Multicenter prospective evaluation of the reliability of the combined use of two models to predict non-sentinel lymph node status in breast cancer patients with metastatic sentinel lymph nodes: the MSKCC nomogram and the Tenon score. Results of the NOTEGS study

Roman Rouzier*,1, Catherine Uzan2, Alexandra Rousseau3, Eugénie Guillot1, Sonia Zilberman3, Charles Meyer4, Pablo Estevez1, Pierre-Francois Dupre5, David Kere6,*, Virginie Doridot7, Gauthier D’halluin8, Xavier Fritel9, Nicolas Pouget1, Clémentine Jankowski10, Chafika Mazouni2, Tabassome Simon3 and Charles Coutant10

1Institut Curie, Paris, Saint-Cloud, France; 2Gustave Roussy, Villejuif, France; 3APHP, Paris, France; 4Hôpitaux Civils de Colmar, Colmar, France; 5CHU Brest, Brest, France; 6Institut Jean Godinot, Reims, France; 7Centre Médical République, Clermont-Ferrand, France; 8Centre Clinical Soyaux, Angoulême, France; 9Université de Poitiers, CIC 1402, CHU de Poitiers, Poitiers, France and 10Centre Georges-François Leclerc, Dijon, France

Background: The purpose of this study was to prospectively evaluate the combined use of The Memorial Sloan Kettering Cancer Center nomogram and Tenon score to select, in patients with metastatic sentinel lymph node (SN), those at low risk of metastatic non-SN for whom additional axillary lymph node dissection (ALND) could be avoided.

Methods: From January 2011 to July 2012, a prospective non-interventional nationwide study was conducted (NCT01509963). We sought to identify the false reassurance rate (FRR, a negative test result is false) in patients with both a ≤10% probability of metastatic non-SN with the MSKCC nomogram and a Tenon score ≤3.5 (low risk): the proportion of patients with metastatic non-SN at additional ALND. Our hypothesis was that these patients would have a FRR ≤5%.

Results: Data on 2822 patients with breast cancer from 53 institutions were prospectively recorded. At least one SN was metastatic (isolated tumour cells, micro- or macrometastases) in 696 patients (24.7%). Among patients with ALND and complete data to calculate combined risk (n = 504), 67 and 437 patients had low and high combined risk, respectively. Patients at low risk had less ALND (47%) compared to patients at high risk (P < 0.001). This study did not meet its primary objective because the FRR in patients with low risk was 16.4% (11 out of 67) (95% confidence interval (CI): 9.7–23.1%). In the high-risk group, 33.9% (148 out of 437) (95% CI: 29.6–38.4%) had non-SN metastases (P = 0.004).

Conclusions: In this controlled prospective study, metastatic SN patients with both a ≤10% probability of metastatic non-SN with the MSKCC nomogram and a Tenon score ≤3.5 failed to identify patients at low risk of metastatic non-SN when completion ALND was not systematic.
In breast cancer patients having at least one positive sentinel lymph node (SN) on final histology, 40–70% of them have no metastatic non-SN. This fact supports the notion that axillary lymph node dissection (ALND) is not required for these patients (Chu et al, 1999; Coutant et al, 2009a).

Several mathematical models have been developed to predict non-SN status in breast cancer patients with SN metastasis (Coutant et al, 2009a; Zhu et al, 2013).

The Memorial Sloan Kettering Cancer Center nomogram (MSKCC nomogram) and Tenon score outperform other methods in academic studies (Coutant et al, 2009a), but their exportability at multiple geographic locations and practice settings has never been reported. Moreover, the combined use of two predictors can help to optimise their predictive values, especially for patients who had an indeterminate probability of an event (Stephenson et al, 2005).

The purpose of this study was to prospectively evaluate the combined use of the MSKCC nomogram and the Tenon score to select patients with metastatic SN who were at low risk of metastatic non-SN and in whom ALND could be avoided.

### MATERIALS AND METHODS

**Patient eligibility and entry procedures.** The NOTEGS study was a prospective non-interventional nationwide study. From January 2012 to July 2013, data on 3157 patients with breast cancer from 53 institutions (university affiliated, general, regional hospital, non-profit private hospital and private practice) were recorded. In France, a combination of a radioactive colloid and patent blue dye is the recommended technique to detect the SN (INCa, 2010). At the time the study was conducted, detailed histological examination of SNs with multilevel section and immunohistochemistry was also recommended (INCa, 2010). During the SLN procedure, palpable nodes could be removed but were not considered as sentinel if they were neither blue nor radioactive. They were not considered as sentinel in the current study.

The eligibility criteria were patients aged over 18 years old with untreated invasive T1–2 breast cancer with indication for the SN procedure and protocol adherence. The SN biopsy and pathological SN examination methods were performed as previously described (Coutant et al, 2009b).

Because of the results of the ACOSOG Z0011 (Giuliano et al, 2011) and IBCSG 23-01 (Galimberti et al, 2013) trials, an additional ALND was not mandatory in the case of metastatic SN.

**Study oversight.** The study was approved by the ethical committee (Comité de Protection des Personnes Ile-de-France IV) and the French data protection authority (Commission Nationale de l’Informatique et des Libertés). It was registered at Clinicaltrials.gov (NCT01509963).

**Risk calculation combining the MSKCC nomogram and the Tenon score.** Calculation of risk by the MSKCC nomogram has been published in 2005 (Van Zee et al, 2003). The calculation must be performed using the calculator developed by the authors, which is easily accessible on the website http://www.mskcc.org/mskcc/html/5794.cfm.
Calculating the Tenon score involves three variables: (1) existence of macrometastasis in a non-SN, and if macrometastases are present then 2 points are assigned, but otherwise; (2) histological size of invasive tumour, for which 3 points are given if the tumour is >20 mm, 1.5 points if the size is 10–20 mm and 0 points if it is <10 mm; (3) ratio between the number of metastatic SN and the number of harvested SN, for which 2 points are given if the ratio is 1, and 1 point if between 0.5 and 1, and 0 points if <0.5. The scores for the three variables are then added together to calculate the Tenon score (Barranger et al, 2005).

According to the MSKCC nomogram and the Tenon score, we defined two risk groups:

- Group at low risk of metastatic non-SN: probability with the MSKCC nomogram of ≤10% and a Tenon score ≤3.5.
- Group at risk of metastatic non-SN: probability with the MSKCC nomogram of >10% and/or Tenon score >3.5.

**End point.** The end point was the false reassurance rate (FRR) (i.e., the false-negative results/all negative results) in the group at low risk of metastatic non-SN (i.e., both a ≤10% probability of metastatic non-SN with the MSKCC nomogram and a Tenon score ≤3.5 (i.e., low risk).

**Statistical considerations.** The sample size calculation was based on the FRR rate in patients considered at low risk for metastatic non-SN with the combined use of the MSKCC nomogram and the Tenon score (i.e., the proportion of positives that yield negative test outcomes with the test). A 5% ± 5% rate (<10%) was considered to be clinically acceptable (Kohrt et al, 2008; Poirier et al, 2008).

Thus, with a risk α = 5% and β = 20%, data from 235 women with metastatic SN and considered to be at low risk of metastatic non-SN were necessary. To calculate the total number of patients to include, we made two assumptions: (1) 25% (i.e., factor 4) of patients have metastatic SN, and (2) 33% (i.e., factor 3) of these patients will be predicted to have a low risk for metastatic non-SN with the combined use of the MSKCC nomogram and the Tenon score. We then multiplied the number of patients to be included by 12 (i.e., factors 4 and 3), to ensure the recruitment of 235 women with metastatic SN at low risk of metastatic non-SN required to get a precision of 5%. In all, we had to include 2820 patients.

False reassurance rate, sensitivities, specificities, predictive values positive and negative were evaluated with 95% confidence intervals (CI).

The performance of both models was quantified with respect to discrimination and, for MSKCC nomogram, to calibration as we have previously reported (Werkoff et al, 2009; Coutant et al, 2009a).

Univariate and multivariate analyses were performed using a logistic regression model. Odds ratios were evaluated with 95% CI. All analyses were performed using the R package with the Design, Hmisc, rms and verification libraries (http://lib.stat.cmu.edu/R/CRAN/). Test results were considered significant when the P-value <0.05.

**RESULTS**

**Patients and pathological data.** We enrolled 3157 patients. Three hundred thirty-five patients were excluded: 227 patients did not

| Table 1. Clinical and pathological data for the 2822 patients with invasive breast cancer having SN procedure, and the group of 696 patients who did have at least one metastatic SN |
|-----------------|-----------------|-----------------|
| **Entire cohort (n = 2822)**                  | **At least one metastatic SN – complete data for the combined use of two models (n = 504)** |
| Age, years                | 60.5 (25–98)  | 59.2 (26–98)   |
| Invasive tumour size at final histology, mm | 15.7 (1–90)    | 19.7 (1–70)    |
| Histology                  |                |                |
| Invasive ductal carcinoma  | 2263           | 404            |
| Invasive lobular carcinoma | 373            | 69             |
| Other                      | 186            | 31             |
| Tumour grade               |                |                |
| Well differentiated, grade 1 | 819            | 94             |
| Moderately differentiated, grade 2 | 1404          | 269            |
| Poorly differentiated, grade 3 | 561            | 141            |
| Not determined             | 86             | 11             |
| Lymphovascular space involvement | 2069          | 277            |
| No                          | 407            | 178            |
| Yes                        | 136            | 49             |
| Not determined             | 340            | 92             |
| Oestrogen/progesterone receptor status | 2493          | 458            |
| Positive                   | 290            | 45             |
| Negative                   | 39             | 18             |
| Not determined             | 346            | 95             |
| Her-2/neu status           |                |                |
| Overexpressed/amplified    | 245            | 53             |
| Negative                   | 2460           | 436            |
| Not determined             | 117            | 15             |
| Mean no of SN per patient (range) | 2.23 (1–14)   | 2.22 (1–12)   |
| No of patients with positive non-SN | 180           | 159            |

Abbreviation: SN = sentinel lymph node.
have SN procedure, 10 had no SN detected and 98 underwent a SN procedure for in situ carcinoma. Finally, 2822 patients underwent a SN procedure for invasive breast cancer (Figure 1).

The patients and pathological data are listed in Table 1. Among them, 696 patients (24.7%) had at least one metastatic SN. Axillary lymph node dissection was performed in 518 patients (74.4%). At least one metastatic non-SN was identified in 167 patients (32.2%).

**Performance of the combined use of the Tenon score and the MSKCC nomogram.** Among the 696 patients, 170 did not have completion ALND and 8 were lost to follow-up. Among the 518 patients with ALND, the Tenon score was calculated for all patients, and MSKCC probabilities and the combined use of the two predictors were calculated for 504 patients.

One hundred eighty-three patients (35.3%) were at low risk according to the Tenon score, 93 (18.5%) were at low risk with the MSKCC nomogram (Supplementary Table 2) and 67 (13.3%) at low risk with combined predictions (Table 2).

Among the 67 patients at low risk with combined predictions, 11 had at least one metastatic non-SN. The FRR of combined predictions was 16.4% (95% CI: 9.7–23.1%). Sensitivity and specificity were 93.1% (95% CI: 88.6–96.2%) and 16.2% (95% CI: 14.1–17.7%), respectively.

**Table 2. Performance of the combined use of the two predictors (Tenon score, the Memorial Sloan Kettering Cancer Center nomogram) to predict non-SN status in breast cancer patients with metastatic SN**

| Non-SN status | Combined Low risk | Combined High risk |
|---------------|-------------------|--------------------|
| Negative      | 56                | 289                |
| Positive      | 11                | 148                |
| FRR (95% CI)  | 16.4% (9.7–23.1)  |                    |
| Sensitivity (95% CI) | 93.1% (88.6–96.2) | |
| Specificity (95% CI) | 16.2% (14.1–17.7) | |
| PPV (95% CI)  | 33.9% (32.2–35.0) |        |
| NPV (95% CI)  | 83.6% (72.9–91.0) |        |

Abbreviations: CI = confidence interval; FRR = false reassurance rate; NPV = negative predictive value; PPV = positive predictive value; SN = sentinel lymph node.

**DISCUSSION**

Until 2011, SLNB had to be completed by ALND if the SLN was metastatic (Lyman et al., 2005; INCa, 2010). Nomograms have been developed to quantify the likelihood of identifying additional positive axillary nodes, but their use has not been universally accepted. In the current prospective study, we evaluated the exportability of two well-established nomograms in a multicentric study including a large variety of settings. We failed to demonstrate that their use is robust because the FRR was over 10%. Moreover, only a minority of patients were eligible for omission of ALND by the use of models.
Given the excellent outcomes for modern breast cancer management to reduce and individualise surgery, systemic therapy and radiation therapy are among the driving forces of clinical studies. The development of nomograms to identify women who are at low or high risk of residual non-sentinel node disease after a positive sentinel node biopsy is in line with this evolution. While they have been widely validated in retrospective or unicentric studies, few large prospective controlled studies have been conducted. Our study is the one of the largest studies to evaluate the performance of models to predict the non-SLN rate. The interpretation of previous studies is limited by a selection bias the patients who did not have complementary axillary dissection in case of metastatic SLN were not evaluated. The strength of our study is that the data includes all patients before the initial SLN procedure. This eliminates the risk of inclusion bias in contrast with previously published studies (Zhu et al, 2013).

The primary outcome was not met as the FRR of combined predictions was 16.2% (95% CI: 8.9–26.8%). The MSKCC nomogram and the Tenon score had similar performances with an AUC of 0.65 (95% CI: 0.62–0.67) and 0.63 (95% CI: 0.61–0.66), respectively. Zhu et al (2013) have recently conducted a meta-analysis to determine which nomogram is best for predicting non-SN metastasis in breast cancer patients; the Cambridge, Mayo, MDA, MSKCC, Stanford and Tenon models were validated in 2156, 2431, 843, 8143, 3700 and 3648 patients, respectively. The pooled AUCs for the Cambridge, MDA, MSKCC, Mayo, Tenon and Stanford models were 0.721, 0.706, 0.715, 0.728, 0.720 and 0.688, respectively. The main reason for the relatively low AUCs in our study is probably related to the omission of ALND in selected patients as an additional ALND was not mandatory. Specific data of these patients are reported in Supplementary Tables 3 and 4. Theoretically, AUC measures from ROC plots are independent of prevalence. However, we observe in our study the spectrum effect, which has been widely discussed in the literature since the initial paper by (Ransohoff and Feinstein (1978)). In line with our finding, the meta-analysis by Zhu et al (2013) revealed that the SLN micrometastasis rate was associated with improved predictive accuracy. In our study, a systematic ALND in case of positive SN was not mandatory. The proportion of patients without ALND was statistically greater in patients with micrometastasis or ITC: 83.5% vs 35.4% (Supplementary Table 3). This artificially shifts the rate of positive non-SLN by excluding low-risk patients and thus negatively impacting the accuracy of the models. However, the low specificity and PPV of these scores are clearly a matter of concern in terms of clinical utility.

The other finding is the low proportion of patients selected by the MSKCC nomogram and the Tenon score for omission of ALND (26.9% and 46.7%, respectively). The American College of Surgeons Oncology Group led the multicenter Z0011 trial to determine the effects of ALND on overall survival in patients with one or two positive SLN (Giuliano et al, 2011). The use of SLND alone compared with ALND did not result in inferior survival and locoregional control. For micrometastasis only, IBCSG 23-01 provided similar results (Galimberti et al, 2013) Using the Z0011 eligibility criteria, ~70% of patients are eligible for omitting completion ALND (Delpech et al, 2013). The rate was 13.5% in our study using the combined approach, far beyond a selection based on Z011 criteria that has been endorsed by most guidelines. This over selection is clearly a limitation of the use of models in clinical practice. The absence of benefit in terms of survival between SLNB alone and complete ALND in Z011, IBCSG 23-01 and in all non-randomised studies limits the additional information gained from ALND to the number of nodes containing metastases (Giuliano et al, 2011; Galimberti et al, 2013; Ram et al, 2014; Bonneau et al, 2015). However, this prognostic information, obtained at the cost of an increase in morbidity, is unlikely to change systemic therapy decisions. Moreover, there is no improvement in axillary recurrence and DFS by ALND when the nodal invasion is micrometastatic or limited to a few lymph nodes. This suggests that this limited burden of disease is likely to be controlled with systemic therapy and RT. Other authors have suggested that it is more important to identify patients at risk of pN2 disease (≥4 metastatic nodes) (Werkoff et al, 2009; Gooch et al, 2014). Other models, and particularly those that integrate extracapsular extension of the tumour in the SN, or those designed to predict the risk of ≥4 metastatic nodes, have to be tested in this series.

In conclusion, we demonstrated in this controlled prospective trial that in metastatic SN patients with both a ≤10% probability of metastatic non-SN with the MSKCC nomogram and a Tenon score ≤3.5, the FRR was statistically over 5% and resulted in a selection that was not compatible with clinical practice. Even evaluated separately, the discrimination, calibration and clinical utility of the Tenon score and the MSKCC nomogram was moderate. The authors declare no conflict of interest.

REFERENCES

Barranger E, Coutant C, Flahault A, Delpech Y, Darai E, Uzan S (2005) An axilla scoring system to predict non-sentinel lymph node status in breast cancer patients with sentinel lymph node involvement. Breast Cancer Res Treat 91: 113–119.

Bonneau C, Hequet D, Estevez JP, Pouget N, Rouzier R (2015) Impact of axillary dissection in women with invasive breast cancer who do not fit the Z0011 ACOSOG trial because of three or more metastatic sentinel lymph nodes. Eur J Surg Oncol 41: 998–1004.

Chu KU, Turner RR, Hansen NM, Brennan MB, Bilchik A, Giuliano AE (1999) Do all patients with sentinel node metastasis from breast carcinoma need complete axillary node dissection? Ann Surg 229: 536–541.

Coutant C, Olivier C, Lambaudie E, Fodrinier M, Marchal F, Guillemin F, Seince N, Thomas V, Levêque J, Barranger E, Darai E, Uzan S, Houvenaeghel G, Rouzier R (2009a) Comparison of models to predict nonsentinel lymph node status in breast cancer patients with metastatic sentinel lymph nodes: a Prospective Multicenter Study. J Clin Oncol 27: 2800–2808.

Coutant C, Rouzier R, Fodrinier E, Marchal F, Guillemin F, Seince N, Seince N, Rodrigues A, Darai E, Uzan S, Barranger E (2009b) Validation of the Tenon breast cancer score for predicting non-sentinel lymph node status in breast cancer patients with sentinel lymph node metastasis: a prospective multicenter study. Breast Cancer Res Treat 113: 537–543.

Delpech Y, Bricou A, Lousqury R, Hudry D, Jankowski C, Willecoq C, Thoury A, Loustalot C, Coutant C, Barranger E (2013) The Exportability of the ACOSOG Z0011 Criteria for omitting axillary lymph node dissection after positive sentinel lymph node biopsy findings: a Multicenter Study. Ann Surg Oncol 20: 2556–2561.

Galimberti V, Cole BF, Zarrida S, Viale G, Luini A, Veronesi P, Baratella P, Chifu C, Sargenti M, Intra M, Gentilini O, Mastropasqua MG, Massarut G, Massarat S, Garbay JR, Zgajnar J, Galatius H, Recalcati A, Littlejohn D, Bamert M, Colleoni M, Price KN, Regan MM, Goldhirsh A, Coates AS,
