European Journal of Obstetrics & Gynecology and Reproductive Biology 167 (2013) 81–85

Contents lists available at SciVerse ScienceDirect

European Journal of Obstetrics & Gynecology and Reproductive Biology

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

Human epididymis protein 4 for differential diagnosis between benign gynecologic disease and ovarian cancer: a systematic review and meta-analysis

JiaYing Lin [a,*], JinBao Qin [b], Voramon Sangvatanakul [a]

[a] Clinic for Gynecology, Charite-Universitatsmedizin Berlin, Berlin, Germany
[b] Department of Vascular Surgery, The Ninth People's Hospital, Affiliated to Shanghai Jiao Tong University, School of Medicine, Shanghai, China

ARTICLE INFO

Article history:
Received 9 June 2012
Received in revised form 30 July 2012
Accepted 31 October 2012

Keywords:
Ovarian neoplasms
Human epididymis protein 4
CA125
Risk of ovarian malignancy algorithm
Meta analysis

ABSTRACT

Objectives: Human epididymis protein 4 (HE4) is a new biomarker for the detection of ovarian cancer. The objective of this review was to assess by meta-analysis the overall diagnostic accuracy of HE4 assay in differentiating malignant ovarian tumors from benign gynecology diseases.
Study design: The MEDLINE, EMBASE, and Cochrane Library databases were searched for studies published up to June 2012 that evaluated HE4 accuracy. Meta-analysis was used to calculate sensitivity, specificity, the positive likelihood ratio (PLR), the negative likelihood ratio (NLR) and the area under curve (AUC).
Results: A total of 11 studies with 3395 patients who fulfilled all inclusion criteria were considered in the analysis. No publication bias was found. HE4 had a pooled sensitivity of 0.74 (95% confidence interval (CI), 0.72–0.76) and a pooled specificity of 0.87 (95% CI, 0.85–0.89). Overall, the positive likelihood ratio was 8.04 (95% CI, 4.89–13.21) and the negative likelihood ratio was 0.27 (95% CI, 0.22–0.34). When HE4 was combined with CA125, the sensitivity was higher than that of HE4 alone at the expense of lower specificity.
Conclusions: The measurement of serum HE4 is a useful method for differential diagnosis between benign gynecologic disease and ovarian cancer.

1. Introduction

Ovarian cancer is the third most common tumor of the female genital tract after carcinomas of the cervix and endometrium, and continues to be the leading cause of death from gynecologic malignancies [1]. In clinical practice, discrimination between benign disease and malignant ovarian tumor in patients with an adnexal mass still remains a challenge for gynecologists, and is important due to the high mortality rate in patients diagnosed with advanced cancer [2].

Early diagnosis and timely surgery and/or chemotherapy are considered the most efficient principles of ovarian cancer therapy. Currently, a combination of physical examination, serum CA125 level, and imaging affords the highest positive predictive value [3–5]. CA125 measurement is an important component in the workup of a woman with an adnexal mass, but its utility is hindered by low specificity, especially in pre-menopausal women where CA125 is elevated above normal in common benign conditions, such as pelvic endometriosis, follicular cysts, cystadenoma, tubo-ovarian abscess, and pregnancy [6]. CA125 is also elevated in less than half of early-stage epithelial ovarian cancer (EOC), and is not expressed in approximately 20% of EOC, resulting in decreased sensitivity [7]. There is therefore a pressing need for novel markers that are sensitive and specific, and can improve the diagnosis of ovarian cancer when used in combination with CA125 or can replace it.

Recently, human epididymis protein 4 (HE4) has been proposed as a new tumor marker for ovarian cancer. HE4 has been suggested to have a diagnostic sensitivity similar to that of CA125, but an increased diagnostic specificity in patients with gynecologic malignancies compared with those having benign gynecologic diseases [8,9]. Results from previous studies are controversial and inconclusive, however, because most of them are single studies with limited sample sizes and use a variety of methods for determining the performance of HE4. In the current study, a meta-analysis of all available studies was conducted to evaluate the performance of serum HE4 measurement in the diagnosis of ovarian cancer. To our knowledge there has been no previous similar meta-analysis.

2. Materials and methods

2.1. Literature search

Original and review articles published until June 2012 that analyzed the diagnostic performance of HE4 were systematically

* Corresponding author at: Hindenburgdamm 30, 12203 Berlin, Germany. Tel.: +49 03529 5345; fax: +49 03529 5333.
E-mail address: jiaying.lin@charite.de (J. Lin).

0301-2115/$ – see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ejogrb.2012.10.036

searched in the MEDLINE, EMBASE, and Cochrane databases. The following keywords were used: (“HE4” OR “WFDC2”) and (“ovarian carcinoma” or “ovarian cancer” or “carcinoma of ovary”) and (“sensitivity” or “specificity” or “false negative” or “false positive” or “diagnosis” or “detection” or “accuracy”). We evaluated all associated publications to retrieve the most eligible studies. Their reference lists were searched manually to find other relevant publications.

2.2. Selection of studies

The eligibility criteria for the meta-analysis of the studies included the following: (1) both the sensitivity and specificity of HE4 levels for the diagnosis of ovarian cancer were provided, or HE4 values were provided in a scatter plot form, allowing test results to be extracted for each individual; (2) 50 or more patients were included; and (3) the study design included women with ovarian cancer and benign gynecologic diseases, and evaluated the contribution of HE4.

Articles were excluded when data were insufficient to construct a 2×2 table of the test result (serum HE4 concentration). The 2×2 tables were constructed independently by two of the authors (JY.L. and JB.Q.). In the event of disagreement, the judgment of a third author (V. S.) was decisive.

2.3. Data extraction

The final set of English articles was assessed independently by two observers (JY.L. and JB.Q.). The observers were blinded to publication details, and differences between them were resolved by consensus. Data retrieved from the reports included the name of the author, publication year, participant characteristics, test method, cutoff value, sensitivity, specificity, and study quality score.

We assessed the quality of the included studies by the criteria selected from the Quality Assessment for Studies of Diagnostic Accuracy checklist for the assessment of diagnostic studies [10]: study design (prospective or retrospective), patient selection (consecutive or not), blinding (blind or not to the interpretation of index text), assay method, study size, etc. The numbers of true-positive (TP), false-negative (FN), false-positive (FP), and true-negative (TN) results in the detection of ovarian cancer were extracted on a per-patient or per-lesion basis.

2.4. Statistical analysis

We calculated the sensitivity and specificity for each study, and then pooled the results per the DerSimonian Liard random effect model [11]. We also calculated the area under the curve (AUC), which represents an overall summary measure of the curve and the test's overall ability to accurately distinguish cases from non-cases.

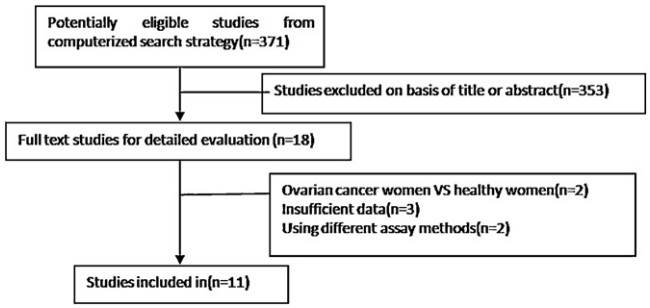

Fig. 1. Results of search strategy.

Heterogeneity was also assessed by the likelihood ratio I^2 index and X^2 test. The I^2 index is a measure of the total variation percentage across studies due to heterogeneity beyond chance; values over 50% indicate heterogeneity [12]. In the X^2 test, $p < 0.05$ was considered as having apparent heterogeneity. If heterogeneity existed [13], a random effect model was used in the primary meta-analysis to obtain a summary estimate for sensitivity with 95% confidence interval (CI). We also tested potential publication biases. All these statistical analyses were undertaken using the Meta-Disc (Version 1.4) [13]. Meta-Disc, produced by Javier Zamora, is freeware software that performs a systematic review of studies that evaluate diagnostic and screening tests. p values <0.05 were considered to be statistically significant.

The potential presence of publication bias was assessed with a funnel plot, the Begg test and the Egger test. These analyses were performed by using the commands for the meta-analysis of diagnostic studies in STATA software (Version 12.0; Stata Corporation, College Station, TX, USA).

3. Results

3.1. Literature search and study design characteristics

Our research yielded 371 primary studies. Among them, 353 were excluded after reviewing the title and abstract, and seven articles were excluded after reviewing the full article (Fig. 1). Two articles were excluded because ovarian cancer patients were compared with healthy populations to evaluate the contributions of HE4 [14,15]. Three articles were excluded because there were insufficient data to calculate TP, TN, FP, and FN values [16–18]. Two articles [8,19] were excluded because the EIA/ELISA assays which were used in these two articles (before 2009) were very different from the EIA/ELISA assays run after 2009. Ultimately, a total of 11 studies [20–30] with 3395 patients fulfilled all inclusion criteria and were considered for the analysis (Table 1).

Table 1
Characteristics of the studies included in the meta-analysis.

Author	Year of publication	Location	Storage temperature (°C)	Number of patients'serum	Study design	Blinding	Patients enrollment	Sensitivity (%)	Specificity (%)
Huhtinen	2009	Finland	−20	143	Retrospective	ND	Consecutive	71.4	95
Nolen	2010	USA	ND	790	ND	ND	ND	71.7	85.2
Abdel-Azeez	2010	Egypt	−80	65	ND	ND	ND	82.9	87.5
Gorp	2011	Belgium	−80	389	Prospective	ND	Consecutive	74.5	83.3
Jacob	2011	Switzerland	−80	160	Prospective	ND	ND	83.3	84.6
Holcomb	2011	USA	ND	494	Prospective	Yes	Consecutive	64.7	91.8
Chang	2011	China	−70	202	ND	Yes	Consecutive	73	98.7
Montagnana	2011	Italy	−80	104	Retrospective	ND	Consecutive	76.4	93.6
Park	2011	Korea	−70	323	Retrospective	ND	Consecutive	90.9	64
Molina	2011	Spain	ND	396	ND	ND	ND	79.3	98.9
Partheen	2011	Sweden	−80	329	prospective	ND	Consecutive	78.1	75