VALUE IN HEALTH 16 (2013) 619–628

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Meta-Analysis of the Accuracy of Two Diagnostic Tests Used in Combination: Application to the Ddimer Test and the Wells Score for the Diagnosis of Deep Vein Thrombosis

Nicola Novielli, PhD, Alexander J. Sutton, PhD, Nicola J. Cooper, PhD*

Department of Health Sciences, University of Leicester, Leicester, UK

ABSTRACT

Objectives: It is standard practice for diagnostic tests to be evaluated against gold standards in isolation. In routine clinical practice, however, it is commonplace for multiple tests to be used before making definitive diagnoses. This article describes a meta-analytic modeling framework developed to estimate the accuracy of the combination of two diagnostic tests, accounting for the likely nonindependence of the tests. **Methods:** A novel multicomponent framework was developed to synthesize information available on different parameters in the model. This allows data to be included from studies evaluating single tests or both tests. Different likelihoods were specified for the different sources of data and linked by means of common parameters. The framework was applied to evaluate the diagnostic accuracy of the Ddimer test and the Wells score for deep vein thrombosis, and the results were compared with those of a model in which independence of tests was assumed. All models were evaluated by using Bayesian Markov chain Monte Carlo simulation methods. **Results:** The results showed the importance of allowing for the (likely) nonindependence of tests in the meta-analysis model when evaluating a combination of diagnostic tests. The analysis also highlighted the relatively limited impact of those studies that evaluated only one of the two tests of interest. **Conclusions:** The models developed allowed the assumption of independence between diagnostic tests to be relaxed while combining a broad array of relevant information from disparate studies. The framework also raises questions regarding the utility of studies limited to the evaluation of single diagnostic tests.

Keywords: Bayesian methods, diagnostic accuracy, evidence synthesis.

Introduction

Accurate diagnosis is a prerequisite for the efficient allocation of treatments. Diagnostic tests with perfect or very high accuracy (reference tests) are often expensive and/or invasive; therefore, index tests, which are usually cheaper and less invasive but also less accurate, often play an important role in medical diagnosis. Rarely is the application of one index test sufficient to diagnose a particular condition, and diagnostic strategies involving multiple tests are often used in routine clinical practice. Where multiple tests are used for diagnosis, however, it is important to acknowledge that the diagnostic results from the different tests may not be independent of one another, and therefore, when synthesizing evidence to evaluate the accuracy of the combination of tests, this interdependence needs to be taken into account, which is seldom done in practice.

Systematic reviews and, consequently, meta-analyses are routinely used to identify the evidence for medical decision making [1] and, more specifically, for clinical/economic decision analytic modeling [2] because optimal decisions should not be based solely on single study results when multiple studies with relevant data exist [3,4]. Systematic reviews and meta-analyses of diagnostic test accuracy studies have focused on the performance of individual tests, which, at least in part, is due to a large proportion of primary studies focusing on the evaluation of single tests. A recent systematic review of health technology assessment reports [5] found that where economic decision models had been used to evaluate different combinations of tests, the accuracy of each combination was calculated on the basis of the assumption of either 1) conditional independence between tests or 2) the accuracy of the second test to be perfect. There is evidence that when the assumption of dependence between tests is not met, both the meta-analysis (for the estimates of the accuracy rates) and the economic evaluation (informed by the meta-analysis results) have the potential to give misleading conclusions [6]. In this article, we focus solely on clinical effectiveness, with an associated article [7] focusing on cost-effectiveness implications.

This article was presented at the 2nd International Symposium on Methods for Evaluating Medical Tests and Biomarkers in Birmingham, UK, July 1-2, 2010, and at the 19th Cochrane Colloquium in Madrid, Spain, October 19-22, 2011.

* *Address correspondence to:* Alexander J. Sutton, Department of Health Sciences, University of Leicester, University Road, Leicester LE1 7RH, UK.

E-mail: njc21@le.ac.uk.

http://dx.doi.org/10.1016/j.jval.2013.02.007

Table 1 – Data extracted by the systematic review of studies reporting both the Wells score and the Ddimer test.

First author (year)	Wells score accuracy data		Ddimer test accuracy data			
	Stratum	Diseased/total	TP	FP	FN	TN
Type A studies: Complete count data for the Ddimer test and the Wells score						
T1. Shields (2002)	Low	1/41	1	8	0	32
	Moderate	6/44	6	18	0	20
	High	10/17	8	2	2	5
T2. Lennox (1999)	Low	4/88	3	8	1	76
	Moderate	12/67	9	12	3	43
	High	30/45	30	8	0	7
T3. Kearon (2001)	Low	5/206	4	25	1	176
	Moderate	24/188	17	51	7	113
	High	35/49	33	8	2	6
T4. Ruiz-Gimenez (2004)	Low	16/135	15	49	1	70
	Moderate	31/136	31	51	0	54
	High	55/112	54	36	1	21
T5. Yamaki (2005)	Low	1/38	1	20	0	17
	Moderate	22/64	22	23	0	19
	High	35/56	35	9	0	12
T6. Anderson (2000)	Low	4/118	4	17	0	97
	Moderate	9/66	6	9	3	48
	High	15/30	13	2	2	13
T7. Anderson (2002)	Low	20/446	17	113	3	313
	Moderate	76/192	61	93	15	23
	High	94/199	79	55	15	50
T8. Bates (2003)	Low	18/296	18	85	0	193
	Moderate	17/189	16	83	1	89
	High	21/71	21	30	0	20
T9. Rio Sola (1999)	Low	28/32	23	1	5	3
	Moderate	44/55	37	6	7	5
	High	9/14	9	3	0	2
T10. Williams (2005)	Low	6/89	6	42	0	41
	Moderate	18/123	15	59	3	46
	High	11/31	10	16	1	4
T11. Yamaki (2009)	Low	29/505	28	233	1	243
	Moderate	117/237	117	104	0	16
	High	109/141	109	29	0	3
Type B studies: Complete count data for the Wells score and partial count data for the Ddimer test						
T12. Borg (1997)	Low	2/32	NA	NA	NA	NA
	Moderate	4/15	NA	NA	NA	NA
	High	26/29	25	2	1	1
T13. Dewar (2008)	Low	9/166	9	70	0	87
	Moderate	17/166	NA	NA	NA	NA
	High	30/108	NA	NA	NA	NA
T14. Elf (2008)	Low	13/159	12	37	1	109
	Moderate	37/141	NA	NA	NA	NA
	High	33/57	NA	NA	NA	NA
Type C studies: Partial count data for both the Wells score and the Ddimer test						
T15. Aguilar-Franco (2002a)	Low	2/149	2	76	0	71
	Moderate	NA/NA	NA	NA	NA	NA
	High	NA/NA	NA	NA	NA	NA
T16. Walsh (2009)	Low	4/49	4	23	0	22
	Moderate	NA/NA	NA	NA	NA	NA
	High	NA/NA	NA	NA	NA	NA
T17. Aguilar-Franco (2002b)	Low	NA/NA	NA	NA	NA	NA
	Moderate	26/134	26	73	0	35
	High	NA/NA	NA	NA	NA	NA
T18. Bucek (2002)	Low	2/93	2	43	0	48
	Moderate	NA/NA	NA	NA	NA	NA
	High	NA/NA	NA	NA	NA	NA

FN, false negative; FP, false positive; NA, not available/reported; TN, true negative; TP, true positive.