VALUE IN HEALTH 16 (2013) 536–541

Available online at www.sciencedirect.com

SciVerse ScienceDirect

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

Evaluating the Cost-Effectiveness of Diagnostic Tests in Combination: Is It Important to Allow for Performance Dependency?

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

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

ABSTRACT

Objectives: To investigate the importance of accounting for potential performance dependency when evaluating the cost-effectiveness of two diagnostic tests used in combination. **Methods:** Two meta-analysis models were fitted to estimate the diagnostic accuracy of Wells score and Ddimer in combination. The first model assumes that the two tests perform independently of one another; thus, two separate meta-analyses were fitted to the Ddimer and Wells score data and then combined. The second model allows for any performance dependency of the two tests by incorporating published data on the accuracy of Ddimer stratified by Wells score, as well as studies of Ddimer alone and Wells score alone. The results from the two meta-analysis models were input into a decision model to assess the impact that assumptions regarding performance dependency have on the overall cost-effectiveness of the tests. **Results:** The results highlight the importance of accounting for potential performance dependency when evaluating the cost-effectiveness of diagnostic tests used in combination. In our example, assuming the diagnostic performance of the two tests to be independent resulted in the strategy "Wells score moderate/high risk treated for DVT and Wells score low risk tested further with Ddimer" being identified as the most cost-effective at the £20,000 willingness-to-pay threshold (probability cost-effective 0.8). However, when performance dependency is modeled, the most cost-effective strategies were "Ddimer alone" and "Wells score low/moderate risk discharged and Wells score high risk further tested with Ddimer" (probability cost-effective 0.4). **Conclusions:** When evaluating the effectiveness and cost-effectiveness of diagnostic tests used in combination, failure to account for diagnostic performance dependency may lead to erroneous results and non-optimal decision making.

Keywords: conditional diagnostic accuracy, cost-effectiveness, diagnostic tests, evidence synthesis.

Introduction

In the area of diagnostic test performance, evidence-based evaluations are crucial to the decision-making process because early diagnosis can lead to diseases being treated more successfully than if treatment were delayed. Often evaluations are performed by focusing on the accuracy of a single test to diagnose a particular condition [1]; however, in routine clinical practice, a diagnosis is usually based on the results obtained from multiple tests.

A recent review of the National Institute for Health Research health technology assessment reports of decision models for diagnostic tests containing meta-analysis results from 1997 to 2009 [1] found that 6 of the 14 (43%) reports included in the review considered a combination of diagnostic tests strategy in the economic decision modeling part of the report. In these six reports, the accuracy of each combination of diagnostic tests was calculated by either assuming 1) conditional independence between tests or 2) the accuracy of the second test to be perfect (which may be reasonable to assume in some contexts). Where multiple tests are used for diagnosis, however, it is highly likely that the tests will not perform independently (i.e., in the case of two tests, the performance of the second test may differ depending on the results of the first test), and therefore it is important to allow for this in the analysis. In fact, there is evidence that when the assumption of dependence between tests is ignored, this may lead to erroneous disease probability estimates [2], which, if input into an economic decision model, will carry forward into the cost-effectiveness analysis results.

In this article, we investigated the importance of allowing for potential performance dependency when evaluating the cost-effectiveness of two diagnostic tests used in combination (which uses recent advances in meta-analysis methodology outlined in our companion article [3]). This was assessed by observing the impact on the cost-effectiveness results, and subsequent conclusions reached, when performance dependency is first ignored and then incorporated. The article focuses on the example of diagnosing deep vein thrombosis (DVT) by using Wells score and Ddimer.

Motivating Example: Ddimer and Wells Score Tests for the Diagnosis of DVT

DVT is a blood clot in a deep vein (lower limb) that is usually treated with anticoagulants. An accurate diagnosis of DVT is crucial

* *Address correspondence to:* Nicola Cooper, Department of Health Sciences, University of Leicester, University Road, Leicester LE1 7RH, UK.

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

1098-3015/$36.00 – see front matter Copyright © 2013, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

http://dx.doi.org/10.1016/j.jval.2013.02.015

to lower mortality due to venous thromboembolism–related adverse events and also to reduce the impact of side effects from anticoagulant treatment given to patients wrongly diagnosed with DVT.

DVT may be diagnosed by reference tests such as ultrasound and venography, which have high diagnostic accuracy, but such tests are expensive to perform. Therefore, cheaper, quicker but less accurate tests are often used for the diagnosis of DVT. Two of these tests, which will be considered in this article, are Ddimer (i.e., measures the concentration of an enzyme in the blood, the higher the measurement the more likely DVT) and Wells score (devised from an assessment of the clinical features of DVT such as clinical history, symptoms, and signs [4,5]). For use in diagnosis, the latter test is usually categorized into *low* (score <1), *moderate* (score 1 or 2), and *high* (score >2) risk of having DVT. For more details about the diagnostic performance data of Ddimer, Wells score, and Ddimer given Wells score used in the analysis, see Novielli et al. [3].

In a recent review, Goodacre et al. [6] found Ddimer and Wells score to not be accurate enough as stand-alone diagnostic tools but that algorithms containing both Wells score and Ddimer were potentially valuable for diagnosis.

Methods

Diagnostic Accuracy Meta-Analysis Models

To investigate the impact that the assumption of test performance dependency has on the cost-effectiveness results, two statistical analyses were undertaken to obtain the joint diagnostic accuracy for Wells score and Ddimer when used in combination.

The first analysis assumed the diagnostic performance of the two tests to be independent and therefore used the results from two separate meta-analyses to inform the cost-effectiveness decision model. As diagnostic accuracy is usually measured in terms of both sensitivity (i.e., the proportion of actual positives that are correctly identified) and specificity (i.e., the proportion of negatives that are correctly identified), bivariate meta-analyses [7] (which allow for the between-study correlation of sensitivity and specificity potentially induced through varying test thresholds used in the different studies) were fitted to the Wells score data and the Ddimer data separately.

The second analysis used the meta-analytic modeling framework developed by Novielli et al. [3], to account for test performance dependency in the estimation of the diagnostic accuracy of Wells score and Ddimer used in combination. This analysis used a multicomponent meta-analysis framework [8] to incorporate data from studies reporting the accuracy of Ddimer stratified by Wells score, as well as studies of Ddimer alone and Wells score alone. Random effects models were used with different likelihoods required for the different data types but linked together through the use of shared parameters [9,10]. For more details about the data (together with references) and the analysis, see Novielli et al. [3].

Decision Model

The comprehensive cost-effectiveness decision model (i.e., integrating the meta-analysis and decision model into a single coherent framework [11]) for evaluating a single diagnostic test used by Sutton et al. [12] (adapted from Goodacre et al. [6]) was modified to allow for the incorporation of two tests in combination (Fig. 1). This decision model assumed a simplified diagnosis-to-treatment pathway for DVT whereby patients who were diagnosed as positive (on the basis of one of the strategies defined in the next section) were treated with anticoagulants, which potentially may cause harmful side effects such as bleeding at different intensities (i.e., false- and true-positive patients may be subject to nonfatal bleeding, fatal intracranial bleeding, nonfatal intracranial bleeding, or no bleeding when treated with anticoagulants). The accuracy parameters (i.e., false positive, false negative, true positive, and true negative for Wells score and Ddimer) were informed by the meta-analysis models discussed above. All other model parameter values and sources (i.e., prevalence of DVT, risk of pulmonary embolism, quality of life-adjusted life-years per each possible health status, costs, etc.) are reported in Goodacre et al. [6].

Strategies

As mentioned previously, for use in diagnosis, Wells score is usually categorized into *low* (score <1), *moderate* (score 1 or 2), and *high* (score >2) risk of having DVT. For the analyses presented in this article, three different classifications of Wells score were used, that is, 1) WS_1—low (score <1) and moderate/high (score ≥ 1), 2) WS_2—low/moderate (score ≤ 2) and high (score >2), or 3) WS_3—low (score <1), moderate ($1 \leq$score≤ 2), and high (score >2).

In our cost-effectiveness analyses, 10 diagnostic strategies were considered as outlined in Table 1. For each dichotomy of two diagnostic tests, two possible strategies can be defined [13]: 1) “believe the negatives”—only patients diagnosed as positive by the first test received the second test (i.e., $(WS_1$ & $DD)_{BN}$ and $(WS_2$ & $DD)_{BN}$) and 2) “believe the positives”—only patients diagnosed as negative by the first test received the second test (i.e., $(WS_1$ & $DD)_{BP}$ and $(WS_2$ & $DD)_{BP}$).

Note that for every pair of dichotomous (or dichotomized) tests combined according to one of the strategies “believe the negative” or “believe the positive,” the order of the tests does not affect the diagnostic accuracy of the strategy [3] (though which test is conditioned on may affect the estimation of effectiveness parameters in the synthesis model) but may affect the costs incurred. For example, Wells score is usually given first (i.e., to everyone with suspected DVT) because it is less expensive than Ddimer, does not require any specialist technology, and can be carried out by an experienced doctor quickly at initial presentation. Therefore, given that Wells score is less expensive than Ddimer, any sequence of the two tests (i.e., diagnostic strategies 1, 2, 3, and 4 listed above) where Wells score is dichotomized and given first will be dominant from an economic point of view compared with the same equally accurate strategy where Ddimer is given first. Note that for strategy 8, where the first test, Wells score, is not dichotomized, this property does not hold.

Modeling Framework

All analyses were conducted by using a comprehensive decision modeling framework [11] that evaluates both the evidence synthesis models and the decision model within a single coherent framework. The modeling framework was implemented by using Markov Chain Monte Carlo simulation in WinBUGS software [14]. Noninformative (vague) prior distributions were used for all parameters estimated by the statistical model. Graphical tools were used to assess the convergence of the Markov Chain Monte Carlo chains, and sensitivity analyses were performed to assess the influence of the initial values and prior distributions on the results. The WinBUGS code (including the specific prior distributions used) is provided in the Appendix in Supplemental Materials found at http://dx.doi.org/%2010.1016/j.jval.2013.02.015.