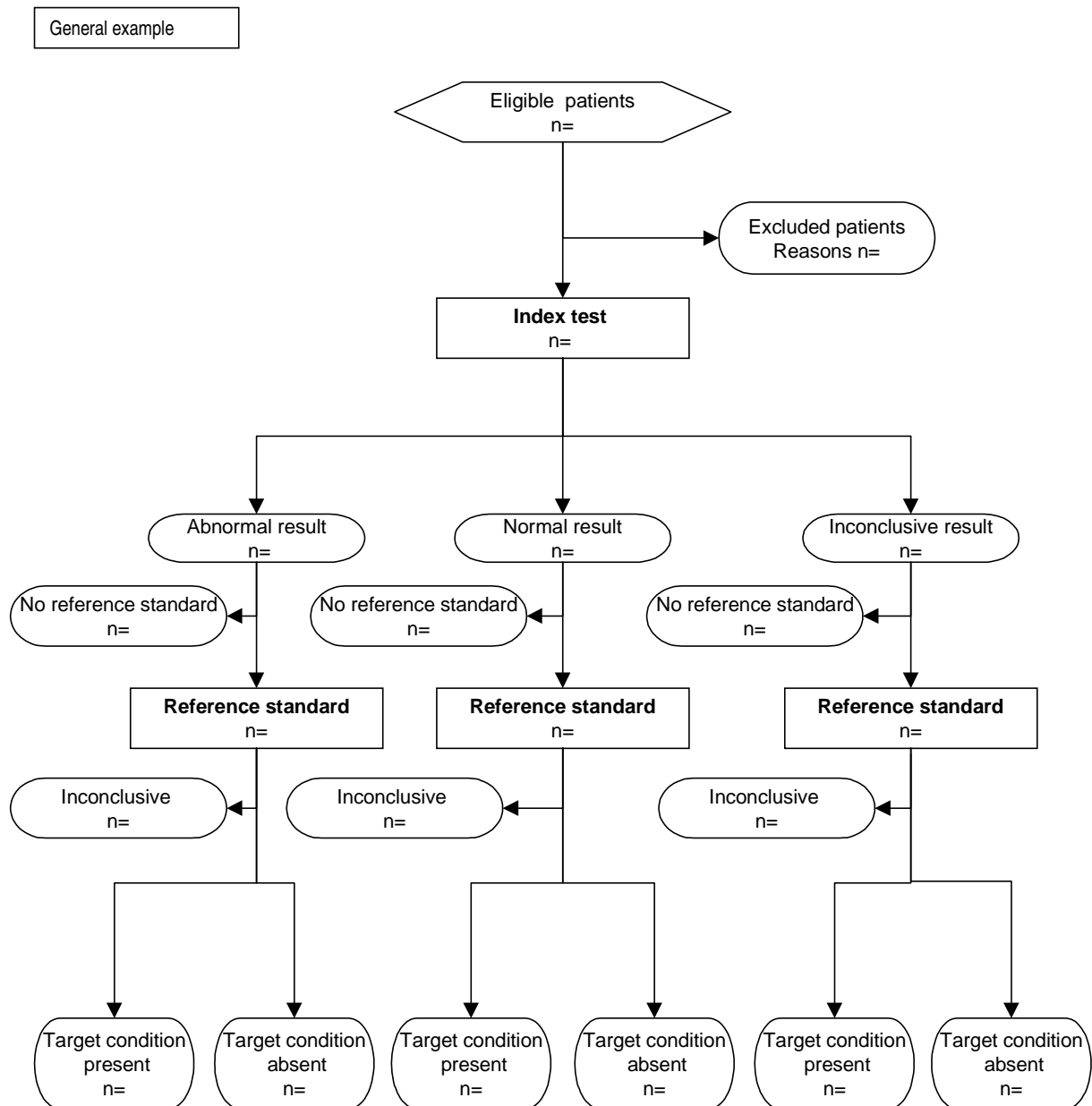


Figure 1. Prototypical flow diagram of a study on diagnostic accuracy.

Test version, November 2001, For evaluation purposes only.

DISCUSSION

The purpose of the STARD checklist is to improve the quality of the reporting of diagnostic studies. The choice of items in the STARD checklist, including the flowchart, is to assure that information on essential elements of the design, the conduct, the execution of tests and results are presented and unambiguously described. We arranged the items under the familiar headings (introduction, methods, results, discussion) of a medical research article. However, the emphasis of STARD is to assure that the various items are all reported, not to dictate the order or place within the article itself. To emphasise this point, we added the column '*reported on page #*' to the checklist.

General considerations in the development of the STARD checklist

The guiding principle in the development of the STARD checklist was the selection of items that would help readers to judge the potential for bias in the study and to appraise the applicability of the findings. Two other general considerations shaped the content and format of the checklist. First, the STARD group believed that one general checklist for studies of diagnostic accuracy, rather than different checklists for each field, would have a better chance of being widely accepted. Although the evaluation of an imaging test differs from that of a lab test, the STARD group felt that these differences were more in degree than in kind. The second consideration was to develop a checklist specifically aimed at studies of diagnostic accuracy. This meant that general issues in the reporting of research findings, like the recommendations contained in the Uniform Requirements for Manuscripts submitted to Biomedical Journals¹⁴, were excluded from the STARD checklist.

Rationale for inclusion of items and the importance of the background document

The decision to include items was based on evidence that linked these items to biased estimates (internal validity) or to variation in measures of diagnostic accuracy (external validity). The evidence varied from narrative articles explaining theoretical principles to papers presenting results from statistical modelling to empirical evidence derived from real-life diagnostic studies. This heterogeneity in evidence was one of the reasons to prepare a background document. In the background document the meaning and rationale of each item will be explained together with a summary of the type and amount of evidence. The existence of such a background document could enhance the use, understanding and dissemination of the STARD checklist. Therefore, we aim to coincide the publication of the STARD checklist with that of the STARD background document. This approach was also motivated by the experience of the CONSORT group where the explanation and explanatory

document was published at time of the revised CONSORT checklist.¹⁵ The members of the CONSORT group felt that part of the critique on the first checklist could have been prevented by the presence of the explanatory document.

Flow diagram

The STARD group put considerable efforts into the development of flow diagrams for diagnostic studies. A flow diagram has the potential to communicate vital information about the design of a study and the flow of participants in a transparent manner.¹⁶ The flow diagram is an essential element in the CONSORT standards for reporting of randomized trials, but its role could even more vital in diagnostic studies given the variety of designs employed in diagnostic research. Flow diagrams in reports of diagnostic accuracy studies could provide insight into:

- the sampling and selection process of participants (external validity)
- the flow of participants in relation to the timing and outcomes of tests, in particular the number of subjects who fail to receive either the index test and/or the reference standard (potential of verification bias^{17,18})
- the number of patients at each stage of the study which identifies the correct denominator for proportions the number of patients at each stage of the study which identifies the correct denominator for rates and proportions (internal consistency)

In this statement, we provide one archetypal flow diagram general, but more examples, reflecting other designs, can be found at the CONSORT Web site at www.consort-statement.org

Evaluation and revision

We plan to measure the impact of the statement on the quality of published reports on diagnostic accuracy using a before-and-after evaluation.¹³ The checklist will be regularly updated and revised when new evidence comes available or based on comments from users of the checklist. Therefore, we welcome any comment, whether content or form, to improve the current version.

APPENDIX

Members of the STARD steering committee

Patrick Bossuyt

Academic Medical Center, Dep of Clinical Epidemiology,
Amsterdam, The Netherlands

Constantine Gatsonis

Brown University, Centre for Statistical Sciences
Providence, United States of America

Les Irwig

University of Sydney, Dep of Public Health &
Community Medicine, Sydney, Australia

David Moher

Thomas C. Chalmers Centre for Sys. Reviews
Ottawa, Ontario, Canada

Riekje de Vet

Free University, Institute for Research in Extramural
Medicine, Amsterdam, The Netherlands

David Bruns

Clinical Chemistry
Charlottesville, United States of America

Paul Glasziou

Mayne Medical School, Dep. of Social &
Preventive Medicine, Herston, Australia

Jeroen Lijmer

Academic Medical Center, Dep of Clinical
Epidemiology, Amsterdam, The Netherlands

Drummond Rennie

Journal of the American Medical Association,
Jacksonville, United States of America

Members of the STARD working group

Doug Altman

Institute of Health Sciences, Centre for Statistics in
Medicine, Oxford, United Kingdom

Colin Begg

Memorial Sloan-Kettering Cancer Center, Dep
Epidemiology & Biostatistics, New York, United States
of America

Harry Büller

Academic Medical Center, Dep of Vascular
Medicine, Amsterdam, The Netherlands

Frank Davidoff

Annals of Internal Medicine
Philadelphia, United States of America

Paul Dieppe

Dept Social Medicine
University of Bristol, Bristol, United Kingdom

Rijk van Ginkel

Academic Medical Center, Dep of Clinical
Epidemiology, Amsterdam, The Netherlands

Gordon Guyatt

McMaster University, Clinical Epidemiology and
Biostatistics, Hamilton, Canada

Richard Horton

The Lancet,
London, United Kingdom

Stuart Barton

British Medical Journal,
BMA House, London, United Kingdom

William Black

Dartmouth Hitchcock Medical Center, Dep of
Radiology, United States of America

Gregory Campbell

US FDA, Center for Devices and Radiological Health
Rockville, United States of America

Jon Deeks

Institute of Health Sciences, Centre for Statistics in
Medicine, Old Road, United Kingdom

Kenneth Fleming

John Radcliffe Hospital, Oxford, United Kingdom

Afina Glas

Academic Medical Center, Dep of Clinical
Epidemiology, Amsterdam, The Netherlands

James Hanley

McGill University, Dep Epidemiology & Biostatistics,
Montreal, Canada

Myriam Hunink

Erasmus Medical Center, Dep Epidemiology &
Biostatistics, Rotterdam, The Netherlands

Jos Kleijnen

NHS Centre for Reviews and Dissemination
York, United Kingdom

Erik Magid

Amager Hospital, Dep Clinical Biochemistry
Copenhagen, Denmark

Matthew McQueen

Hamilton Civic Hospitals, Dep of Laboratory Medicine
Hamilton, Canada

John Overbeke

Nederlands Tijdschrift voor Geneeskunde,
Amsterdam, The Netherlands

Anthony Proto

Radiology Editorial Office, Richmond, United States
of America

David Sackett

Trout centre,
Ontario, Canada

Harold Sox

Dartmouth Hitchcock Medical Center
Dep of Medicine, Lebanon, United States of America

Stephan Walter

McMaster University, Clin Epidemiology and
Biostatistics, Hamilton, Canada

Andre Knottnerus

Maastricht University, Netherlands School of Primary
Care Research, Maastricht, The Netherlands

Barbara McNeil

Harvard Medical School, Dep of Health Care Policy,
Boston, United States of America

Andrew Onderdonk

Channing Laboratory, Boston, United States of
America

Christopher Price

St Bartholemew' s - Royal London School of Medicine
and Dentistry, London, United Kingdom

Hans Reitsma

Academic Medical Center, Dep of Clinical
Epidemiology, Amsterdam, The Netherlands

Gerard Sanders

Academic Medical Center, Dep of Clinical Chemistry
Amsterdam, The Netherlands

Sharon Straus

Mt. Sinai Hospital, Toronto, Canada

REFERENCES

1. Guyatt GH, Tugwell PX, Feeny DH, Haynes RB, Drummond M. A framework for clinical evaluation of diagnostic technologies. *Can Med Assoc J* 1986;134:587-594.
2. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991;11:88-94.
3. Kent DL, Larson EB. Disease, level of impact, and quality of research methods. Three dimensions of clinical efficacy assessment applied to magnetic resonance imaging. *Invest Radiol* 1992;27:245-254.
4. Griner PF, Mayewski RJ, Mushlin AI, Greenland P. Selection and interpretation of diagnostic tests and procedures. Principles and applications. *Ann Intern Med* 1981;94:557-592.
5. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. The selection of diagnostic tests. In: Sackett D, editor. *Clinical Epidemiology*. 2nd ed. Boston/Toronto/London: Little, Brown and Company; 1991. p. 47-57.
6. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* 1978;8:283-298.
7. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA* 1995;274:645-651.
8. Nelemans PJ, Leiner T, de Vet HCW, van Engelshoven JMA. Peripheral arterial disease: Meta-analysis of the diagnostic performance of MR angiography. *Radiology* 2000;217:105-114.
9. Devries SO, Hunink MGM, Polak JF. Summary receiver operating characteristic curves as a technique for meta-analysis of the diagnostic performance of duplex ultrasonography in peripheral arterial disease. *Acad Radiol* 1996;3:361-369.
10. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061-6.
11. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276:637-9.
12. Moher D, Schulz KF, Altman D. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001;285:1987-1991.
13. Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials. A comparative before-and-after evaluation. *JAMA* 2001;285:1992-1995.
14. International Committee of Medical Journal Editors. Uniform Requirements for manuscripts submitted to biomedical journals. *JAMA*. 1997;277:927-934. Available at: ACP Online, <http://www.acponline.org>.
15. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gøtzsche PC, Lang T. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Ann Intern Med* 2001;134:663-694.
16. Egger M, Jüni, Barlett C. Value of flow diagrams in reports of randomized controlled trials. *JAMA* 2001;285:1996-1999.
17. Panzer RJ, Suchman AL, Griner PF. Workup bias in prediction research. *Med Decis Making* 1987;7:115-119.
18. Begg CB. Biases in the assessment of diagnostic tests. *Stat Med* 1987;6:411-423.

**L'ESPOSIZIONE COMPLETA E RIGOROSA
DEGLI STUDI DI ACCURATEZZA DIAGNOSTICA:
L'INIZIATIVA STARD**

The STARD group

Recapiti per commenti e suggerimenti :

E-mail: stard@amc.uva.nl

Phone: +31-20-5666694

Fax: +31-20-6912683

Titolo originale:

Towards complete and accurate reporting of studies on diagnostic accuracy: the STARD initiative.

Traduzione a cura di Pino Giocoli del Gruppo di lavoro EBM, Associazione Microbiologi Clinici Italiani,

Il lettore può inviare domande, commenti e suggerimenti in italiano all'indirizzo:

gioco.en@numerica.it

RIASSUNTO

Obiettivi. L'obiettivo dell'iniziativa “*Standards for Reporting Diagnostic Accuracy*” (STARD) è di migliorare l'esposizione degli studi di accuratezza diagnostica con l'ausilio di una *checklist* e di un diagramma di flusso. Una descrizione completa e accurata permette al lettore di individuare potenziali *bias* e di valutare se i risultati dello studio sono generalizzabili.

Metodi. Il gruppo direttivo dello STARD ha condotto un'approfondita ricerca in letteratura, allo scopo di individuare pubblicazioni sul modo di eseguire studi sui test diagnostici e riferirne i risultati. Un lungo elenco di voci di potenziale interesse è stato ridimensionato durante una riunione di consenso di due giorni, cui hanno partecipato ricercatori, editori e membri di organizzazioni professionali. Un'attenzione speciale è stata dedicata alla preparazione di un diagramma di flusso, ad uso generale per gli studi di accuratezza diagnostica.

Risultati. L'esame di linee guida per la ricerca diagnostica ha fornito 33 elenchi di argomenti, riuniti poi in una lunga lista di 75 voci. Da essa, durante una riunione di consenso, è stata preparata una *checklist* finale di 25 punti chiave. E' stato delineato anche un prototipo di diagramma di flusso che fornisce informazioni sul metodo di reclutamento, l'ordine delle determinazioni e il numero di pazienti sottoposti al test in valutazione, allo standard di riferimento o ad entrambi.

Conclusioni. La valutazione di una ricerca è condizionata dalla chiara esposizione della sua struttura e dei suoi risultati. Il gruppo STARD confida che la *checklist* e il diagramma di flusso migliorino la qualità dei rapporti sugli studi di accuratezza diagnostica.

INTRODUZIONE

Il mondo dei test diagnostici è in continua evoluzione. Nuovi test sono inventati a ritmo serrato e la tecnologia di quelli già in uso viene perfezionata. Malgrado ciò, i clinici possono essere ingannati da risultati affetti da vari gradi di inaccuratezza – a causa di metodologie male impostate o descritte impropriamente - e in conseguenza possono prendere iniziative terapeutiche errate. Una valutazione rigorosa dei test diagnostici prima della loro introduzione nella pratica clinica potrebbe ridurre gli effetti dannosi di risultati sbagliati e i costi sanitari, evitando indagini non necessarie. Gli studi che servono a definire l'accuratezza diagnostica di un test costituiscono una parte vitale di tale processo di valutazione (1-3).

Negli studi di accuratezza diagnostica, i risultati del test da valutare sono paragonati a quelli ottenuti con un metodo di riferimento. A tale scopo si eseguono determinazioni su una medesima serie di persone sospettate di trovarsi in una certa condizione che s'intende rivelare. La parola **test** è riferibile a qualsiasi metodo per ottenere informazioni aggiuntive sullo stato di salute di un soggetto. Include cioè informazioni provenienti da prove di laboratorio, *imaging*, prove funzionali, anatomo patologiche, nonché dalla storia e dall'esame obiettivo. La condizione che interessa, **il target dell'indagine**, può essere propria di una particolare malattia o di una qualsiasi altra situazione che richieda un intervento clinico, ad esempio l'inizio, la modifica o la fine di una terapia.

Sotto questo punto di vista, lo **standard di riferimento** è considerato il miglior metodo disponibile per stabilire la presenza o l'assenza del *target* dell'indagine. Può essere dunque essere un metodo, ma anche una combinazione di metodi, inclusi test di laboratorio, indagini radiologiche, anatomo patologiche, o il monitoraggio dei soggetti nel tempo (*follow-up*). Il termine **accuratezza** si riferisce alla concordanza (espressa quantitativamente) tra l'informazione ottenuta dal test in studio e lo standard di riferimento. L'accuratezza diagnostica può essere espressa in molti modi: sensibilità e specificità, *likelihood ratio*, *odds ratio* diagnostiche o come area sottostante una curva ROC (4-6).

Esistono parecchie potenziali minacce alla validità interna ed esterna di uno studio di accuratezza diagnostica. Una rassegna di questo tipo di studi - pubblicati tra il 1978 e il 1993 in quattro delle più importanti riviste mediche - ha svelato (nel migliore dei casi) una qualità mediocre (7). Inoltre ha dimostrato che spesso mancavano le informazioni su elementi chiave del disegno, dell'esecuzione e dell'interpretazione dei test (7). La carenza di informazioni critiche è stata confermata da alcune metanalisi (8,9). Come per qualsiasi altro tipo di ricerca, gli errori nell'architettura di uno studio possono condurre a risultati affetti da *bias*. Un rapporto di studiosi olandesi ha chiarito che studi diagnostici con certe caratteristiche d'impostazione sono associati a stime viziate e troppo ottimistiche dell'accuratezza diagnostica, in paragone a studi che non posseggono quei difetti (10).

Al *Cochrane Colloquium* tenutosi a Roma nel 1999, il *Cochrane Diagnostic and Screening Test Methods Working Group* ha discusso sul basso livello metodologico degli studi diagnostici e sulla qualità sub-ottimale dei rapporti che servivano a descriverli. Mosso dalla constatazione che studi di cattiva qualità sopravvalutano l'accuratezza diagnostica, il Gruppo ha ritenuto che il primo passo fosse migliorarne la qualità espositiva. In seguito al successo dell'iniziativa CONSORT (11-13), si è deciso di seguirne l'esempio e di sviluppare in modo analogo una *checklist* di punti essenziali, da utilizzare per i rapporti sull'accuratezza diagnostica di un test.

L'obiettivo dell'iniziativa *Standards for Reporting Diagnostic Accuracy* (STARD) è dunque di migliorare la qualità espositiva degli studi di accuratezza diagnostica. Un rapporto completo e accurato permette al lettore di individuare i *bias* potenziali dello studio (**validità interna**) e di esprimere un giudizio sulla possibilità di generalizzarne e applicarne i risultati (**validità esterna**).

METODI

Il gruppo direttivo dello STARD (v. *appendice con il nome dei membri e altri dettagli*) ha condotto una vasta ricerca delle pubblicazioni sul modo di eseguire studi sui test diagnostici ed esporne i risultati. Sono stati utilizzati MEDLINE, EMBASE e il *Cochrane Research and Methods database*. Inoltre, i membri del gruppo hanno esaminato la bibliografia degli articoli, i loro archivi personali e hanno contattato altri esperti nel campo. Dopo una revisione delle pubblicazioni più rilevanti è stato compilato un lungo elenco di voci da includere eventualmente nella *checklist*.

In seguito è stata indetta una riunione di consenso di due giorni, alla quale sono stati invitati ricercatori, editori, esperti di metodologia e membri di organizzazioni professionali. L'intento del *meeting* era di ridurre il lungo elenco dei potenziali elementi della *checklist* e di discutere sul modo migliore di organizzarla e di presentarla.

La conferenza ha compreso sessioni ristrette e sessioni plenarie. I membri di ogni sessione ristretta si sono concentrati nell'esame di un gruppo di voci affini del lungo elenco e le loro osservazioni sono state poi discusse nelle sessioni plenarie. Alla fine del primo giorno era già disponibile una prima stesura della *checklist*. Nel secondo giorno questa versione è stata ancora discussa e modificata da tutti i partecipanti del *consensus meeting*. E' seguito un giro finale di commenti formulati per posta elettronica da tutti membri del gruppo STARD. La versione del documento espressa dalla conferenza è stata quindi sottoposta ad una prova sul campo dai potenziali utilizzatori, che hanno aggiunto altri commenti. Resa disponibile in Internet al CONSORT website (www.consort-statement.org) ha raccolto altre osservazioni. Dopo una discussione finale in seno allo STARD Group, si è giunti alla *checklist* presentata in questo articolo.

RISULTATI

L'esame delle esistenti linee guida per gli studi diagnostici ha fruttato 33 elenchi di argomenti. Sulla base di questo materiale e dei suggerimenti del gruppo direttivo e di tutto il gruppo di lavoro STARD è stata preparata una lunga lista di 75 voci inseribili nella *checklist*. Nella riunione di consenso del 16 e 17 settembre 2000, il gruppo di lavoro ha ridotto il numero dei punti chiave a 25 e introdotto importanti modifiche di struttura e di linguaggio. Dopo i commenti e le critiche nei successivi stadi della valutazione, il risultato è la versione della ***checklist*** presentata nella **tabella 1**.

Il **diagramma di flusso** dà informazioni sui metodi di reclutamento dei pazienti (es. se in serie consecutiva di persone con specifici sintomi, oppure con modalità caso-controllo), sull'ordine di esecuzione delle determinazioni, sul numero di pazienti sottoposti sia al test in studio che al test di riferimento

(v. **figura 1**). Quello mostrato è un prototipo di diagramma di flusso che riguarda il disegno più comunemente adoperato nella ricerca diagnostica.

DISCUSSIONE

Lo scopo della **checklist STARD** è di migliorare la qualità dell'esposizione degli studi di accuratezza diagnostica. La scelta dei suoi elementi e di quelli del diagramma di flusso mira ad assicurare che non vi siano ambiguità nel modo di presentare e di descrivere i punti chiave del disegno di uno studio, la sua esecuzione e i suoi risultati. Abbiamo suddiviso i punti della *checklist* nelle classiche **sezioni di un articolo** di ricerca medica (Introduzione, Metodi, Risultati, Discussione). Il gruppo STARD sottolinea che tutti i punti della lista devono essere considerati, tuttavia non intende indicarne l'ordine né il posto che essi devono occupare nell'articolo. E' per questo che nella *checklist* abbiamo inserito una colonna che indica il **numero della pagina** cui essi si riferiscono.

Considerazioni generali sullo sviluppo della checklist STARD.

Il criterio informatore della *checklist* STARD è stato la scelta di elementi in grado di far da guida al lettore in un giudizio sui potenziali *bias* dello studio e sulla pratica applicabilità dei risultati. Altre due considerazioni generali hanno aiutato a modellarne la forma e il contenuto. Primo: si è ritenuto che una *checklist* unica per gli studi di accuratezza in tutti i campi della diagnostica fosse più gradita di una *checklist* per ciascuno di essi. Infatti, benchè la valutazione di un test di *imaging* differisca da quella di un test di laboratorio, le differenze sono ritenute più di quantità che di qualità. Secondo: questa è una *checklist* specifica, mirata agli studi di accuratezza diagnostica. Ciò significa che le istruzioni generali sul modo di descrivere una ricerca – ad esempio gli *Uniform Requirements for Manuscripts submitted to Biomedical Journals* (14) – non riguardano la nostra *checklist*.

Punti chiave della lista: la loro logica e l'importanza di un documento esplicativo

La decisione di includere certi elementi nella *checklist* è basata sulle loro evidenti relazioni con i vizi di valutazione (*bias*) (validità interna) o con le variazioni nella stima dell'accuratezza diagnostica (validità esterna). L'evidenza rilevata è di livello diverso, secondo la qualità degli articoli - di tipo narrativo, con spiegazioni teoriche, o di elaborazione statistica (*statistical modelling*¹), oppure riportanti dati empirici ricavati da studi in condizioni reali. Questa eterogeneità spiega la necessità di un **documento esplicativo**. In esso il significato e la logica di ciascun elemento saranno commentati e sarà fatta una sintesi del tipo e della quantità dell'evidenza. Un documento di tal

genere è in grado di favorire la comprensione, la disseminazione e l'utilizzo della *checklist* STARD. Perciò noi progettiamo di far coincidere la sua pubblicazione definitiva con quella del relativo commento. Così ha agito il gruppo CONSORT, che ha pubblicato il documento esplicativo insieme alla *checklist* revisionata (15). I membri del gruppo CONSORT ritengono che parte delle critiche alla loro prima *checklist* sarebbero state evitate in presenza di opportune spiegazioni.

¹ *Abbiamo interpellato sul significato del termine H. Reitsma, dello STARD Working Group, che ha dato la seguente spiegazione: "Per statistical modelling intendiamo lo studio degli effetti di omesse verifiche di risultati positivi o negativi in una tabella 2x2 sui parametri dell'accuratezza. Le alterazioni delle caratteristiche del test sono confrontate con la situazione di base (Diamond GA. Med Decis Making 1992;12:22-3)" (NdT).*

Il diagramma di flusso

Un diagramma di questo tipo ha richiesto un notevole impegno da parte del gruppo di lavoro. Un **diagramma di flusso** ha la capacità di comunicare informazioni vitali sull'architettura di uno studio e sulla trasparenza del lavoro eseguito (16). Nel progetto CONSORT sui trial randomizzati il diagramma è un elemento chiave, ma il suo ruolo potrebbe essere ancor più vitale negli studi diagnostici, data la varietà dei disegni impiegati in questo campo di ricerca. L'uso dei diagrammi di flusso per le descrizioni degli studi sull'accuratezza diagnostica può fornire elementi su:

- Il campionamento e la selezione dei partecipanti (validità esterna);
- Il flusso dei partecipanti in funzione del tempo di esecuzione e dei risultati dei test, in particolare il numero dei soggetti che non sono sottoposti al test in studio e/o al test di riferimento (potenziale *verification bias* (17,18);
- Il numero di pazienti che partecipa a ciascuno stadio dello studio, che identifica il corretto denominatore per frequenze e proporzioni (validità interna).

In questo documento noi forniamo un prototipo generale di diagramma di flusso ma altri esempi, relativi a diversi disegni, si trovano al CONSORT Web site (www.consort-statement.org).

Valutazione e revisione

Ci riserviamo di valutare l'impatto del documento sulla qualità degli articoli sull'accuratezza diagnostica per mezzo di uno studio prima-dopo (13). La *checklist* sarà regolarmente aggiornata e revisionata in base alle nuove

evidenze disponibili o ai commenti di chi la usa. Perciò diamo il benvenuto a qualsiasi commento sul contenuto e sulla forma della presente versione.

BIBLIOGRAFIA

1. Guyatt GH, Tugwell PX, Feeny DH, Haynes RB, Drummond M. A framework for clinical evaluation of diagnostic technologies. *Can Med Assoc J* 1986;134:587-594.
2. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991;11:88-94.
3. Kent DL, Larson EB. Disease, level of impact, and quality of research methods. Three dimensions of clinical efficacy assessment applied to magnetic resonance imaging. *Invest Radiol* 1992;27:245-254.
4. Griner PF, Mayewski RJ, Mushlin AI, Greenland P. Selection and interpretation of diagnostic tests and procedures. Principles and applications. *Ann Intern Med* 1981;94:557-592.
5. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. The selection of diagnostic tests. In: Sackett D, editor. *Clinical Epidemiology*. 2nd ed. Boston/Toronto/London: Little, Brown and Company; 1991. p. 47-57.
6. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* 1978;8:283-298.
7. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA* 1995;274:645-651.
8. Nelemans PJ, Leiner T, de Vet HCW, van Engelshoven JMA. Peripheral arterial disease: Meta-analysis of the diagnostic performance of MR angiography. *Radiology* 2000;217:105-114.
9. Devries SO, Hunink MGM, Polak JF. Summary receiver operating characteristic curves as a technique for meta-analysis of the diagnostic performance of duplex ultrasonography in peripheral arterial disease. *Acad Radiol* 1996;3:361-369.
10. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061-6.
11. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276:637-9.
12. Moher D, Schulz KF, Altman D. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001;285:1987-1991.

13. Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials. A comparative before-and-after evaluation. JAMA 2001;285:1992-1995.
14. International Committee of Medical Journal Editors. Uniform Requirements for manuscripts submitted to biomedical journals. JAMA. 1997;277:927-934. Available at: ACP Online, <http://www.acponline.org>.
15. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gøtzsche PC, Lang T. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. Ann Intern Med 2001;134:663-694.
16. Egger M, Jüni, Barlett C. Value of flow diagrams in reports of randomized controlled trials. JAMA 2001;285:1996-1999.
17. Panzer RJ, Suchman AL, Griner PF. Workup bias in prediction research. Med Decis Making 1987;7:115-119.
18. Begg CB. Biases in the assessment of diagnostic tests. Stat Med 1987;6:411-423.

Esempio

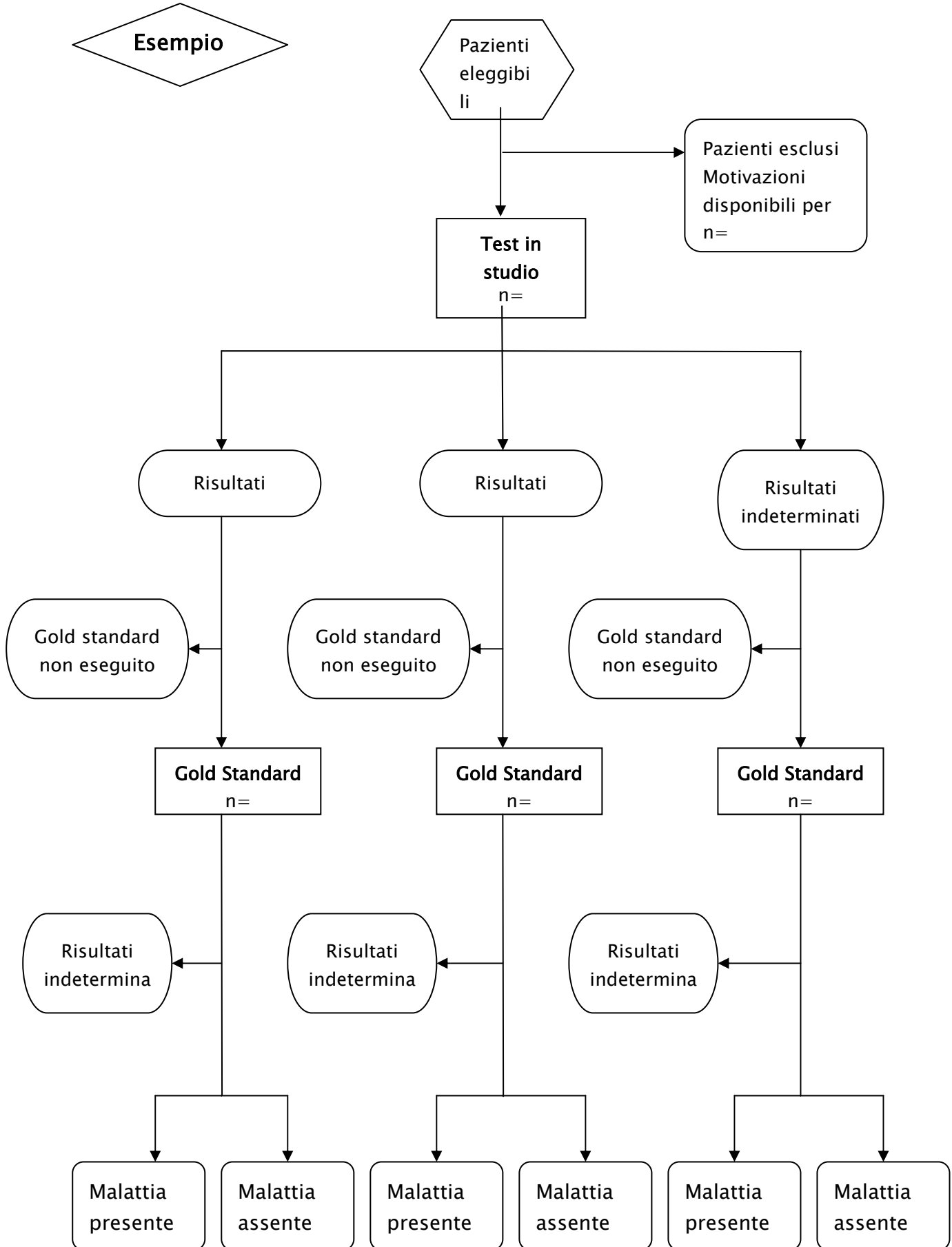


Diagramma di flusso per uno studio sull'accuratezza diaagnostica

Checklist di elementi da verificare per migliorare la presentazione degli studi sull'accuratezza diagnostica

SEZIONE ED ARGOMENTO	N.	DESCRIZIONE	Pag.
TITOLO/RIASSUNTO/ PAROLE CHIAVE	1	Articolo sullo studio dell'accuratezza diagnostica (raccomandato il MeSH heading 'sensitivity and specificity')	
INTRODUZIONE	2	Quesiti che sono oggetto di ricerca, p.es.: <i>stima dell'accuratezza diagnostica o confronto dell'accuratezza tra test o gruppi di partecipanti</i>	
METODI			
Partecipanti	3	La popolazione dello studio: criteri d'inclusione ed esclusione, ambienti e luoghi ove sono stati raccolti i dati	
	4	Reclutamento dei partecipanti: è basato sulla sintomatologia attuale, sui risultati di test già eseguiti, o sul fatto che i partecipanti erano già stati sottoposti al test in studio o a quello di riferimento?	
	5	Campionamento dei partecipanti: è stato eseguito su una serie consecutiva di pazienti definiti con i criteri di selezione (3) e (4)? Se no, specificare come i pazienti sono stati selezionati.	
	6	Raccolta dei dati: l'identificazione dei partecipanti e la raccolta dati sono state effettuate prima che il test in studio e il riferimento fossero eseguiti (studio prospettivo) o dopo (studio retrospettivo)?	
Standard di riferimento	7	Lo standard di riferimento e la sua logica	
Metodologia dei test	8	Specifiche tecniche dei materiali e dei metodi impiegati, inoltre particolari sul come e quando le misure sono state eseguite (e/o citare i riferimenti per il test in studio e il gold standard)	
	9	Definizione e spiegazione delle unità, dei cutoff e/o	

		delle categorie di risultati del test in studio e del test di riferimento	
	10	Il numero, l'addestramento e l'esperienza delle persone (a) che eseguono e (b) interpretano il test in studio e il gold standard	
	11	Precisazioni sul fatto che gl'interpreti del test in studio e del gold standard fossero all'oscuro dei rispettivi risultati ("cecità"); descrizione di qualsiasi informazione loro disponibile	
Metodi statistici	12	Metodi per calcolare i parametri dell' accuratezza diagnostica o per fare paragoni; inoltre i metodi statistici impiegati per quantificare l'incertezza (es. intervalli di confidenza al 95%)	
	13	Metodi per calcolare la riproducibilità, se impiegati	
RISULTATI			
Partecipanti	14	Numero al momento di esecuzione dello studio, inclusi l'inizio e la fine del reclutamento	
	15	Caratteristiche cliniche e demografiche (es. età, sesso, spettro dei sintomi, patologia intercorrente, trattamenti in corso, centro di reclutamento)	
	16	Quanti partecipanti idonei ai criteri d'inclusione sono o non sono stati sottoposti al test in studio e/o al test di riferimento; spiegazioni sul perché alcuni partecipanti sono stati eventualmente esclusi dall'uno o dall'altro (è utile servirsi di un diagramma di flusso)	
Standard di riferimento	17	Intervallo tra esecuzione del test in studio e del test di riferimento; notizie su eventuali trattamenti interposti	
	18	Spettro di gravità della malattia indagata (definirne i criteri) in coloro che ne sono affetti; descrivere altre diagnosi nei partecipanti che invece ne sono esenti	
Risultati	19	Tabella dei risultati del test in studio in confronto a quelli del test di riferimento; per serie continue di risultati (<i>non dicotomici</i>) distribuzione del test in studio in confronto a quelli del test di riferimento	
	20	Risultati indeterminati, risultati mancanti e fuori scala del test in studio stratificati in paragone ai risultati del gold standard; precisare i provvedimenti adottati	
	21	Inconvenienti a carico del test in studio e del test di	

		riferimento	
Interpretazione	22	Stime dell'accuratezza diagnostica e misure dell'incertezza (es. intervalli di confidenza al 95%)	
	23	Stime della variabilità dell'accuratezza diagnostica tra sottogruppi di partecipanti, tra gl' interpreti dei risultati e tra centri di studio (se effettuate)	
	24	Misure della riproducibilità del test (se effettuate)	
DISCUSSIONE			
	25	Applicabilità clinica degli studi eseguiti	

APPENDICE

Membri del gruppo direttivo STARD

Patrick Bossuyt

Academic Medical Center, Dep of Clinical Epidemiology,
Amsterdam, The Netherlands

Constantine Gatsonis

Brown University, Centre for Statistical Sciences
Providence, United States of America

Les Irwig

University of Sydney, Dep of Public Health &
Community Medicine, Sydney, Australia

David Moher

Thomas C. Chalmers Centre for Sys. Reviews
Ottawa, Ontario, Canada

Riekje de Vet

Free University, Institute for Research in Extramural
Medicine, Amsterdam, The Netherlands

David Bruns

Clinical Chemistry
Charlottesville, United States of America

Paul Glasziou

Mayne Medical School, Dep. of Social &
Preventive Medicine, Herston, Australia

Jeroen Lijmer

Academic Medical Center, Dep of Clinical
Epidemiology, Amsterdam, The Netherlands

Drummond Rennie

Journal of the American Medical Association,
Jacksonville, United States of America

Membri del Gruppo di lavoro STARD

Doug Altman

Institute of Health Sciences, Centre for Statistics in
Medicine, Oxford, United Kingdom

Colin Begg

Memorial Sloan-Kettering Cancer Center, Dep
Epidemiology & Biostatistics, New York , United States
of America

Harry Büller

Academic Medical Center, Dep of Vascular
Medicine, Amsterdam, The Netherlands

Frank Davidoff

Annals of Internal Medicine
Philadelphia, United States of America

Paul Dieppe

Dept Social Medicine
University of Bristol, Bristol, United Kingdom

Rijk van Ginkel

Academic Medical Center, Dep of Clinical Epidemiology,
Amsterdam, The Netherlands

Gordon Guyatt

McMaster University, Clinical Epidemiology and
Biostatistics, Hamilton, Canada

Richard Horton

The Lancet,
London, United Kingdom

Stuart Barton

British Medical Journal,
BMA House, London, United Kingdom

William Black

Dartmouth Hitchcock Medical Center, Dep of
Radiology, United States of America

Gregory Campbell

US FDA, Center for Devices and Radiological Health
Rockville, United States of America

Jon Deeks

Institute of Health Sciences, Centre for Statistics in
Medicine, Old Road, United Kingdom

Kenneth Fleming

John Radcliffe Hospital, Oxford, United Kingdom

Afina Glas

Academic Medical Center, Dep of Clinical
Epidemiology, Amsterdam, The Netherlands

James Hanley

McGill University, Dep Epidemiology & Biostatistics,
Montreal, Canada

Myriam Hunink

Erasmus Medical Center, Dep Epidemiology &
Biostatistics, Rotterdam, The Netherlands

Jos Kleijnen

NHS Centre for Reviews and Dissemination
York, United Kingdom

Erik Magid

Amager Hospital, Dep Clinical Biochemistry
Copenhagen, Denmark

Matthew McQueen

Hamilton Civic Hospitals, Dep of Laboratory Medicine
Hamilton, Canada

John Overbeke

Nederlands Tijdschrift voor Geneeskunde,
Amsterdam, The Netherlands

Anthony Proto

Radiology Editorial Office, Richmond, United States
of America

David Sackett

Trout centre,
Ontario, Canada

Harold Sox

Dartmouth Hitchcock Medical Center
Dep of Medicine, Lebanon, United States of America

Stephan Walter

McMaster University, Clin Epidemiology and
Biostatistics, Hamilton, Canada

Andre Knottnerus

Maastricht University, Netherlands School of Primary
Care Research, Maastricht, The Netherlands

Barbara McNeil

Harvard Medical School, Dep of Health Care Policy,
Boston, United States of America

Andrew Onderdonk

Channing Laboratory, Boston, United States of
America

Christopher Price

St Bartholemew's - Royal London School of Medicine
and Dentistry, London, United Kingdom

Hans Reitsma

Academic Medical Center, Dep of Clinical Epidemiology,
Amsterdam, The Netherlands

Gerard Sanders

Academic Medical Center, Dep of Clinical Chemistry
Amsterdam, The Netherlands

Sharon Straus

Mt. Sinai Hospital, Toronto, Canada

APPLICAZIONE PRATICA DELL'APPROCCIO EBM IN LABORATORIO: UN CASO DI ARTRITE REUMATOIDE.

Renzo Gianello Paolo Bartolaminelli** * responsabile laboratorio analisi chimico-cliniche e microbiologiche, Fondazione "S. Maugeri"-Clinica del Lavoro e della Riabilitazione- I.R.C.C.S.- sede di Lumezzane (Bs) **responsabile Reparto Ortopedia- Casa di Cura ex-Poliambulanza "S.Gnutti"-Lumezzane (Bs).*

► DESCRIZIONE DEL CASO

Nel mese di febbraio dello scorso anno (2002) si è presentata all' ambulatorio divisionale della nostra casa di cura la sign.ra M.M., di anni 67, lamentando dolore di intensità variabile durante il giorno a carico delle articolazioni delle mani e dei polsi.

All'esame obiettivo non si riscontravano deformazioni delle estremità, né arrossamento o tumefazione, con ROM (Range Of Motion) nella norma.

Sulla base della evidente simmetria del quadro clinico vengono prescritti gli esami ematochimici di routine (in particolare, gli indici di flogosi), e consigliata una visita reumatologica nel sospetto di una patologia di questo tipo.

In seguito a ciò, la paziente viene sottoposta a terapia anti-infiammatoria e cortisonica per os , benchè siano presenti patologie concomitanti, quali il Diabete Mellito di II° tipo (di modesta entità), e la signora sia HCV positiva.

Viene eseguita, inoltre, una indagine radiologica di polsi e mani che rivela una modesta artrosi inter-falangea con minima deformazione della interlinea articolare.

La paziente gode di relativa buona salute, continuando la terapia indicata, per circa un anno, quando torna alla nostra osservazione per la comparsa di una tumefazione al ginocchio sinistro, accompagnata a dolore intenso ed impotenza funzionale.

Data la presenza di un notevole idrartro, si decide di eseguire una artrocentesi che dà esito a circa 80 cc di liquido torbido sul quale vengono eseguiti: es.chimico-fisico, microscopico, batteriologico e citologico.

Subito dopo si effettua infiltrazione cortisonica con beneficio da parte della paziente.

► DATI DI LABORATORIO

Nel mese di febbraio 2002, come già detto, vengono eseguiti gli esami di routine che rivelano una alterazione degli indici infiammatori, quali: VES 44 mm, FR: 113 U/ml, PCR : 13,7 MG/l.

Su indicazione dello specialista immunologo, nel mese di aprile dello stesso anno la paziente esegue poi: ricerca crioglobuline (con modesta positività), ANA Test (con esito negativo), anti-ENA (negativo), anti-Cardiolipina (negativo), VES (35 mm), PCR (16,8 mg/L) ed APF (risultato positivo).

Sulla base di questi dati e del quadro clinico lo Specialista conclude con una diagnosi di Artrite Reumatoide in fase iniziale e consiglia la terapia con FANS e cortisonici.

Dopo circa un mese viene eseguito un successivo controllo della VES (50 mm), che viene ripetuta nel mese di novembre (41 mm).

L'esame del liquido prelevato con artrocentesi rivela la presenza di un rilevante numero di leucociti (47.800 /mm³), prevalentemente polimorfonucleati (78%), mentre l'esame citologico conferma il reperto flogistico acuto e molti ragociti.

Per quanto riguarda le patologie concomitanti, la glicemia si è sempre mantenuta entro valori accettabili, o solo modestamente aumentati, analogamente alle transaminasi.

CONSIDERAZIONI DIAGNOSTICHE ED APPROCCIO EBM AL CASO CLINICO

Il semplice dato degli esami ematochimici di routine come la VES, il fattore reumatoide e la PCR permettono già, inizialmente, di operare una efficace diagnosi differenziale tra le artropatie sieropositive (di cui poi tratteremo) e quelle sieronegative (come l'osteoartrite e le altre malattie degenerative) oppure la spondilite anchilosante, l'artropatia psoriasica, la sindrome di Reiter e le artropatie entesopatiche, per citare le forme più comuni.

Assai più complessa può essere invece la distinzione nell'ambito delle artriti sieropositive, anche se una anamnesi approfondita, insieme ai risultati radiologici e biochimici, è, di solito, sufficiente per districarsi adeguatamente tra le forme più frequenti (AR, connettiviti sistemiche, polimialgia reumatica, gotta, artriti settiche).

Per questo motivo la paziente è stata inviata dallo specialista Reumatologo che, data la negatività dei tests immunologici, ha ipotizzato una A.R. dell'anziano, consigliando una terapia cortisonica.

Nella fase iniziale della malattia, cioè nel momento in cui la paziente è giunta alla nostra osservazione, una forma di A.R. non poteva che essere considerata possibile, in quanto non ancora pienamente rispondente ai criteri diagnostici stabiliti nel 1987 dall'ARA (American Rheumatism Association) ma tuttora assai validi: infatti, nel nostro caso specifico erano presenti solo: l'artrite simmetrica, l'artrite delle articolazioni interfalangee prossimali e metacarpofalangee, l'artrite delle art. radiocarpiche e, infine, la positività del fattore reumatoide.

Per tutto quanto detto ci è sembrato utile, al fine di dare un significativo contributo alla diagnosi precoce, applicare al caso in oggetto l'approccio basato sull'Evidence Based Medicine (EBM).

A questo proposito abbiamo focalizzato la nostra attenzione su alcuni esami di laboratorio che, nel corso degli anni hanno consolidato la loro posizione nell'ambito della diagnosi di Artrite Reumatoide, e cioè il F.R. (Fattore Reumatoide) e una analisi molto utile benché poco nota tra i clinici: l'A.P.F. (Fattore Perinucleare).

Prima, però, di prendere in considerazione questi due esami, è necessario, al fine di una adeguata applicazione del Nomogramma di Fagan, stabilire, nel caso specifico, la probabilità pre- test cioè la prevalenza della malattia in questione.

Per fare questo è di fondamentale importanza attingere quanto più possibile dalla letteratura finora pubblicata sull'argomento.

Molto sinteticamente, l'Artrite Reumatoide è una malattia cronica diffusa in tutto il mondo con una prevalenza, nella popolazione generale, intorno al 2%; inoltre, dato molto importante nel nostro caso, predilige il sesso femminile con un rapporto 4:1 ed esordisce più frequentemente dopo i 40 anni di età.

D'altro canto, è pur vero che la forma di A.R. nell'anziano (> 60 anni) presenta una prognosi assai migliore rispetto alla forma classica, essendo caratterizzata da scarsa erosività, frequenti remissioni spontanee ed una ottima risposta agli steroidi.

Sulla base di questi dati e sul fatto che la nostra paziente presenta almeno tre criteri diagnostici stabiliti dall'American Rheumatism Association (A.R.A.), è possibile indicare una probabilità pre- test sufficientemente elevata, ragionevolmente attorno al 20-25%.

Detto questo passiamo ad analizzare in modo più approfondito gli esami di laboratorio che sono stati scelti per la diagnosi precoce e come essi possono essere utilizzati nell'ottica EBM:

A) F.R. (Fattore Reumatoide)

Il Fattore Reumatoide rappresenta il primo autoanticorpo identificato nell'A.R.; per la precisione, si dovrebbe parlare di Fattori Reumatoidi in quanto si tratta di anticorpi anti-gammaglobuline di diverso tipo: infatti, essi possono essere IG-M (FR classico) oppure IG-G o IG-A.

In genere, il meccanismo patogenetico con cui tali immunoglobuline partecipano alle lesioni tipiche della malattia è dato dall'attivazione complementare che innesca importanti reazioni flogistiche a livello articolare(1).

Da quanto riportato dai testi di riferimento più comunemente utilizzati (1) (2), i F.R. si riscontrano in circa il 70-80% (1, 2) dei casi di A.R.; pertanto, tali fattori presentano una sensibilità relativamente bassa per la malattia in esame e, inoltre, possono essere occasionalmente isolati in molte altre patologie (quali: connettivite mista, LES, TBC, LUE, infezioni virali, leucemie, linfomi, ecc.), come anche in soggetti sani nel 2-10% dei casi (3), e soprattutto, anziani nel 15-20% dei casi (4).

Di conseguenza non è opportuno richiedere tale esame di laboratorio a "tappeto", ma solo in soggetti opportunamente selezionati.

Una analisi più approfondita della letteratura più recente, condotta presumibilmente nel rispetto di una più rigorosa selezione della casistica, sembra però offrire un quadro assai più variegato dei dati. Questo ci ha spinto ad operare delle scelte degli studi pubblicati ed a considerare "criticamente" gli stessi studi in base al tipo di evidenza suggerito dalla metodologia EBM.

I risultati di questa revisione critica sono schematizzati come segue:

ESAME:

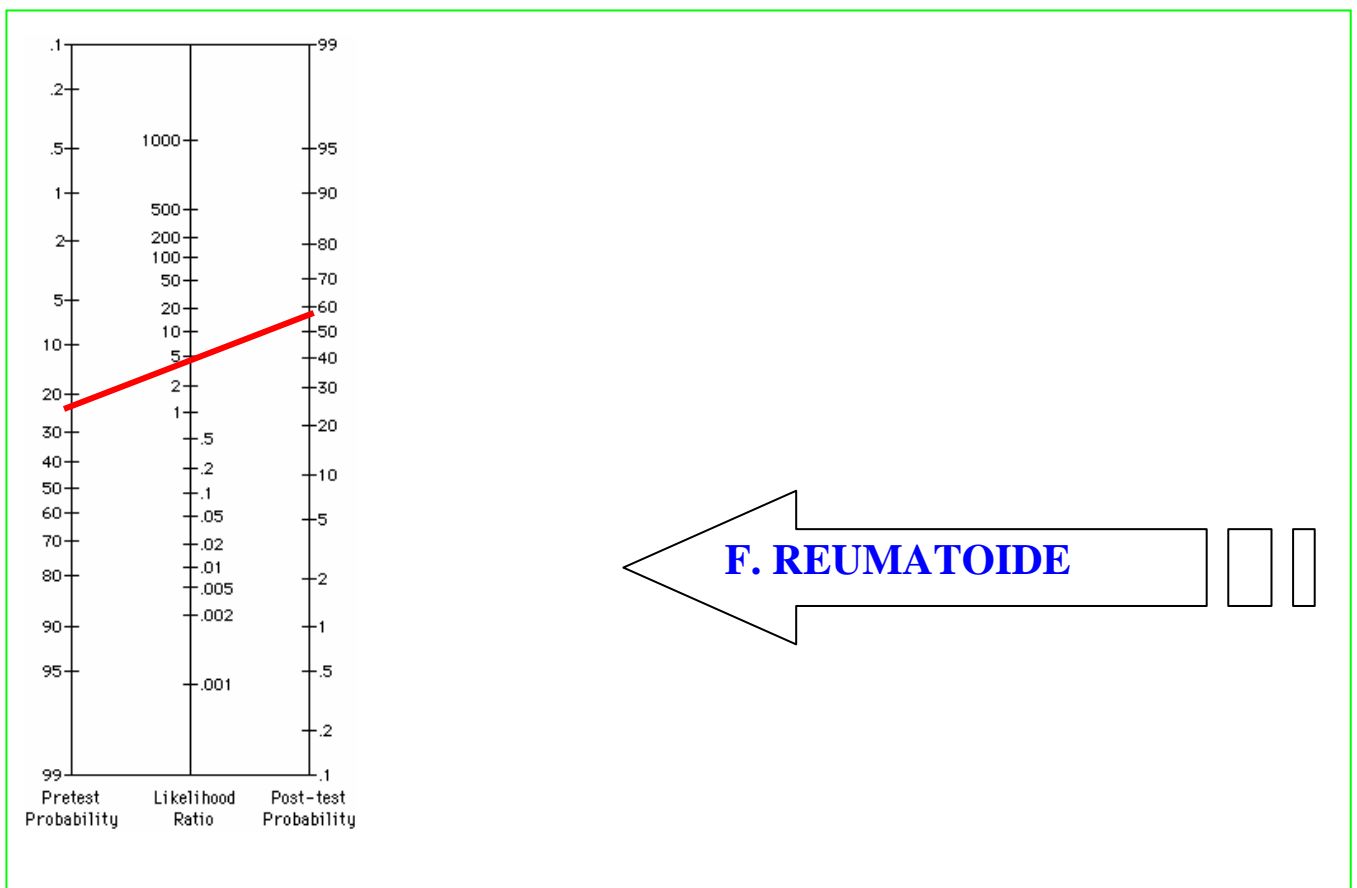
F.R. (Fattore Reumatoide)

LIVELLO o Tipo di evidenza dello studio	SENSIBI- LITA' (%)	SPECIFICI- TA' (%)	LR +	LR -
VON ESSEN R. et al. – (1993) (5) LIVELLO III	72	86	5,1	0,3
LI H. et al. – (2000) (6) LIVELLO II	44,5	72,3	1,6	0,8
SCHELLENKEN S.G.A. et al. – (2000) (7) LIVELLO II	54	91	6,0	0,5
BIZZARO N. et al. – (2001) (8) LIVELLO III	62	84	3,9	0,5
GOLDBACH – HANSKY R. et al. – (2000) (9) LIVELLO III	66	87	5,0	0,4
BAS S. et al. – (2002) (10) LIVELLO II	75	74	3,0	0,3

Considerando questi dati complessivamente, e cercando di dare un “peso” maggiore ai risultati evidenziati dagli studi di livello II°, presumibilmente di maggior valore clinico, possiamo concludere che il Fattore Reumatoide presenta una sensibilità piuttosto bassa (61% circa), una specificità buona (81% circa), un valore di LR + modesto (3,8% circa) ed un altrettanto modesto valore di LR – (0,5% circa).

Tutto ciò indica che l’esame diagnostico in oggetto è utile come test di primo impatto, considerato, anche, il suo relativo basso costo, in pazienti che presentano sintomi sospetti di A.R.(o patologie correlate).

Per quanto riguarda l’applicazione pratica nel caso da noi preso in esame, l’utilizzo del Nomogramma di Fagan permette di ottenere una probabilità post-test di circa il 58%; pertanto non particolarmente significativa.



D’altro canto non avremmo potuto aspettarci un risultato migliore trattandosi di un test, date le caratteristiche appena descritte, con basso potere discriminativo, sia come capacità di determinazione della malattia che come capacità di esclusione della stessa. Di conseguenza è stato necessario individuare un test di secondo livello che avesse delle potenzialità migliori.

Il test che analizzeremo adesso sembra, perciò, avere tali caratteristiche

B) A.P.F. (Fattore Perinucleare o Anti Perinuclear Factor):

l'A.P.F. fa parte di un gruppo di anticorpi, prevalentemente IGG, diretti contro sostanze di origine proteica, o peptidi, denominati Pro-FILAGGRINA che sono presenti in granuli sferici cherato-jalini del citoplasma delle cellule epiteliali.

Oltre all'A.P.F., allo stesso gruppo appartengono anche gli autoanticorpi detti AKA (anti-cheratina), AFA (anti filaggrina), Anti-CCP (anti-citrullina) e ne sono in studio molti altri, tanto che è stato coniato il termine: autoanticorpi anti –sistema della Filaggrina (Citokeratine Filament Aggregating Protein) (1).

Tali anticorpi sono presenti nel 3% dei soggetti sani (14) ed in altre artropatie (es. Artrite Psoriasica) (15).

Questi autoanticorpi sono all'esame degli studiosi da più di 10 anni e, pertanto, analogamente al Fattore Reumatoide, siamo stati indotti ad una analisi la più approfondita possibile della letteratura finora pubblicata, tenendo sempre presente l'approccio EBM riguardo il tipo di evidenza ed il conseguente livello degli studi pubblicati.

Ciò ha prodotto i seguenti dati che poi vedremo di esaminare criticamente:

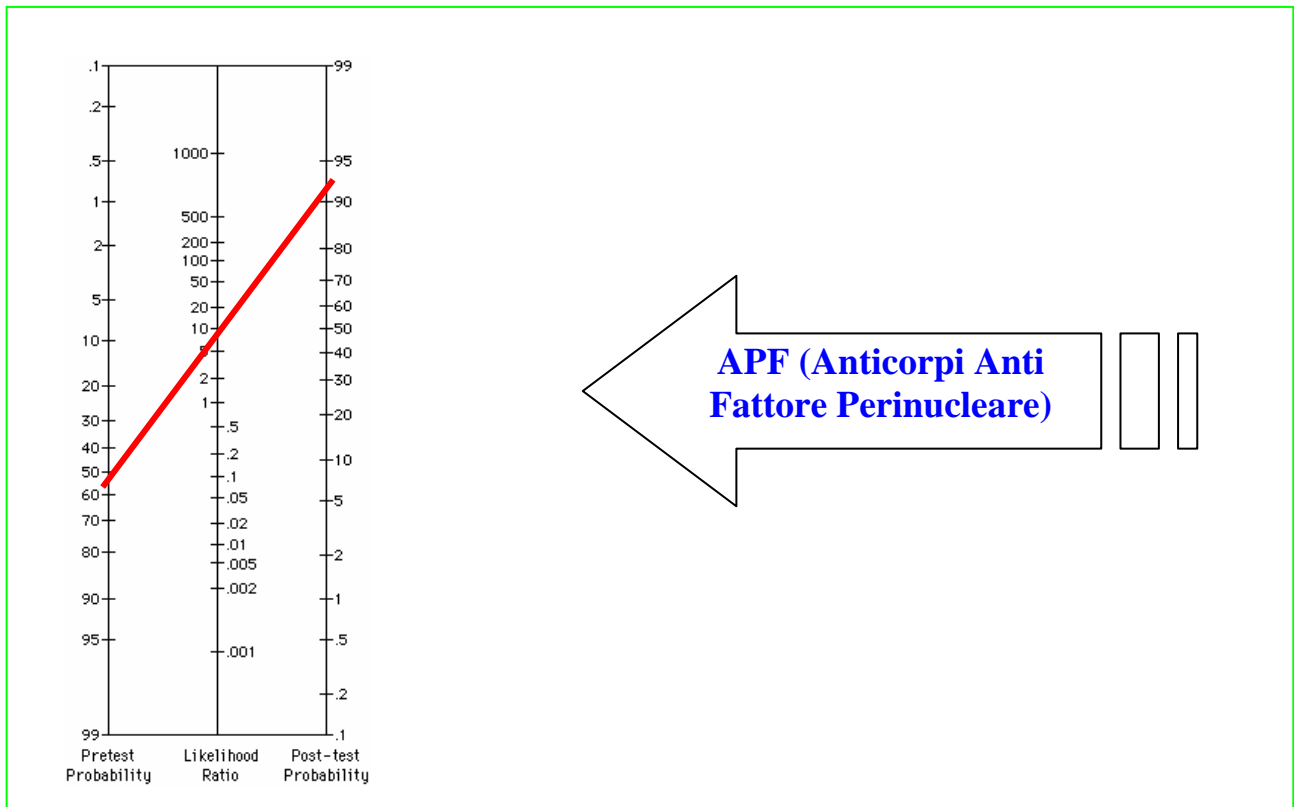
ESAME: A.P.F. (Fattore Perinucleare)

LIVELLO o tipo di evidenza dello studio	SENSIBILITA' (%)	SPECIFICITA' (%)	LR +	LR -
JANSENS X. et al. – (1988) (11) LIVELLO III	86,6	96,2	22,7	0,14
BERTHELOT J.M. et al. – (1995) (12) LIVELLO II	95	82	5,3	0,06
VINCENT C. et al. – (1999) (13) LIVELLO III	52	97	17,3	0,49
LI H. et al. – (2000) (16) LIVELLO III	35,2	91,4	4,0	0,70
BAS S. et al. – (2002) (10) LIVELLO II	46	94	7,7	0,57

Pertanto, considerando i dati complessivamente, e cercando di dare un maggior “peso” agli studi di livello II°, presumibilmente più significativi, possiamo affermare che l'A.P.F. presenta una sensibilità piuttosto bassa (60% circa), una specificità molto buona (91% circa), un valore di LR+ discreto (circa 8) ed un LR- modesto (circa 0,4).

Nel nostro caso specifico, essendo l'esame A.P.F. positivo, applicando il Nomogramma di Fagan, si ottengono dei risultati molto interessanti.

Infatti, assumendo come punto di partenza (probabilità pre-test) la probabilità post-test già calcolata con il F.R. otteniamo una probabilità post-test del 92% circa:



Naturalmente, ciò non rappresenta una diagnosi di certezza, ma sicuramente un notevole passo avanti in un campo, quello della patologia auto-immune, dove spesso la diagnosi viene raggiunta assai tardivamente.

I promettenti risultati ottenuti con l'A.P.F., hanno spinto numerosi autori verso la ricerca di simili marcatori di A.R., nel tentativo di aumentare specificità e sensibilità dell'analisi (7) (8) (10).

Per esempio, molto promettente è risultato l'anticorpo ANTI-CCP (anti-Citrullina) ottenendo i seguenti dati: Sensibilità: 68%, Specificità: 96%, LR+: 17, LR-: 0,33.

Per quanto riguarda l'APF il suo valore diagnostico è stato ulteriormente sottolineato dall'evidenza che pazienti con A.R. e con F.R. negativo presentano spesso (tra il 9-21% a seconda delle casistiche) un APF positivo (6) (11) e tale positività si associa più frequentemente ad una prognosi peggiore, con precoce evoluzione erosiva delle lesioni articolari(5) (11) (16), con una più rapida progressione radiologica del danno (17) (18) (19), tanto da indurre molti studiosi ad ipotizzare un precoce ed aggressivo intervento terapeutico, in media da iniziare entro 3 mesi, in questo tipo di pazienti.

Successivamente, l'efficacia della terapia può essere facilmente controllata periodicamente utilizzando i comuni e poco costosi parametri di flogosi quali VES, Fibrinogeno e PCR che si sono dimostrati molto utili a questo scopo (20) (21) (22).

Infine è da sottolineare come una diagnosi precoce della malattia, ed un suo altrettanto precoce trattamento al fine di prevenire le future lesioni articolari, e la conseguente disabilità, non può che

avere un importante impatto socio-economico: infatti dati recenti raccolti dalla Società Italiana di Reumatologia, in collaborazione con il CERGAS dell'Università Bocconi, hanno stimato un costo medio annuo per paziente affetto da A.R. in più di 12mila Euro; di questi, in particolare, il 35% è dovuto ai ricoveri, il 29% alla terapia, il 9% alla diagnostica ed il 27% è a carico del malato stesso.

E' stato stimato, infine, che più tempo trascorre tra l'esordio e la diagnosi della malattia maggiore sarà la gravità della stessa, con conseguente notevole impatto economico. I costi annui diretti ed indiretti per paziente, infatti, nella prima fase della malattia aumentano rispettivamente a 1640 e 2070 Euro, nella seconda a 2900 e 9570 Euro, nella terza a 4240 e 12.180 Euro, e nella quarta a 5700 e 17.250 Euro.

BIBLIOGRAFIA

- 1) BATTERLE C. – LE MALATTIE AUTOIMMUNI – (2001) – PAG. 458- 459 – EDIZ. PICCIN.
- 2) TODESCO S., GAMBARI P. – MALATTIE REUMATICHE – (1998) 2° EDIZ. – PAG. 156 ED. MC GRAW – HILL.
- 3) TOZZOLI R., BIZZARRO N. – LA DIAGNOSTICA DI LABORATORIO DELLE MALATTIE AUTOIMMUNI SISTEMICHE – VOL IX N. 1, 1998 – PAG. 171-
- 4) ROMAGNANI S. ET AL.. – MALATTIE DEL SISTEMA IMMUNITARIO – (2000) 2° EDIZ. – PAG. 343 – ED. MC GRAW – HILL.
- 5) VON ESSEN R. ET AL. – PROSPECT FOR AN ADDITIONAL LABORATORY CRITERION FOR RHEUMATOID ARTHRITIS – (1993) SCAND. J. RHEUMATOL. 22, 267 – 272.
- 6) LI H. ET AL. – SPECIFIC ANTIBODIES FOR THE EARLY DIAGNOSIS OF RHEUMATOID ARTHRITIS – ZHONGHA YI XUE ZA ZHI (2000) 80; 20 – 4.
- 7) SCHELLENKENS GA. ET AL. – THE DIAGNOSTICS PROPERTIES OF RHEUMATOID ARTHRITIS ANTIBODIES RECOGNIZING A CYCLIC CITRULLINATED PEPTIDE – ARTHRITIS RHEUM. (2000) 43; 155 – 63-
- 8) BIZZARRO N. ET AL. – DIAGNOSTIC ACCURACY OF THE ANTI-CITRULLINE ANTIBODY ASSAY FOR RHEUMATOID ARTHRITIS – CLIN. CHEM. (2001) 47; 1089 – 93-
- 9) GOLDBACH – MANSKY R. ET AL. – RHEUMATOID ARTHRITIS ASSOCIATED AUTOANTIBODIES IN PATIENTS WITH SYNOVITIS OF RECENT ONSET – ARTHRITIS RES. (2000) 2; 236 – 43-
- 10) BAS S. ET AL. – DIAGNOSTIC TESTS FOR RHEUMATOID ARTHRITIS : COMPARISON OF ANTI-CYLLIC CITRULLINATED PEPTIDE ANTIBODIES, ANTI-KERATINE ANTIBODIES AND IGM-RHEUMATOID FACTORS – RHEUMATOLOGY (2002) 41; 809 – 14-
- 11) JANSSENS X. ET AL. – THE DIAGNOSTIC SIGNIFICANCE OF THE ANTIPERINUCLEAR FACTOR FOR RHEUMATOID ARTHRITIS – J. RHEUMATOL. (1988); 15; 1346 – 40-
- 12) BERTHELOT J. M. – SPECIFICITY OF ANTIPERINUCLEAR FACTOR FOR RHEUMATOID ARTHRITIS IN RHEUMATOID FACTOR – POSITIVE SERA – BR. J. RHEUMATOL. (1995); 34; 716 – 720-
- 13) VINCENT C. ET AL. – ANTI-PERINUCLEAR FACTOR COMPARED WITH THE SO CALLED “ANTI-KERATIN” ANTIBODIES AND ANTIBODIES TO HUMAN EPIDERMIS FILAGGRIN, IN THE DIAGNOSIS OF ARTHRITIDES – ANN. RHEUM. DIS. (1999); 58; 42 – 8-
- 14) YOUINOU P. ET AL. – THE ANTIPERINUCLEAR FACTOR. I. CLINICAL AND SEROLOGIC ASSOCIATIONS- CLIN. EXP. RHEUMATOL. (1990) 8; 259 – 264-
- 15) CALZAVARA – PINTON P.G. ET AL. – INCIDENCE OF ANTIPERINUCLEAR FACTOR IN PATIENTS WITH PSORIATIC ARTHRITIS – ADV. EXP.MED.BIOL. (1999) 455;215 – 220-

- 16) KURKI P. ET AL. – ANTIBODY TO STRATUM CORNEUM (AKA) (ANTKERATIN ANTIBODY) AND ANTI – PERINUCLEAR FACTOR : MARKERS FOR PROGRESSIVE RHEUMATOID ARTHRITIS – SCAND. J. RHEUMATOL. (1997) 26; 346 – 9 –
- 17) COMBE B. ET AL. – PROGNOSTIC FACTORS FOR RADIOGRAPHIC DAMAGE IN EARLY RHEUMATOID ARTHRITIS : A MULTIPARAMETER PROSPECTIVE STUDY – ARTHRITIS RHEUM (2001) 44 ; 1736 – 43 –
- 18) GENEVAY S. ET AL. – AN EIGHT YEAR PROSPECTIVE STUDY OF OUTCOME BY ANTIPERINUCLEAR FACTOR AND ANTIKERATIN ANTIBODIES AT ONSET OF RHEUMATOID ARTHRITIS - ANN. RHEUM. DIS. (2002) 61; 734 – 6-
- 19) MEYER O. ET AL. – ANTICITRULLINATED PROTEIN/PEPTIDE ANTIBODY ASSAYS IN EARLY RHEUMATOID ARTHRITIS FOR PREDICTING FIVE YEARS RADIOGRAPHIC DAMAGE – ANN. RHEUM. DIS. (2003) 62; 120 – 6-
- 20) ARVIDSON NG. ET AL. – DISEASE ACTIVITY IN RHEUMATOID ARTHRITIS: FIBRINOGEN IS SUPERIOR TO THE ERYTHROCYTE SEDIMENTATION RATE – SCAND. J. CLIN. LAB. INVEST. (2002) 62; 315 – 9-
- 21) WOLFE F. – COMPARATIVE USEFULNESS OF C-REACTIVE PROTEIN AND ERYTHROCYTE SEDIMENTATION RATE IN PATIENTS WITH RHEUMATOID ARTHRITIS – J.RHEUMATOL- (1997) 24; 1477 – 85-
- 22) DEVLIN J. ET AL. – THE ACUTE PHASE AND FUNCTION IN EARLY RHEUMATOID ARTHRITIS. C-REACTIVE PROTEIN LEVELS CORRELATE WITH FUNCTIONAL OUTCOME – J. RHEUMATOL. (1997) 24; 9 – 13-