Tissue microarray in prognostic studies on vulva cancer

Fons, G.

Citation for published version (APA):
Chapter 5

External validation of COX-2 expression in vulvar cancer

Guus Fons
Jacobs van der Velden
Matthé PM Burger
Ate GJ van der Zee
Harry Hollema
Fiebo JW ten Kate

Submitted
Abstract

Aim of the study
In previous studies from our group we found overexpression of COX-2 as determined by immunohistochemistry to be related to worse survival. However, reproducibility of immunohistochemical tests in general is limited and cut-off points as determined in the studied population may not be valuable in different independent populations. The aim of the current study was to assess the reproducibility of the association between overexpression of COX-2 and survival in an independent but similar study group of vulvar cancer patients from another hospital.

Methods
COX-2 staining patterns as determined on tissue micro array (TMA) with material of 116 vulvar cancer patients, treated at the University Medical Centre Groningen, were related to survival and compared with the COX-2 staining patterns on the TMAs with material of 126 patients, treated at the Academic Medical Centre in Amsterdam.

Results
Both groups were comparable concerning the size of tumor, stage of disease, presence of lymph node metastases, extra capsular spread, vascular space involvement, adjuvant radiotherapy and disease-specific and disease-free survival. Any association between disease-specific survival or disease-free survival and strong COX-2 expression could not be demonstrated for the cases from Groningen. In the group from Amsterdam strong COX-2 expression was again significantly associated with both disease-specific (HR 2.44, 95% CI 1.16-5.13, \( P=0.02 \)) and disease-free survival (HR 2.24, 95% CI 1.15-4.38, \( P=0.02 \)).

Conclusion
The association between strong COX-2 expression and survival in 126 vulvar cancer patients from Amsterdam could not be reproduced in an independent but similar study group of 116 patients from Groningen.
Introduction

Squamous cell cancer of the vulva is a rare disease that mainly affects elderly women. Tumor diameter, lymph node involvement and vascular space involvement are the most important prognostic factors for survival. (1-3) Nevertheless, a substantial number of cancer-related deaths occur in patients without these risk factors. Additional markers are needed to predict the outcome more accurately. The development of tissue micro array (TMA) gives us the opportunity to test the prognostic significance of protein markers on a large scale. In a previous study we found overexpression of COX-2 to be significantly associated with survival. (4) Although this association was confirmed by a validation study in a similar population of the same hospital (article in press), external validation still is imperative for two main reasons. First, the reproducibility of immunohistochemical tests in general is limited because of the lack of standardization in procedures. Second, since cut-off points are determined as the best values in the populations under study, it is possible that test results worsen as initial cut-off points are used in a different, independent population. (5) The objective of the current study was to assess the reproducibility of the association between overexpression of COX-2 and survival in an independent but similar population of vulvar cancer patients.

Materials and methods

Patients

A TMA with material of 298 vulvar cancer patients treated at the University Medical Centre Groningen between 1984 and 2001 was used for external validation. Staining patterns on this TMA were compared with the staining patterns on two TMAs with material from 50 and 80 patients respectively, treated at the Academic Medical Centre in Amsterdam between 1994 and 2003.

The data of all patients who were not treated according to the current standard protocol, consisting of radical local excision of the tumor with unilateral or bilateral lymphadenectomy, were excluded. Sixty-five patients were excluded because they were treated with primary radiotherapy or a lymph node dissection was omitted. Hundred-seventeen patients underwent an en bloc resection of the groin nodes, the data of these patients were not taken into account either. The data of 116 patients from Groningen were reviewed.

Four patients from Amsterdam were excluded having had a lymph node debulking instead of a lymphadenectomy. The data of the remaining 126 patients from Amsterdam were reviewed.
All patients were treated with curative intent. Patients with more than 1 nodal metastasis and/or extra capsular growth were treated with adjuvant external beam radiation on the groins and pelvis. Patient characteristics are shown in Table 1.

**Immunohistochemistry**

Formalin fixed paraffin-embedded tumor tissue samples from all patients were available for constructing TMAs. In both institutes the same construction-procedure of the TMA was followed. One representative Haematoxylin and Eosin (H&E) slide of each tumor was selected. Three representative areas of interest with infiltrative carcinoma were encircled on each slide. In the corresponding paraffin block, 0.6 mm cores were punched out. These cores, each 3-4 mm high, were then embedded in the donor block using a manually operated tissue micro array device (Beecher Instruments, Silver Springs, MD). The spacing between the cores was 1mm. The recipient block was sectioned at 4μm and transferred to glass slides. The glass slides of Groningen and Amsterdam were processed at the same time, using the avidin-biotin method for immunostaining. The sections were deparaffinized with xylol, and re-hydrated through series of graded alcohol and submitted to antigen retrieval by pressure-cooking for 10 minutes in Tres-EDTA buffer (pH 9.0). All the slides were incubated with the monoclonal COX-2 antibody (Cayman Chemical,160112) at a dilution of 1:800 overnight at 4°C.

**Immunohistochemical scoring**

Two observers scored the staining results (FJK, GF). Tumor cells were marked as negative (<10% of cells show weak staining or cells show no staining), weak positive (10-50% of cells show weak staining or <10% of cells show strong staining) and strong positive (>50% of cells show moderate staining or >10% of cells show strong staining). This scoring system is slightly different from the scoring system used in the first TMA study with the material of 50 patients with vulvar cancer. Originally the intensity of staining was not taken into account. In the previous validation study, in which the staining patterns on TMA and whole slides were compared, the agreement improved after introducing the level of intensity into the scoring system.

Score results of cores of 1 tumor were combined into 1 score. If the scores of 3 individual cores of 1 tumor differed, the one that occurred most often determined the final score. Non-assessable cores were either lost during processing or contained less than 10% tumor cells. If 2 of 3 cores were not assessable the case was excluded. Strong focal expression was considered positive: if one core showed strong expression and the other core did not, the case was considered positive. For statistical analysis scores were dichotomized,
by combining a negative and a weak score into a negative score. This strategy is in accordance with the strategy developed in the first TMA study. \(^{(6)}\)

**Table 1. Patient characteristics**

<table>
<thead>
<tr>
<th></th>
<th>Groningen N=116</th>
<th>Amsterdam N=126</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age median (min-max)</td>
<td>69 (34-93)(^1)</td>
<td>73 (37-92)</td>
<td>0.02</td>
</tr>
<tr>
<td>Follow-up months median (min – max)</td>
<td>53 (1-114)(^1)</td>
<td>42 (1-174)</td>
<td>0.21</td>
</tr>
<tr>
<td>Size of tumor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 4cm</td>
<td>89 (77)(^2)</td>
<td>93 (74)(^2)</td>
<td></td>
</tr>
<tr>
<td>&gt;=4 cm</td>
<td>26 (22)</td>
<td>33 (26)</td>
<td>0.52</td>
</tr>
<tr>
<td>Unknown</td>
<td>1 ( 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIGO</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IA</td>
<td>2 ( 2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IB</td>
<td>24 (21)</td>
<td>27 (21)</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>43 (37)</td>
<td>50 (40)</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>32 (27)</td>
<td>28 (22)</td>
<td></td>
</tr>
<tr>
<td>IVA</td>
<td>13 (11)</td>
<td>20 (16)</td>
<td></td>
</tr>
<tr>
<td>IVB</td>
<td>2 ( 2)</td>
<td>1 ( 1)</td>
<td>0.83</td>
</tr>
<tr>
<td>VSI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>93 (80)</td>
<td>101 (80)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>16 (14)</td>
<td>25 (20)</td>
<td>0.30</td>
</tr>
<tr>
<td>Unknown</td>
<td>7 ( 6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lymph node metastases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>70 (60)</td>
<td>81 (64)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>46 (40)</td>
<td>45 (36)</td>
<td>0.53</td>
</tr>
<tr>
<td>Extra capsular growth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>100 (86)</td>
<td>99 (79)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>15 (13)</td>
<td>27 (21)</td>
<td>0.09</td>
</tr>
<tr>
<td>Unknown</td>
<td>1 ( 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radiotherapy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>77 (66)</td>
<td>92 (73)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>39 (34)</td>
<td>34 (27)</td>
<td>0.26</td>
</tr>
<tr>
<td>Recurrence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>69 (60)</td>
<td>81 (64)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>44 (38)</td>
<td>45 (36)</td>
<td>0.22</td>
</tr>
<tr>
<td>Unknown</td>
<td>3 ( 2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) range; \(^2\) percentages
Statistical analysis

Outcome parameters were disease-free survival (DFS) and disease-specific survival (DSS). DFS was defined as the time in months from the date of surgery to the date of histological proven recurrent disease or, in the case of censored observation, the date of last contact. DSS was defined as the time in months from the date of surgery to the date of death, corrected for causes of death other than vulvar cancer. Survival is calculated with the Kaplan-Meier product limit method. Cox’s proportional hazards model was used for the analysis of the impact of COX-2 expression on survival both for the group of patients from Groningen and for the one from Amsterdam. Differences between the 2 groups regarding a continuous or qualitative variable were analyzed with the Mann-Whitney U test or chi-square test, respectively. Calculations were performed with SPSS for Windows (version 15.02, SPSS Inc, UK). P-values <0.05 were considered to be significant.

The COX-2 score results on the TMA of Amsterdam were compared with the score results of our previous TMA study. Kappa (κ) was used as a measure of agreement corrected for agreement by chance. (7)

Results

Patient characteristics

The data of 116 patients from Groningen and 126 patients from Amsterdam were evaluated. Both groups were comparable concerning the size of the tumor (< or ≥ 4 cm), FIGO stage of disease, presence of lymph node metastases, extra capsular spread, vascular space involvement (VSI) and adjuvant radiotherapy (Table 1.). Patients from Amsterdam were on average 4 years older than those from Groningen (P=0.02). Median follow-up was 53 months in Groningen and 42 months in Amsterdam.

In the group from Groningen DFS and DSS at 5 years were 62% and 80% respectively. In the group from Amsterdam the DFS and DSS at 5 years were 63% and 71% respectively. No significant differences were found regarding both the DFS and DSS in the groups from Groningen and from Amsterdam (P=0.79 and P=0.19, respectively).

Immunohistochemistry

On both the TMAs from Groningen and Amsterdam positive controls were positive and negative controls were negative. COX-2 expression score results are shown in Table 2. Four cases from Amsterdam and 18 cases from Groningen were excluded because of non-assessable cores.

Six of 98 cases (6%) on the TMA of Groningen showed strong positive expression, while 22 of 122 cases (18%) on the TMA of Amsterdam showed strong expression (P<0.001).
Table 2. COX-2 expression

<table>
<thead>
<tr>
<th>COX-2 expression</th>
<th>Amsterdam First staining N=126 N (%)</th>
<th>Amsterdam Second staining N=126 N (%)</th>
<th>Groningen N=116 N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>46 (37)</td>
<td>42 (33)</td>
<td>44 (38)</td>
</tr>
<tr>
<td>Weak</td>
<td>56 (44)</td>
<td>58 (46)</td>
<td>48 (41)</td>
</tr>
<tr>
<td>Strong</td>
<td>18 (14)</td>
<td>22 (18)</td>
<td>6 (5)</td>
</tr>
<tr>
<td>Non-assessable</td>
<td>6 (5)</td>
<td>4 (3)</td>
<td>18 (16)</td>
</tr>
</tbody>
</table>

The association between COX-2 score results and survival was calculated for both the group from Groningen and from Amsterdam. Any association between strong COX-2 expression, and DSS (\(P=0.49\)) or DFS (\(P=0.65\)) could not be demonstrated for the cases from Groningen. In the group from Amsterdam strong COX-2 expression was associated with both DSS (HR=2.44, 95% CI 1.16-5.13, \(P=0.02\)) and DFS (HR 2.24, 95% CI 1.15-4.38, \(P=0.02\)).

COX-2 score results on the TMA from Amsterdam were compared with the initial score results on the same TMA. An absolute agreement existed in 112 of 118 cases (95%), available for this assessment. \(\kappa\) was 0.82 (very good).

**Discussion**

Our study shows that the association between strong COX-2 expression and survival in 126 vulvar cancer patients from Amsterdam could not be reproduced in an independent but similar study group of 116 patients from Groningen. This lack of reproducibility might be caused by differences between the groups of patients or by technique-associated differences.

All well-known clinicopathologic factors of prognostic significance for patients with vulvar cancer were equally present in both groups (Amsterdam – Groningen) except for age and time interval of treatment. The patients from Groningen were on average 4 years younger than those from Amsterdam. Age is not an independent factor of prognostic significance in vulvar cancer patients.(8) The time interval of treatment is considerably longer in the group from Groningen (1984-2001) than in the group from Amsterdam (1995-2003). Since all selected patients had the standard treatment it is unlikely that this difference in treatment period is important for the outcome.

The cut-off points for COX-2 expression have been defined as the best discriminating values in the initially studied population. Since this initial study group was rather small (N=50), it can be expected that the best cut-off point in another population will be different, and consequently worse test results will be found when the original cut-off...
point is used. This phenomenon might partly explain the lack of reproducibility of our results in this external validation study.

The number of cases with strong COX-2 expression is significantly lower on the TMA from Groningen than on the one from Amsterdam. Although no significant differences, except for age, between both groups of patients exist it cannot be completely ruled out that differences in characteristics of the patients underlie this phenomenon. The difference between the numbers of positive cases might also point to differences between the work-up of the material in Groningen and Amsterdam. The steps of processing, embedding and fixation of tissues were performed in different laboratories at different moments, long before the production of the TMA. The influence of differences within these procedures on the final results cannot be excluded. As in both pathological laboratories the same fixative is used the difference in work-up is limited to a difference in time between surgical removal of the specimen and fixation. As far as this could be verified it is not the case.

The age of the tissue blocks could also be a factor of influence since some antigens show loss of reactivity in very old tissue blocks. As 4 of 6 strong positive cases from Groningen were dated before 1995, this is not a convincing explanation. Both TMAs from Amsterdam and Groningen were constructed according to the same protocol. The unstained sections from Groningen were stored at room temperature for about two weeks in contrast with those from Amsterdam. The latter were stored for 2 to 4 years at -80°C and 2 days at room temperature. As the scores on the TMA of Amsterdam corresponded very well with the initial scores on the same TMA it is unlikely that antigenicity is reduced significantly under these conditions. Several studies have demonstrated the deleterious effects of storing pre-cut tissue slides at room temperature. The loss of immunoreactivity differs per antigen. The range of markers, for which antigen loss on stored paraffin slides is a problem, is not known. But, it is obvious that the difference in duration of storage at room temperature potentially accounts for a difference in intensity of expression.

As the unstained sections of the TMA from Groningen and Amsterdam underwent the same procedure for antigen retrieval and immunostaining simultaneously, it seems very unlikely that differences in immunostaining arose in this phase.

The problem of lack of reproducibility of the association between expression patterns of COX-2 and clinical pathological variables and survival also applies to other tumors. The prognostic significance of COX-2 expression in patients with squamous cell cancer of the esophagus has been extensively studied. In 13 papers, cut-off levels for overexpression varied from >10% to >50% of cells. Strong expression of COX-2 was found in 7% to 90% of the cases. Seven authors investigated the association between expression of COX-2 and survival. In four studies, COX-2 overexpression was not a significant factor for survival, in 3 it was associated with reduced survival. In 1 study the
opposite was found: low expression of COX-2 in 36 patients treated with neo-adjuvant chemotherapy was associated with shorter survival. As far as we can ascertain there hasn’t been any form of internal or external validation in the studies mentioned. In general the need for validation of new tests is widely recognized. This applies to immunohistochemical tests in particular because standardization in the technical procedures is lacking. In the current study, the only technique related factor possibly underlying the difference in intensity of expression of COX-2 in the two groups, is the difference of duration of storage at room temperature of the unstained sections of the TMA from Groningen and Amsterdam. Apart from that, it is the determination of cut off points in a rather small study group that is probably another significant factor for the lack of reproducing previous results in this external validation study. Before the decision about the prognostic significance of COX-2 expression in vulvar cancer can be taken, further research has to be done. TMA can play a significant role in future validation studies. However, cut-off levels have to be determined in a larger study population and all phases of the technical procedure of the construction of the TMAs have to be geared to one another as good as possible.
References


(5) van den Bruel A, Aertgeerts B, Buntinx F. Results of diagnostic accuracy studies are not always validated. J Clin Epidemiol 2006; 59(6):559-566.


