Pathology - Research and Practice 216 (2020) 153025

Contents lists available at ScienceDirect

Pathology - Research and Practice

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

Review

Diagnostic accuracy of p53 immunohistochemistry as surrogate of TP53 sequencing in endometrial cancer

Antonio Raffone[a], Antonio Travaglino[b,*], Marco Cerbone[a], Caterina De Luca[c], Daniela Russo[b], Anna Di Maio[d], Margot De Marco[e,**], Maria Caterina Turco[e], Luigi Insabato[b], Fulvio Zullo[a]

[a] Gynecology and Obstetrics Unit, Department of Neuroscience, Reproductive Sciences and Dentistry, School of Medicine, University of Naples Federico II, Naples, Italy
[b] Anatomic Pathology Unit, Department of Advanced Biomedical Sciences, School of Medicine, University of Naples Federico II, Naples, Italy
[c] Molecular Biology Unit, Department of Public Health, School of Medicine, University of Naples Federico II, Naples, Italy
[d] RSA Oasi San Francesco, Castellammare di Stabia, Italy
[e] Department of Medicine, Surgery and Odontology Schola Medica Salernitana, University of Salerno, Baronissi, SA, 84081, Italy

ARTICLE INFO

Keywords:
Prognosis
Treatment
Endometrium
Risk assessment
ProMisE
Molecular

ABSTRACT

Aberrant p53 immunohistochemical expression is used to identify the copy-number-high/*TP53*-mutant subgroup of endometrial cancer (EC). We aimed to determine the diagnostic accuracy of p53 immunohistochemistry as surrogate for *TP53* sequencing through a systematic review and meta-analysis.

Electronic databases were searched from their inception to June 2019. All studies assessing p53 expression and *TP53* mutations in EC were included. Diagnostic accuracy was assessed based on area under the curve (AUC). Immunohistochemical criteria used to define aberrant p53 expression were "overexpression" and "overexpression or complete absence". Subgroup analysis was based on the sequencing technique adopted (Polymerase Chain Reaction + sequencing, or next generation sequencing, NGS).

Thirteen observational studies with 727 endometrial cancers were included. Both "overexpression" and "overexpression or complete absence" showed high diagnostic accuracy (AUC = 0.9088 and 0.9030, respectively). The subgroup with "overexpression" and NGS showed the best results, with very high diagnostic accuracy (AUC = 0.9927).

In conclusion, immunohistochemistry for p53 is a highly accurate surrogate of *TP53* sequencing. Overexpression of p53 in $\geq 70-80\%$ showed the best accuracy in predicting *TP53* mutations. Further studies in this field should adopt optimized immunohistochemical procedures and take into account less common p53 patterns (e.g. cytoplasmic expression).

1. Introduction

Endometrial cancer (EC) is the most prevalent gynecologic cancer in the developed world, and the fourth most common cancer in women overall [1]. Although EC often shows good outcomes, women with advanced disease or more aggressive subtypes may not be curable with adjuvant therapy [2–4]. In the last decades, mortality rates of EC have increased globally [1]; this has been attributed to an inaccurate and little reproducible risk stratification, which has led to overtreatment and undertreatment of thousands of women [5–7].

In 2013, The Cancer Genome Atlas (TCGA) Research Network has proposed a reclassification of EC that will likely affect post-surgical adjuvant treatment for women with aggressive tumors. This reclassification divided ECs into four molecular subgroups correlated with prognosis: POLE/ultramutated, microsatellite-instability/hypermutated, copy-number-low, and copy-number-high. The copy-number high subgroup was characterized by *TP53* mutations and showed the worst prognosis [8].

Given the costs, complexity and time required for such molecular classifier [2,7], immunohistochemistry has been proposed as a more widely applicable surrogate of molecular techniques [9–13]. In this scenario, p53 immunohistochemistry has been proposed as a surrogate test to identify the copy-number high group, as the aberrant expression of p53 reflects *TP53* mutations [2,14–16].

However, the criteria to define p53 expression as "aberrant" have not been consistent among studies and over time. In fact, although an

* Corresponding author at: Anatomic Pathology Unit, Department of Advanced Biomedical Sciences, School of Medicine, University of Naples Federico II, Via Sergio Pansini, 5, Naples, 80131, Italy.
** Corresponding author at: Department of Medicine, Surgery and Odontology Schola Medica Salernitana, University of Salerno, Baronissi, SA, 84081, Italy.
E-mail addresses: antonio.travaglino.ap@gmail.com (A. Travaglino), mdemarco@unisa.it (M. De Marco).

https://doi.org/10.1016/j.prp.2020.153025
Received 20 April 2020; Received in revised form 14 May 2020; Accepted 15 May 2020

overexpression of p53 overexpression has always been considered as aberrant [2,6,14–26], the cut-offs of intensity of staining and percentage of stained nuclei have been variable; furthermore, other p53 patterns have recently been found associated with *TP53* mutations, such as cytoplasmic expression and complete absence [6,7,14,15,24–27].

Objective of this systematic review and meta-analysis was to assess the diagnostic accuracy of immunohistochemistry for p53 as surrogate for *TP53* sequencing in EC, according to the different criteria adopted in the literature.

2. Materials and methods

2.1. Study protocol

Methods for electronic search, study selection, risk of bias assessment, extraction and analysis of data were defined a priori. Two authors (AR, AT) independently performed all review steps. Disagreements were resolved by discussion with all authors. The study was reported following the Preferred Reporting Item for Systematic Reviews and Meta-analyses (PRISMA) statement [28] and the Synthesizing Evidence from Diagnostic Accuracy Tests (SEDATE) guidelines [29].

Before data extraction, the study protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO) (registration No.: CRD42019133621) following the PRISMA guidelines for protocols (PRIMSA-P).

2.2. Search strategy

MEDLINE, Google Scholar, Web of Sciences, Scopus, Cochrane Library, EMBASE and ClinicalTrial.gov were searched from their inception to June 2019, by using a combination of the following text words and all their synonyms found on Medical SubHeading (MeSH) vocabulary: “TP53″; “p53″; “tumor protein 53″; “endometrium”; “endometrial cancer”; “endometrioid adenocarcinoma”; “serous”; “undifferentiated”; “clear cell”; “endometrium”; “immunohistochemistry”; “immunohistochemical”; “marker”; “prognosis”; “Atlas”; “cancer”; “genome”; “TCGA”; “PORTEC”; “TransPORTEC”; “Proactive Molecular Risk Classifier”; “ProMisE”. Relevant references from each selected study were also evaluated.

2.3. Study selection

All peer-reviewed, retrospective or prospective studies evaluating the association between p53 immunohistochemistry and *TP53* mutations were included in the systematic review. Exclusion criteria were: sample size < 5 cases; reviews; case reports. Studies not allowing comparisons between immunohistochemistry and molecular analysis were excluded. In case of overlapping data between two studies (i.e. same institution and period of enrollment, same immunohistochemical and molecular findings), the study evaluating the smaller sample was excluded.

2.4. Risk of bias within studies assessment

The risk of bias within studies was assessed according to the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) [30]. For each study, 4 domains related to the risk of bias were assessed: 1) Patient selection (i.e. if the patients were consecutive or randomly selected); 2) Index test (i.e. if p53 immunohistochemistry was unbiased, e.g. blinded evaluation and clearly stated criteria to assess p53 expression); 3) Reference standard (i.e. if the methods for molecular analysis were unbiased, e.g. blinded evaluation and clearly stated criteria to assess *TP53* mutations); 4) Flow and Timing (i.e. if all patients were assessed with both index and reference standard; if all patients were assessed with the same tests, if the latency time between index and reference standard did not affect the results). Reviewer’s judgments were “low risk,” “unclear risk” or “high risk of bias” for each domain if data regarding the domain were “reported and adequate”, “reported but inadequate” and “not reported” respectively. Concerns about applicability were also assessed for the domains 1, 2 and 3 (i.e. if the criteria used are right but do not fit the objective of our study).

2.5. Data extraction

Original data were extracted without modification. Two by two contingency tables were created for each study, reporting two qualitative variables:

- immunohistochemical expression of p53 (index test), alternatively dichotomized as “normal expression” *vs* “overexpression”, and “normal expression” *vs* “overexpression or complete absence”;
- *TP53* mutational status (reference standard), dichotomized as “wild type” *vs* “mutated”.

For 3 studies that not defined a cut-off for p53 overexpression, data regarding the index test were extracted by using the following criteria:

- for the study that used two semiquantitative scale (0–4) for intensity and overall distribution of immunostaining, “overexpression” was considered for a score of at least 2 (intensity)/3 (distribution) or 3 (intensity)/2 (distribution) [17];
- for the study that used two semiquantitative scale for intensity (0–3) and overall distribution (0–4) of immunostaining, “overexpression” was considered for a score of at least 6 [18];
- for the study that classified overexpression of p53 as diffuse (100 % of nuclei) or focal (30 % of nuclei), only a diffuse immunostaining was considered as “overexpression” [19].

Absent immunohistochemical expression of p53 was defined as complete absence of p53 nuclear staining in the tumor cells for studies adopting a qualitative immunohistochemical evaluation, or as an immunostaining score of 0 for studies adopting a quantitative or semiquantitative immunostaining score of p53 expression.

2.6. Data analysis

TP53-mt cancers with aberrant p53 expression were considered as true positive; *TP53*-wt cancers with normal expression of p53 were considered as true negative; *TP53*-wt cancers with aberrant p53 expression were considered as false positive; *TP53*-mt cancers with normal expression of p53 were considered as false negative.

Sensitivity, specificity, positive and negative likelihood ratios (LR + and LR-) and diagnostic odds ratio (DOR) were calculated for each study and as pooled estimate. Values were reported graphically on forest plots with 95 % confidence interval (CI).

Post-test probabilities of mutated and wild type *TP53* mutational status were calculated and graphically reported using a Fagan’s nomogram with 95 % CI, for both “overexpression” and “overexpression or complete absence” p53. The pre-test probability (prevalence of TP53 mutations in EC) of 23 % derived from TGCA results [8].

Statistical heterogeneity amongst the included studies was assessed using the Higgins I^2 statistic; heterogeneity was classified as null for $I^2 = 0\%$, minimal for $0\% < I^2 \leq 25$ %, low for $25 < I^2 \leq 50$ %, moderate for $50 < I^2 \leq 75$ % and high for $I^2 > 75$ %, as previously described [31].

The random effect model of DerSimonian and Laird was adopted independently from the heterogeneity, as recommended for meta-analysis of diagnostic accuracy by the SEDATE guidelines [29].

Area under the curve (AUC) was calculated on summary receiver operating characteristic (SROC) curves. The diagnostic usefulness was categorized as absent for $AUC \leq 0.5$, low for $0.5 < AUC \leq 0.75$, moderate for $0.75 < AUC \leq 0.9$, high for $0.9 < AUC < 0.97$, very

high for AUC ≥ 0.97.

Additional analysis was performed separating data into 4 subgroups based on the criteria adopted to define an aberrant immunohistochemical expression of p53 ("overexpression" *vs* "overexpression or complete absence") and the sequencing techniques adopted to diagnose *TP53* mutations (Polymerase Chain Reaction, PCR + sequencing *vs* Next Generation Sequencing, NGS), as a higher sensitivity in detecting mutations has been reported for NGS, when compared to older techniques [32]. Sensitivity, specificity, LR+, LR-, DOR, AUC on SROC curves, and post-test probabilities were calculated for each subgroup.

The data analysis was performed using Review Manager 5.3 (Copenhagen: The Nordic Cochrane Centre, Cochrane Collaboration, 2014) and Meta-DiSc version 1.4 (Clinical Biostatistics Unit, Ramon y Cajal Hospital, Madrid, Spain).

3. Results

3.1. Study selection

4958 articles were identified through database search. 785 articles remained after duplicate removal. 133 articles remained after titles screening. 54 articles were evaluated for eligibility after abstracts screening. Finally, 13 observational studies with 727 patients were included in the systematic review [2,6,16–26]. The whole process of study selection is reported in detail in Supplementary Fig. 1.

3.2. Study characteristics

Most EC (65.9 %) were endometrioid adenocarcinoma, while other histotypes were: serous (17.8 %), clear cell (7.9 %), mixed (4.9 %), undifferentiated (3.2 %), carcinosarcoma (0.3 %). Grading was 1 in 30 % of EC, 2 in 28.8 %, 3 in 41 %.

Histologic specimens were obtained by hysterectomy in 7 studies and by biopsy in only 1 patient in 1 study, while in 6 studies the sampling method was unreported. DNA or RNA was extracted from paraffin-embedded tissue in 6 studies, and from fresh frozen tissue or paraffin-embedded tissue in 1 study, while the former was not reported in 6 studies. Molecular analysis included polymerase chain reaction (PCR) and exon sequencing in 6 studies, retro-polymerase chain reaction (RT-PCR) and exon sequencing in 1 study, and next generation sequencing (NGS) in 5 studies. Single-strand conformation polymorphism (SSCP) was performed as a screening test in 1 study. After PCR or RT-PCR, exons sequenced were 5–8 in all studies; in 2 studies, also exons 4 [21] and 11 [23] were sequenced.

Characteristics of the included studies are shown in detail in Supplementary Tables 1, 2 and 3.

3.3. Risk of bias within studies assessment

Regarding the risk of bias within studies assessment, for the "patient selection" domain, 9 studies were classified at unclear risk of bias because they did not report if the patients were consecutive or randomly

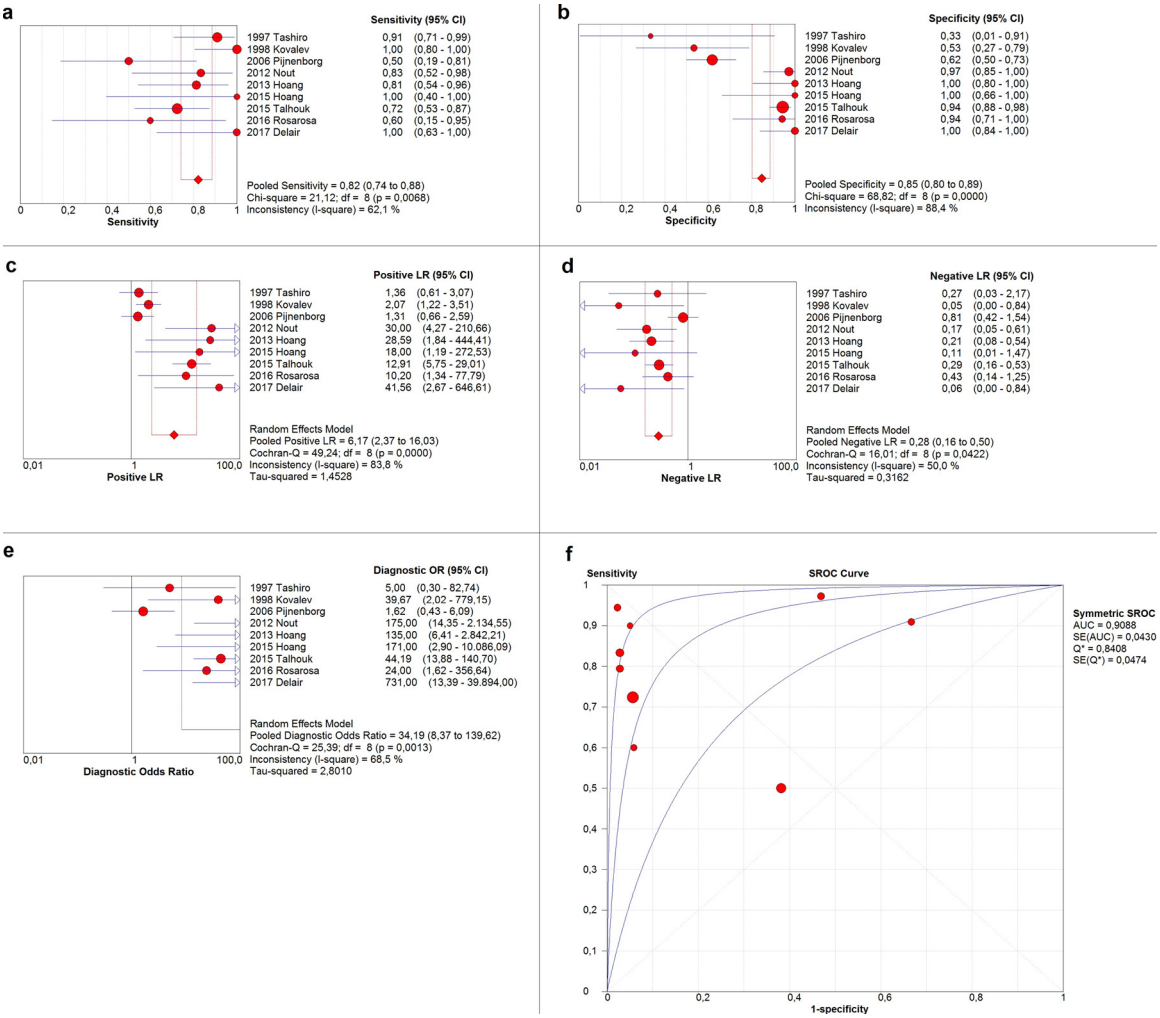

Fig. 1. Forest plots of individual studies and pooled sensitivity (a), specificity (b), positive and negative likelihood ratios (c and d), diagnostic odds ratio (e) with SROC curves (f) of p53 "overexpression" as surrogate of *TP53* mutations.

selected. Concerns were considered unclear for 10 studies, given that patients selection was restricted to: serous EC [18,19,22,24], serous and mixed EC [6], endometrioid EC [20,23], undifferentiated EC [25], clear cell EC [26]. Moreover, it was impossible to exclude overlapping data between two studies for 4 patients [6,24].

For the "index test" domain, 7 studies were considered at unclear risk of bias because a blinded evaluation of p53 immunohistochemistry was not reported [6,17,19,20,22,24,25].

For the "reference standard" domain, 7 studies were considered at unclear risk of bias because a blinded evaluation of *TP53* mutations was not reported [6,19,21,22,24–26]. Three studies were categorized a high risk of bias because 1 study performed molecular analysis for *TP53* mutations in only 8 of 107 cases [16], 1 in only cases with a score of at least 1 for both intensity and overall distribution of immunostaining [17], and 1 in only cases with aberrant p53 immunohistochemistry [20].

For the "flow and timing" domain, 3 studies were considered at high risk of bias, because not all patients were assessed with both index and reference standard [16,17,20].

All the remaining judgments were "low risk of bias".

Results of risk of bias among studies assessment are graphically reported in Supplementary Fig. 2.

3.4. Diagnostic accuracy analysis

Of 13 studies included in the systematic review, 4 studies were excluded from meta-analysis of diagnostic accuracy because 3 showed high risk of bias in two domains [16,17,20] and 1 analyzed only case with *TP53* mutations [22].

In the analysis of diagnostic accuracy as a surrogate of *TP53* mutations, p53 "overexpression" showed a pooled sensitivity of 0.82 (95 % CI, 0.74 − 0.88), with moderate heterogeneity among studies (I^2 = 62.1 %). Pooled specificity was 0.85 (95 % CI, 0.80 − 0.89) with high heterogeneity (I^2 = 88.4 %). Pooled positive and negative likelihood ratios were 6.17 (95 % CI, 2.37–16.03) and 0.28 (95 % CI, 0.16 − 0.50) respectively, with high heterogeneity (I^2 = 83.8 %) and low heterogeneity (I^2 = 50 %) respectively. Pooled DOR was 34.19 (95 % CI, 8.37–139.62), with high heterogeneity (I^2 = 68.5 %). The overall diagnostic accuracy was high, with an AUC of 0.9088 (Fig. 1). In the case of a positive test (p53 overexpression at immunohistochemistry), the post-test probability of TP53 mutations was 65 % (95 % CI, 58–71 %), while in the case of a negative test (p53 normal expression at immunohistochemistry), the post-test probability was 8% (95 % CI, 6–10 %) (Supplementary Fig. 3a)

On the other hand, "overexpression or complete absence" of p53 showed a pooled sensitivity of 0.85 (95 % CI, 0.77 − 0.91), with high heterogeneity among studies (I^2 = 69 %). Pooled specificity was 0.78 (95 % CI, 0.72-0.83) with high heterogeneity (I^2 = 91.2 %). Pooled positive and negative likelihood ratios were 3.75 (95 % CI, 1.54–9.09) and 0.27 (95 % CI, 0.14 − 0.54) respectively, with high heterogeneity (I^2 = 90.6 %) and low heterogeneity (I^2 = 48.1 %) respectively. Pooled DOR was 21.65 (95 % CI, 4.87–96.22), with high heterogeneity

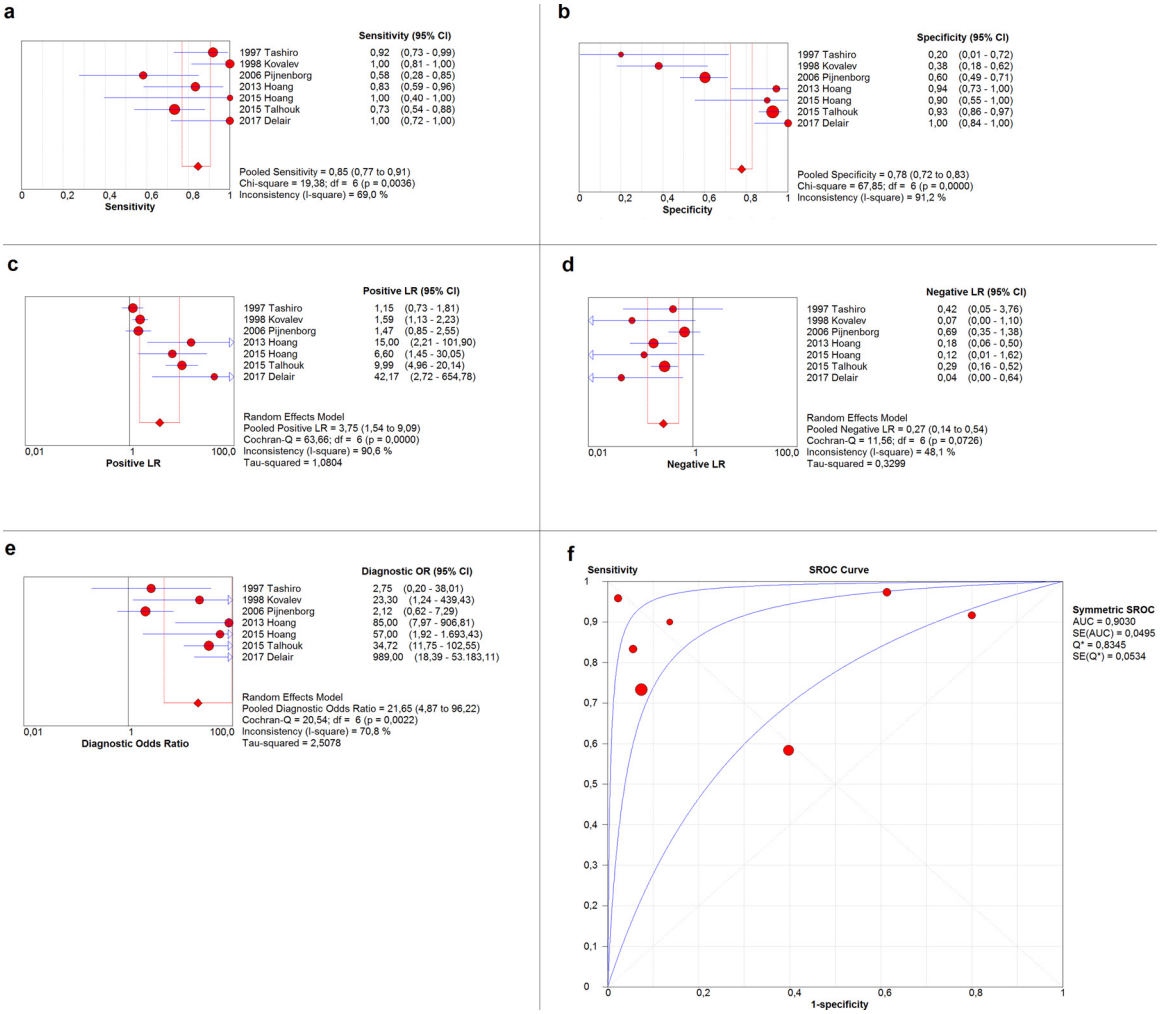

Fig. 2. Forest plots of individual studies and pooled sensitivity (a), specificity (b), positive and negative likelihood ratios (c and d), diagnostic odds ratio (e) with SROC curves (f) of "overexpression or complete absence" of p53 as surrogate of TP53 mutations.

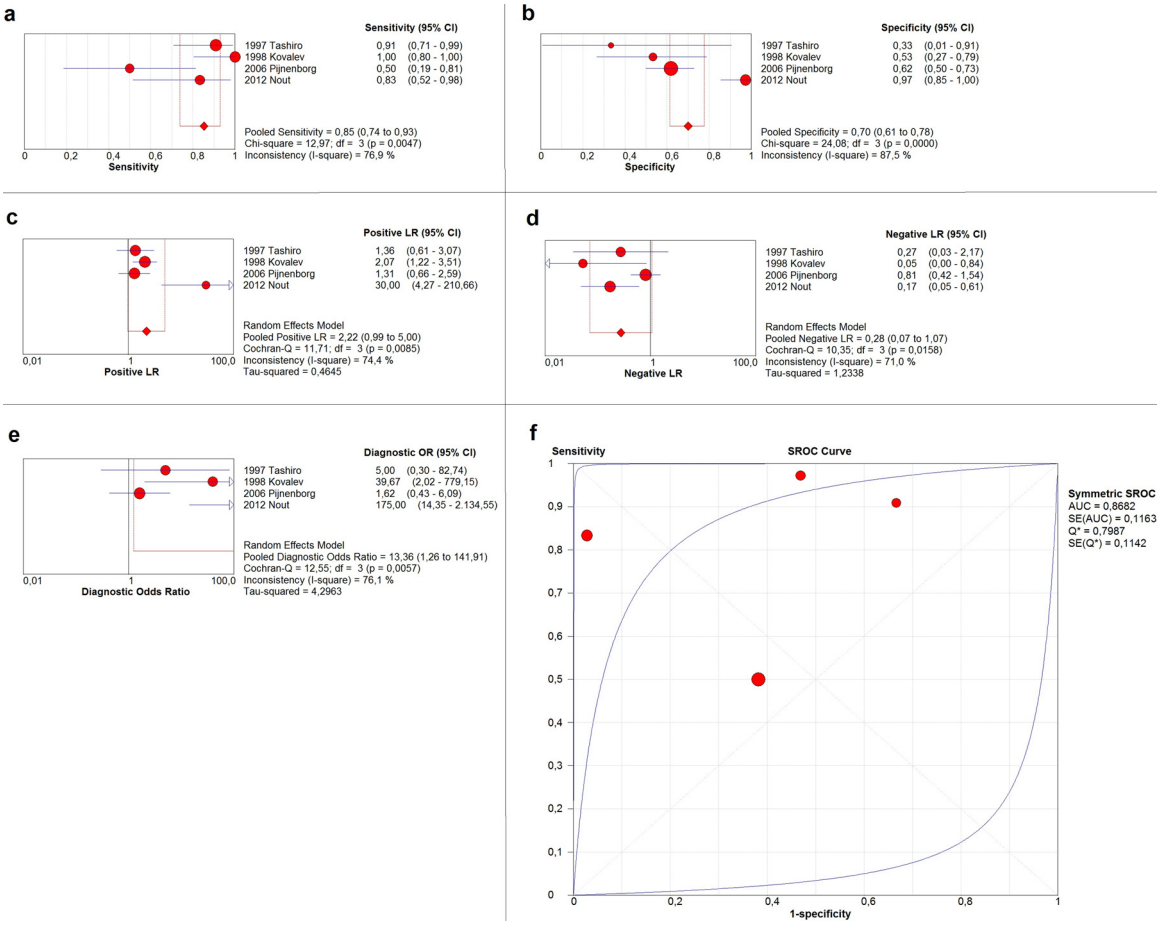

Fig. 3. Forest plots of individual studies and pooled sensitivity (a), specificity (b), positive and negative likelihood ratios (c and d), diagnostic odds ratio (e) with SROC curves (f) of p53 "overexpression" as surrogate of TP53 mutations detected by PCR + sequencing (subgroup 1).

(I^2 = 70.8 %). The overall diagnostic accuracy was high, with an AUC of 0.9030 (Fig. 2). In the case of a positive test (p53 overexpression or complete absence at immunohistochemistry), the post-test probability of *TP53* mutations was 53 % (95 % CI, 47 %–59 %), while in the case of a negative test (p53 normal expression at immunohistochemistry) the post-test probability was 7% (95 % CI, 5–11 %) (Supplementary Fig. 3b).

3.5. Additional analysis

With regard to subgroups analysis, 4 studies were included in the subgroup 1 ("overexpression" and PCR + sequencing) [18,19,21,23], 5 in subgroup 2 ("overexpression" and NGS) [2,6,24–26], 3 in subgroup 3 ("overexpression or complete absence" of p53 and PCR + sequencing) [18,19,21], and 4 in subgroup 4 ("overexpression or complete absence" of p53 and NGS) [2,6,24,26].

In subgroup 1, p53 "overexpression" showed a pooled sensitivity of 0.85 (95 % CI, 0.74 − 0.93) as surrogate of *TP53* mutations detected by PCR + sequencing, with high heterogeneity among studies (I^2 = 76.9 %). Pooled specificity was 0.70 (95 % CI, 0.61 − 0.78) with high heterogeneity (I^2 = 87.5 %). Pooled positive and negative likelihood ratios were 2.22 (95 % CI, 0.99–5.00) and 0.28 (95 % CI, 0.07–1.07) respectively, with moderate heterogeneity (I^2 = 74.4 %) and moderate heterogeneity (I^2 = 71 %) respectively. Pooled DOR was 13.36 (95 % CI, 1.26–141.91), with high heterogeneity (I^2 = 76.1 %). The overall diagnostic accuracy was moderate, with an AUC of 0.8682 (Fig. 3). In the case of a positive test (p53 overexpression at immunohistochemistry), the post-test probability of TP53 mutations was 40 % (95 % CI, 37–43 %), while in the case of a negative test (p53 normal expression at immunohistochemistry), the post-test probability was 8% (95 % CI, 6–10 %) (Supplementary Fig. 3c).

In subgroup 2, p53 "overexpression" showed a pooled sensitivity of 0.79 (95 % CI, 0.67 − 0.88) as surrogate of *TP53* mutations detected by NGS, with low heterogeneity among studies (I^2 = 45.5 %). Pooled specificity was 0.96 (95 % CI, 0.92 − 0.98) with minimal heterogeneity (I^2 = 13.3 %). Pooled positive and negative likelihood ratios were 14.48 (95 % CI, 7.35–28.55) and 0.27 (95 % CI, 0.17 − 0.42) respectively, with null heterogeneity (I^2 = 0%) and null heterogeneity (I^2 = 0%) respectively. Pooled DOR was 57.51 (95 % CI, 22.29–148.37), with null heterogeneity (I^2 = 0%). The overall diagnostic accuracy was very high, with an AUC of 0.9927 (Fig. 4). In the case of a positive test (p53 overexpression at immunohistochemistry), the post-test probability of TP53 mutations was 81 % (95 % CI, 77 %–86 %), while in the case of a negative test (p53 normal expression at immunohistochemistry), the post-test probability was 7% (95 % CI, 6–9 %) (Supplementary Fig. 3d).

In subgroup 3, "overexpression or complete absence" of p53 showed a pooled sensitivity of 0.87 (95 % CI, 0.75 − 0.95) as surrogate of *TP53* mutations detected by PCR + sequencing, with high heterogeneity among studies (I^2 = 82.7 %). Pooled specificity was 0.54 (95 % CI, 0.44 − 0.64) with moderate heterogeneity (I^2 = 65.6 %). Pooled positive and negative likelihood ratios were 1.42 (95 % CI, 1.11–1.82) and 0.42 (95 % CI, 0.12–1.50) respectively, with null heterogeneity (I^2 = 0%) and low heterogeneity (I^2 = 39.1 %) respectively. Pooled DOR was 3.31 (95 % CI, 0.95–11.56), with minimal heterogeneity (I^2 = 16.1 %). The overall diagnostic accuracy was low, with an AUC of 0.6248 (Fig. 5). In the case of a positive test (p53 overexpression or complete absence at immunohistochemistry), the post-test probability

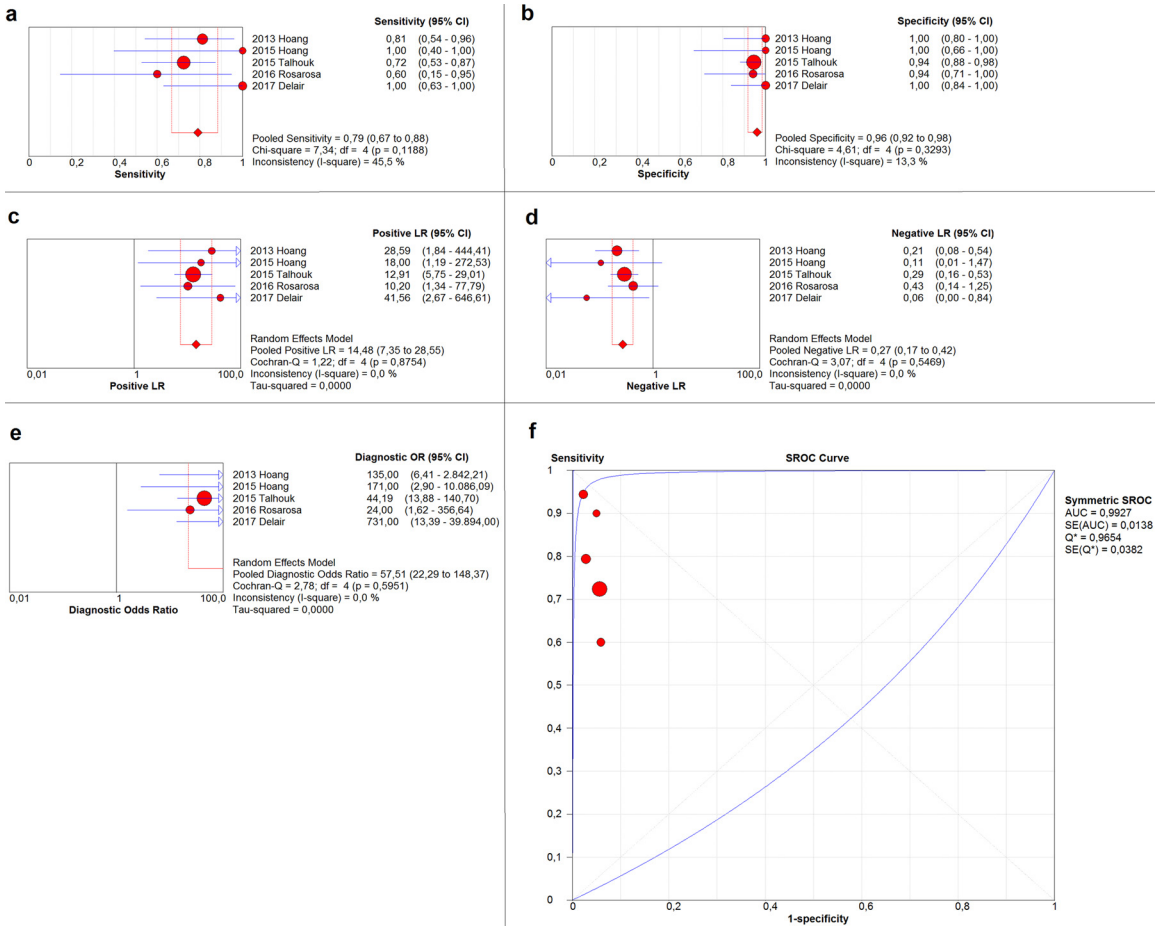

Fig. 4. Forest plots of individual studies and pooled sensitivity (a), specificity (b), positive and negative likelihood ratios (c and d), diagnostic odds ratio (e) with SROC curves (f) of p53 "overexpression" as surrogate of TP53 mutations detected by Next Generation Sequencing (subgroup 2).

of TP53 mutations was 30 % (95 % CI, 28–32 %), while in the case of a negative test (p53 normal expression at immunohistochemistry), the post-test probability was 11 % (95 % CI, 8–14 %) (Supplementary Fig. 3e).

In subgroup 4, "overexpression or complete absence" of p53 showed a pooled sensitivity of 0.83 (95 % CI, 0.71 − 0.91) as surrogate of *TP53* mutations detected by NGS, with high heterogeneity among studies (I^2 = 76.9 %). Pooled specificity was 0.94 (95 % CI, 0.89 − 0.97) with minimal heterogeneity (I^2 = 4.4 %). Pooled positive and negative likelihood ratios were 10.42 (95 % CI, 5.78–18.80) and 0.23 (95 % CI, 0.14 − 0.38) respectively, with null heterogeneity (I^2 = 0%) and null heterogeneity (I^2 = 0%) respectively. Pooled DOR was 49.29 (95 % CI, 19.64–123.71), with null heterogeneity (I^2 = 0%). The overall diagnostic accuracy was high, with an AUC of 0.9688 (Fig. 6). In the case of a positive test (p53 overexpression or complete absence at immunohistochemistry), the post-test probability of TP53 mutations was 75 % (95 % CI, 71–80 %), while in the case of a negative test (p53 normal expression at immunohistochemistry), the post-test probability was 6% (95 % CI, 5–8%) (Supplementary Fig. 3f).

4. Discussion

4.1. Main findings and interpretation

Our study showed that both "overexpression" and "overexpression or complete absence" of p53 were highly accurate immunohistochemical surrogates of *TP53* mutations in EC, with an AUC of 0.9088 and 0.9030, respectively. AUC of both "overexpression" and "overexpression or complete absence" of p53 further increased adopting only NGS for detecting *TP53* mutation as reference standard (AUC of 0.9927 and 0.9688, respectively). In particular, diagnostic accuracy became very high for "overexpression" of p53.

The possibility of using immunohistochemistry to predict genetic alterations is a long-standing issue [33–43]. In the case of EC, this issue has become even more of interest after the TCGA findings, which is expected to revolutionize the risk stratification in EC [1,8,44]. To date, management of patients with EC is still linked to post-surgical staging histologic examination (principally histotype, tumor grade and stage) [45–53]. However, histotype and grade of EC have shown poor reproducibility, even when evaluated by expert pathologists [5,6,54,55]. The poor reproducibility in histologic examination seems to regard the endometrium even more than other tissues [56–60]. The Proactive Molecular Risk Classifier for Endometrial Cancer (ProMisE) is a validate classifier which assigns EC specimens to one of four prognostic subgroups reflecting the four TCGA subgroups [2,7]. The ProMisE adopts the immunohistochemical expression of mismatch repair proteins as a surrogate of microsatellite status and p53 expression as a surrogate of copy-number status, allowing a wider applicability of the TCGA classification in the common practice [2,7].

At this point a new problem takes over: to define the accuracy and interpretation criteria of these surrogate immunohistochemical markers. In fact, an aberrant p53 immunohistochemistry has long since been defined as p53 overexpression [16–23]; however, different semiquantitative scores or qualitative interpretations of immunostaining have been used to classify p53 as overexpressed [2,6,16–26]. More recently, it has emerged that other less common patterns of p53 immunostaining have an underlying p53 mutation, i.e. complete negativity and cytoplasmic expression [6,24–27]. In the common practice,

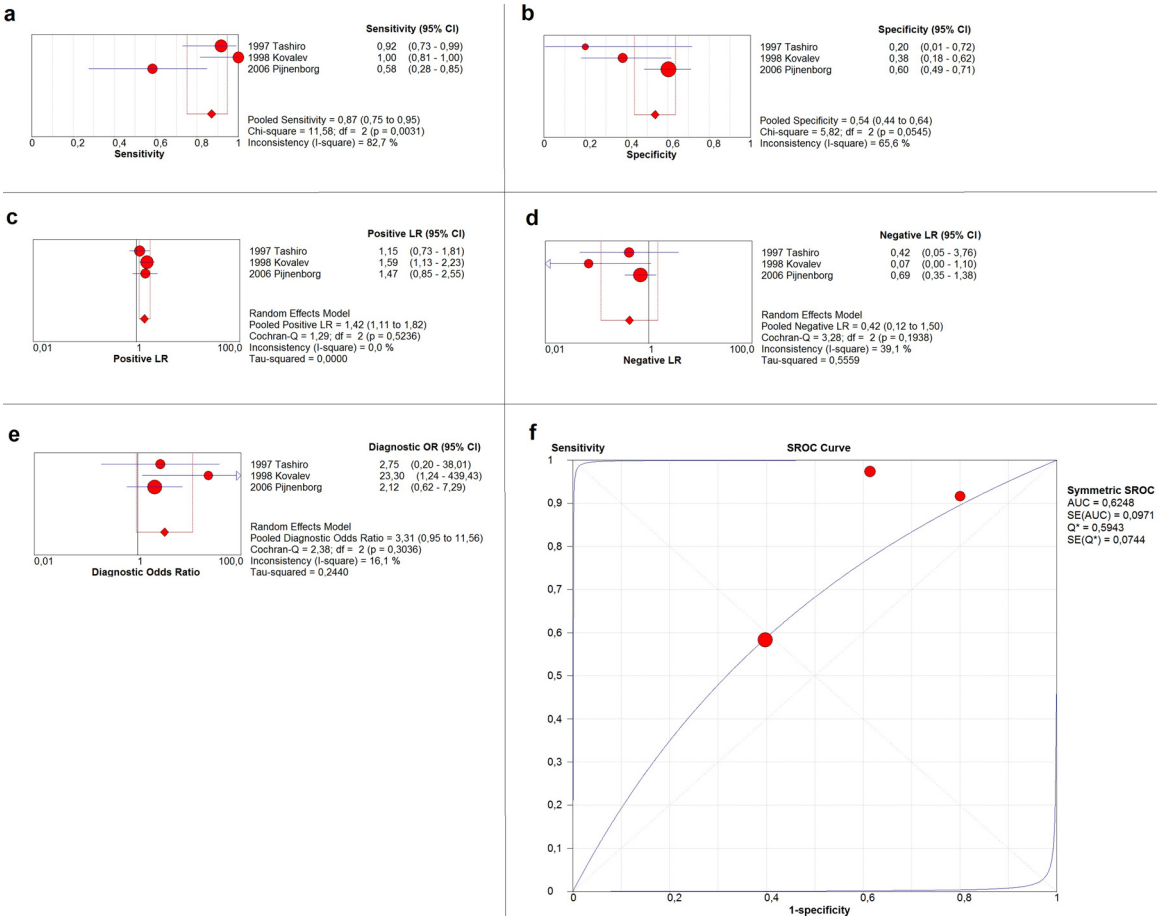

Fig. 5. Forest plots of individual studies and pooled sensitivity (a), specificity (b), positive and negative likelihood ratios (c and d), diagnostic odds ratio (e) with SROC curves (f) of "overexpression or complete absence" of p53 as surrogate of TP53 mutations detected by PCR + sequencing (subgroup 3).

the new problem about univocal immunostaining interpretation should be resolved in order to introduce immunohistochemical surrogates.

Although we found a high diagnostic accuracy as surrogate of *TP53* mutations for both "overexpression" and "overexpression or complete absence" of p53, the statistical heterogeneity among studies was at least moderate for most accuracy parameters in the overall analysis. We considered that such heterogeneity could be due to the differences in the thresholds of p53 expression adopted as index test and in the molecular technique adopted as reference standard. Since NGS is more sensitive and reproducible in detecting *TP53* mutations than previous sequencing techniques [32], we performed a subgroups analysis based on whether NGS was adopted. As expected, the heterogeneity decreased to minimal/null in the subgroups that adopted NGS as reference standard, while the AUC increased for both "overexpression" and "overexpression or complete absence" of p53. These more recent studies also adopted higher thresholds of p53 expression to define "overexpression"; indeed, they made a diagnosis of p53 overexpression only in the case of strong nuclear staining in $\geq 70-80\%$ of tumor cells [2,6,24–26]. This finding is in accordance with the recent evidence that p53 may be positive in most tumor cells without implying a *TP53* mutation; in these cases, the alternation of weak and strong intensity should favor a diagnosis of wild-type expression [27]. On the other hand, a diagnosis of p53 overexpression should be based on a p53 positivity in almost all tumor cell nuclei with consistently strong intensity; such pattern is found in about two-thirds of *TP53*-mutant tubo-ovarian serous carcinomas [27].

Unexpectedly, the accuracy appeared slightly higher for p53 "overexpression" than for "overexpression or complete absence". This would be in contrast with the evidence that a completely negative p53 staining reflects an underlying *TP53* mutation, being found in about one fourth of *TP53*-mutant tubo-ovarian serous carcinomas [27]. Possible explanations for such a finding might be that a correct interpretation of p53 immunohistochemistry needs an optimized immunohistochemical procedure [27]. In fact, the boundary between a wild-type expression and a complete absence of p53 might be thin, especially tumors that are weakly proliferative [6,27]. Furthermore, technical artefacts and failure to assess positive internal controls (such as stromal cells and lymphocytes) might lead to an erroneous diagnosis of complete p53 absence [27]. We might hypothesize that, in the common practice, the wrong application of the "complete absence" criterion might decrease the accuracy in predicting *TP53* mutations. Further research may be necessary to clarify this point.

A more recently described abnormal p53 pattern is cytoplasmic expression, which is found in about 4% of *TP53*-mutant tubo-ovarian serous carcinomas [27]. Such pattern was not described in the studies included in this meta-analysis, probably due to its rarity. A diagnosis of cytoplasmic p53 pattern would require the presence of a definite cytoplasmic staining, while an equivocal blush should be ignored [27]. Therefore, the correct interpretation of such pattern is strongly dependent on the immunohistochemical protocol adopted, which should not be too weak [27].

4.2. Strengths and limitations

To the best of our knowledge, this may be the first systematic review and meta-analysis on this topic. We calculated the diagnostic accuracy of p53 immunohistochemistry as surrogate of *TP53* mutations, comparing different criteria to define aberrant p53 immunohistochemistry.

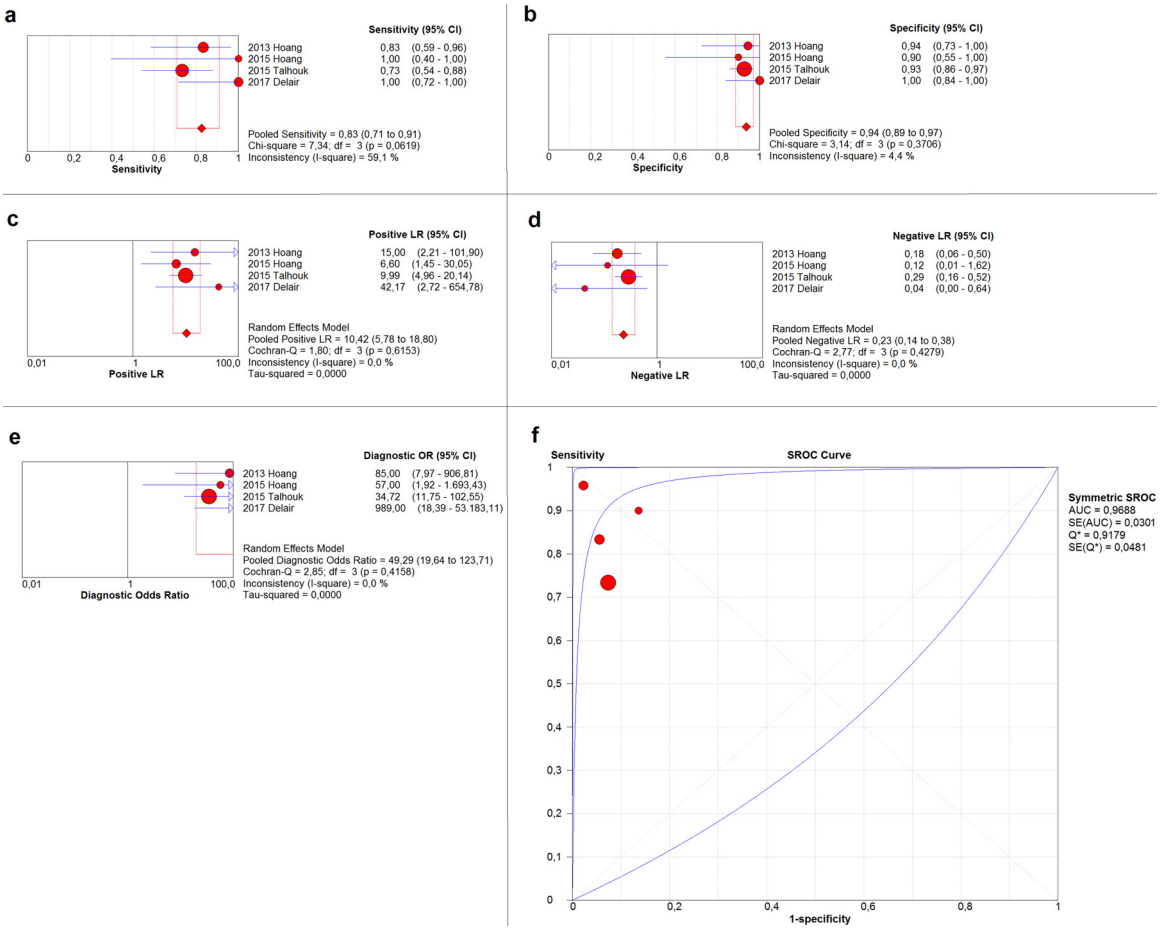

Fig. 6. Forest plots of individual studies and pooled sensitivity (a), specificity (b), positive and negative likelihood ratios (c and d), diagnostic odds ratio (e) with SROC curves (f) of "overexpression or complete absence" of p53 as surrogate of TP53 mutations detected by Next Generation Sequencing (subgroup 4).

A limitation of the overall analysis might be the different criteria used to define p53 overexpression in the included studies. Nonetheless, all studies that adopted NGS and showed the best results used similar criteria to define p53 overexpression (i.e. diffuse strong nuclear staining in $\geq 70-80\%$ of the tumor cells) and showed low statistical heterogeneity, resulting in accordance with recent evidence in this field [27]. In order to refine the accuracy of p53 immunohistochemistry, further studies on this topic should adopt the recommended optimized immunohistochemical procedures and consider also the less common pattern of p53 expression, such as cytoplasmic expression of the protein [27].

5. Conclusion

Aberrant expression of p53 is a highly accurate immunohistochemical surrogate of *TP53* mutations. Overexpression of p53 (i.e. strong nuclear positivity in $\geq 70-80\%$ of tumor cell nuclei) showed the best accuracy in predicting *TP53* mutations, especially using NGS a reference standard.

Further studies in this field should adopt optimized immunohistochemical procedures, particularly to allow a correct interpretation of complete p53 negativity. Furthermore, less common staining patterns, such as cytoplasmic expression, should also be taken into account for a more thorough definition of the accuracy of p53 immunohistochemistry.

Hopefully, the refinement of p53 immunohistochemistry will lead to a great improvement in the risk stratification of EC in the common practice.

Author contribution

AR and AT independently assessed electronic search, eligibility of the studies, inclusion criteria, risk of bias, data extraction and data analysis. Disagreements were resolved by discussion among all authors (AR, AT, MC, CDL, ADM, MDM, MCT, LI, FZ). AR, AT, MC, CDL, LI and FZ conceived the study; MC, CDL, ADM, MDM, MCT, LI and FZ worked on the design of the study; AR, AT, MC, CDL, ADM, MDM and MCT worked on the manuscript preparation; AT, AR and DR worked on the manuscript revision; LI and FZ supervised the whole study.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent

Not applicable.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the

online version, at doi:https://doi.org/10.1016/j.prp.2020.153025.

References

[1] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2015, CA Cancer J. Clin. 65 (1) (2015) 5–29 Jan-Feb.

[2] A. Talhouk, M.K. McConechy, S. Leung, et al., A clinically applicable molecular-based classification for endometrial cancers, Br. J. Cancer 113 (2) (2015) 299–310.

[3] P. Giampaolino, A. Di Spiezio Sardo, A. Mollo, et al., Hysteroscopic endometrial focal resection followed by levonorgestrel intrauterine device insertion as a fertility-sparing treatment of atypical endometrial hyperplasia and early endometrial Cancer: a retrospective study, J. Minim. Invasive Gynecol. 26 (4) (2019) 648–656 May - Jun.

[4] A. Raffone, A. Travaglino, G. Saccone, et al., Management of women with atypical polypoid adenomyoma of the uterus: a quantitative systematic review, Acta Obstet. Gynecol. Scand. (2019), https://doi.org/10.1111/aogs.13553 Feb 3.

[5] C.B. Gilks, E. Oliva, R.A. Soslow, Poor interobserver reproducibility in the diagnosis of high-grade endometrial carcinoma, Am. J. Surg. Pathol. 37 (2013) 874–881.

[6] L.N. Hoang, M.K. McConechy, M. Kobel, et al., Histotype-genotype correlation in 36 high-grade endometrial carcinomas, Am. J. Surg. Pathol. 37 (2013) 1421–1432.

[7] A. Talhouk, M.K. McConechy, S. Leung, et al., Confirmation of ProMisE: a simple, genomics-based clinical classifier for endometrial cancer, Cancer 123 (5) (2017) 802–813.

[8] Cancer Genome Atlas Research Network, et al., Integrated genomic characterization of endometrial carcinoma, Nature 497 (7447) (2013) 67–73.

[9] G. Kim, K.C. Kurnit, B. Djordjevic, et al., Nuclear β-catenin localization and mutation of the CTNNB1 gene: a context-dependent association, Mod. Pathol. (2018).

[10] A. Travaglino, A. Raffone, G. Saccone, et al., Immunophenotype of atypical polypoid adenomyoma of the uterus: diagnostic value and insight on pathogenesis, Appl. Immunohistochem. Mol. Morphol. (2019), https://doi.org/10.1097/PAI.0000000000000780 Jun 6.

[11] A. Raffone, A. Travaglino, M. Mascolo, L. Insabato, F. Zullo, Predictive accuracy of hormone receptors in conservatively treated endometrial hyperplasia and early endometrioid carcinoma, Acta Obstet. Gynecol. Scand. 99 (1) (2020) 140 Jan.

[12] A. Raffone, A. Travaglino, A. D'Antonio, et al., Immunohistochemistry for BAG3 in cervical precancerous lesions, Acta Obstet. Gynecol. Scand. (2019), https://doi.org/10.1111/aogs.13754 Oct 20 [Epub ahead of print].

[13] A. Travaglino, A. Raffone, M. Mascolo, et al., TCGA molecular subgroups in endometrial Undifferentiated/Dedifferentiated carcinoma, Pathol. Oncol. Res. (2019), https://doi.org/10.1007/s12253-019-00784-0 Dec 6 [Epub ahead of print].

[14] E. Stelloo, R.A. Nout, E.M. Osse, et al., Improved risk assessment by integrating molecular and clinicopathological factors in early-stage endometrial cancer-combined analysis of the PORTEC cohorts, Clin. Cancer Res. 22 (16) (2016) 4215–4224.

[15] E. Stelloo, T. Bosse, R.A. Nout, et al., Refining prognosis and identifying targetable pathways for high-risk endometrial cancer; a TransPORTEC initiative, Mod. Pathol. 28 (6) (2015) 836–844.

[16] M.F. Kohler, A. Berchuck, A.M. Davidoff, Overexpression and mutation of p53 in endometrial carcinoma, Cancer Res. 52 (6) (1992) 1622–1627 Mar 15.

[17] M.W. Jones, S. Kounelis, C. Hsu, Prognostic value of p53 and K-ras-2 topographic genotyping in endometrial carcinoma: a clinicopathologic and molecular comparison, Int. J. Gynecol. Pathol. 16 (4) (1997) 354–360 Oct.

[18] H. Tashiro, C. Isacson, R. Levine, p53 gene mutations are common in uterine serous carcinoma and occur early in their pathogenesis, Am. J. Pathol. 150 (1) (1997) 177–185 Jan.

[19] S. Kovalev, N.D. Marchenko, B.G. Gugliotta, Loss of p53 function in uterine papillary serous carcinoma, Hum. Pathol. 29 (6) (1998) 613–619 Jun.

[20] Y.Z. Feng, T. Shiozawa, A. Horiuchi, Intratumoral heterogeneous expression of p53 correlates with p53 mutation, Ki-67, and cyclin A expression in endometrioid-type endometrial adenocarcinomas, Virchows Arch. 447 (5) (2005) 816–822 Nov Epub 2005 Jul 14.

[21] J.M. Pijnenborg, L. van de Broek, G.C. Dam de Veen, TP53 overexpression in recurrent endometrial carcinoma, Gynecol. Oncol. 100 (2) (2006) 397–404 Feb Epub 2005 Nov 4.

[22] L. Jia, Y. Liu, X. Yi, Endometrial glandular dysplasia with frequent p53 gene mutation: a genetic evidence supporting its precancer nature for endometrial serous carcinoma, Clin. Cancer Res. 14 (8) (2008) 2263–2269. Apr 15;.

[23] R.A. Nout, T. Bosse, C.L. Creutzberg, Improved risk assessment of endometrial cancer by combined analysis of MSI, PI3K-AKT, Wnt/β-catenin and P53 pathway activation, Gynecol. Oncol. 126 (3) (2012) 466–473, https://doi.org/10.1016/j.ygyno.2012.05.012 Sep Epub 2012 May 15.

[24] L.N. Hoang, M.K. McConechy, B. Meng, Targeted mutation analysis of endometrial clear cell carcinoma, Histopathology. 66 (5) (2015) 664–674. Apr.

[25] J.M. Rosa-Rosa, S.1 Leskelä, E. Cristóbal-Lana, Molecular genetic heterogeneity in undifferentiated endometrial carcinomas, Mod. Pathol. 29 (11) (2016) 1390–1398 Nov.

[26] D.F. DeLair, K.A.1 Burke, P. Selenica, The genetic landscape of endometrial clear cell carcinomas, J. Pathol. 243 (2) (2017) 230–241 Oct.

[27] M. Köbel, B.M. Ronnett, N. Singh, R.A. Soslow, C.B. Gilks, W.G. McCluggage, Interpretation of P53 immunohistochemistry in endometrial carcinomas: toward increased reproducibility, Int. J. Gynecol. Pathol. (2018), https://doi.org/10.1097/PGP.0000000000000488 Mar 6 [Epub ahead of print].

[28] D. Moher, L. Shamseer, M. Clarke, et al., Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement, Syst. Rev. 4 (2015) 1.

[29] A. Sotiriadis, S.I. Papatheodorou, W.P. Martins, Synthesizing evidence from diagnostic accuracy tests: the SEDATE guideline, Ultrasound Obstet. Gynecol. 47 (3) (2016) 386–395.

[30] P.F. Whiting, A.W. Rutjes, M.E. Westwood, et al., QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies, Ann. Intern. Med. 155 (8) (2011) 529–536.

[31] A. Raffone, A. Travaglino, A. Santoro, et al., Accuracy of one-step nucleic acid amplification in detecting lymph node metastases in endometrial Cancer, Pathol. Oncol. Res. (2019), https://doi.org/10.1007/s12253-019-00727-9 Aug 23 [Epub ahead of print].

[32] S. Behjati, P.S. Tarpey, What is next generation sequencing? Arch. Dis. Child. Educ. Pract. Ed. 98 (6) (2013) 236–238 Dec.

[33] A. Raffone, A. Travaglino, G. Saccone, et al., Loss of PTEN expression as diagnostic marker of endometrial precancer: a systematic review and meta-analysis, Acta Obstet. Gynecol. Scand. 98 (3) (2019) 275–286 Mar.

[34] A. Travaglino, A. Raffone, G. Saccone, et al., PTEN immunohistochemistry in endometrial hyperplasia: which are the optimal criteria for the diagnosis of precancer? APMIS 127 (4) (2019) 161–169. Apr.

[35] A. Travaglino, A. Raffone, G. Saccone, et al., Immunohistochemical nuclear expression of β-Catenin as a surrogate of CTNNB1 exon 3 mutation in endometrial Cancer, Am. J. Clin. Pathol. 151 (5) (2019) 529–538. Apr 2.

[36] A. Raffone, A. Travaglino, G. Saccone, et al., PTEN expression in endometrial hyperplasia and risk of cancer: a systematic review and meta-analysis, Arch. Gynecol. Obstet. 299 (6) (2019) 1511–1524 Jun.

[37] A. Raffone, A. Travaglino, G. Saccone, et al., PAX2 in endometrial carcinogenesis and in differential diagnosis of endometrial hyperplasia. A systematic review and meta-analysis of diagnostic accuracy, Acta Obstet. Gynecol. Scand. 98 (3) (2019) 287–299 Mar.

[38] A. Travaglino, A. Raffone, G. Saccone, et al., Immunohistochemical predictive markers of response to conservative treatment of endometrial hyperplasia and early endometrial cancer: a systematic review, Acta Obstet. Gynecol. Scand. 98 (9) (2019) 1086–1099 Sep.

[39] A. Raffone, A. Travaglino, G. Saccone, et al., Should progesterone and estrogens receptors be assessed for predicting the response to conservative treatment of endometrial hyperplasia and cancer? A systematic review and meta-analysis, Acta Obstet. Gynecol. Scand. 98 (8) (2019) 976–987. Aug.

[40] A. Raffone, A. Travaglino, G. Saccone, et al., Diagnostic and prognostic value of ARID1A in endometrial hyperplasia: a novel marker of occult cancer, APMIS 127 (9) (2019) 597–606 Sep.

[41] A. Travaglino, A. Raffone, G. Saccone, et al., Nuclear expression of β-catenin in endometrial hyperplasia as marker of premalignancy, APMIS 127 (11) (2019) 699–709 Nov.

[42] A. Travaglino, A. Raffone, M. Mascolo, et al., Clear cell endometrial carcinoma and the TCGA classification, Histopathology 76 (2) (2020) 336–338 Jan.

[43] A. Raffone, A. Travaglino, A. D' Antonio, et al., BAG3 expression correlates with the grade of dysplasia in squamous intraepithelial lesions of the uterine cervix, Acta Obstet. Gynecol. Scand. 99 (1) (2020) 99–104 Jan.

[44] A. Raffone, A. Travaglino, M. Mascolo, et al., TCGA molecular groups of endometrial cancer: pooled data about prognosis, Gynecol. Oncol. (2019), https://doi.org/10.1016/j.ygyno.2019.08.019 Aug 29. pii: S0090-8258(19)31471-31474 [Epub ahead of print].

[45] C.L. Creutzberg, W.L. van Putten, P.C. Koper, et al., Surgery and postoperative radiotherapy versus surgery alone for patients with stage-1 endometrial carcinoma: multicentre randomised trial. PORTEC Study Group. Post Operative Radiation Therapy in Endometrial Carcinoma, Lancet 355 (9213) (2000) 1404–1411. Apr 22.

[46] J. Fanning, Long-term survival of intermediate risk endometrial cancer (stage IG3, IC, II) treated with full lymphadenectomy and brachytherapy without teletherapy, Gynecol. Oncol. 82 (2) (2001) 371–374. Aug.

[47] H.M. Keys, J.A. Roberts, V.L. Brunetto, et al., A phase III trial of surgery with or without adjunctive external pelvic radiation therapy in intermediate risk endometrial adenocarcinoma: a Gynecologic Oncology Group study, Gynecol. Oncol. 92 (3) (2004) 744–751 Mar.

[48] A. Mariani, S.C. Dowdy, W.A. Cliby, et al., Prospective assessment of lymphatic dissemination in endometrial cancer: a paradigm shift in surgical staging, Gynecol. Oncol. 109 (1) (2008) 11–18. Apr.

[49] J.S. Kwon, F. Qiu, R. Saskin, M.S. Carey, Are uterine risk factors more important than nodal status in predicting survival in endometrial cancer? Obstet. Gynecol. 114 (4) (2009) 736–743 Oct.

[50] N. Colombo, E. Preti, F. Landoni, et al., Endometrial cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up, Ann. Oncol. 24 (Suppl 6) (2013) vi33–vi38 Oct.

[51] M.M. AlHilli, A. Mariani, J.N. Bakkum-Gamez, et al., Risk-scoring models for individualized prediction of overall survival in low-grade and high-grade endometrial cancer, Gynecol. Oncol. 133 (3) (2014) 485–493 Jun.

[52] S. Bendifallah, G. Canlorbe, P. Collinet, et al., Just how accurate are the major risk stratification systems for early-stage endometrial cancer? Br. J. Cancer 112 (5) (2015) 793–801 Mar 3.

[53] T.W. Kong, S.J. Chang, J. Paek, Y. Lee, M. Chun, H.S. Ryu, Risk group criteria for tailoring adjuvant treatment in patients with endometrial cancer: a validation study of the Gynecologic Oncology Group criteria, J. Gynecol. Oncol. 26 (1) (2015) 32–39 Jan.

[54] H. Guan, A. Semaan, S. Bandyopadhyay, et al., Prognosis and reproducibility of new and existing binary grading systems for endometrial carcinoma compared to FIGO grading in hysterectomy specimens, Int. J. Gynecol. Cancer 21 (4) (2011) 654–660 May.

[55] G. Han, D. Sidhu, M.A. Duggan, et al., Reproducibility of histological cell type in high-grade endometrial carcinoma, Mod. Pathol. 26 (12) (2013) 1594–1604 Dec.

[56] W.G. McCluggage, My approach to the interpretation of endometrial biopsies and curettings, J. Clin. Pathol. 59 (8) (2006) 801–812.
[57] A. Travaglino, A. Raffone, G. Saccone, et al., Significant risk of occult cancer in complex non-atypical endometrial hyperplasia, Arch. Gynecol. Obstet. 300 (5) (2019) 1147–1154 Nov.
[58] A. Raffone, A. Travaglino, G. Saccone, et al., Endometrial hyperplasia and progression to cancer: which classification system stratifies the risk better? A systematic review and meta-analysis, Arch. Gynecol. Obstet. 299 (5) (2019) 1233–1242 May.
[59] A. Travaglino, A. Raffone, G. Saccone, et al., Complexity of glandular architecture should be reconsidered in the classification and management of endometrial hyperplasia, APMIS. 127 (6) (2019) 427–434 Jun.
[60] A. Travaglino, A. Raffone, G. Saccone, et al., Congruence Between 1994 WHO Classification of Endometrial Hyperplasia and Endometrial Intraepithelial Neoplasia System, Am. J. Clin. Pathol. 153 (1) (2020) 40–48 Jan 1.