Performance of Salamanca refinement of the T3-AJCC8 versus the Brigham and Women's Hospital and Tübingen alternative staging systems for high-risk cutaneous squamous cell carcinoma

Laura Puebla-Tornero, MD,[a] Luis Antonio Corchete-Sánchez, MSc, PhD,[b,c,d] Alberto Conde-Ferreirós, MD,[a] Natalia García-Sancha, MSc,[b,e] Roberto Corchado-Cobos, MSc,[b,e] Concepción Román-Curto, MD, PhD,[a,b] and Javier Cañueto, MD, PhD[a,b,e]
Salamanca, Spain

Introduction: The Brigham and Women's Hospital and the Tübingen cutaneous squamous cell carcinoma (SCC) stratification systems propose different criteria from the American Joint Committee on Cancer, eighth edition. Our group identified prognostic subgroups within T3 stage according to the American Joint Committee on Cancer eighth edition's classification, the most common classification for high-risk cutaneous SCCs.

Objective: To compare the performance and prognostic accuracy of these staging systems in a subset of high-risk cutaneous SCCs.

Methods: Homogeneity, monotonicity, and McNemar tests for pairwise comparisons were assessed. Distinctiveness and relative risk of poor outcome were calculated by stage. Prognostic accuracy was compared with respect to quality (Akaike and Bayesian information criteria), concordance (Harrell C-index and Gönen and Heller concordance probability estimate), and predictive accuracy (sensitivity, specificity, negative predictive value, positive predictive value, and global accuracy).

Results: The Brigham and Women's Hospital and Salamanca systems were more distinctive, homogeneous, and monotonic than the Tübingen system. The Tübingen system was the most specific, whereas the Salamanca and Brigham and Women's Hospital systems were more sensitive. Negative predictive value was high in all 3 systems, but positive predictive value and accuracy were low overall.

Conclusions: Alternative staging systems may partially overcome the heterogeneity and low prognostic accuracy of the American Joint Committee on Cancer, eighth edition and enable high-risk cutaneous SCCs to be stratified more reliably, but their prognostic accuracy is still low. Considering the accumulation of risk factors may improve high-risk cutaneous SCC risk stratification. (J Am Acad Dermatol 2021;84:938-45.)

Key words: AJCC8; cutaneous squamous cell carcinoma; prognosis; skin cancer; staging.

INTRODUCTION

Cancer staging systems, which stratify tumors into prognostic subgroups by risk factors, are useful for managing patients. Existing cutaneous squamous cell carcinoma (SCC) staging systems do not accurately predict poor prognosis based on T stage,[1-3] making

From the Departamento de Dermatología,[a] Instituto de Investigación Biomédica de Salamanca (IBSAL),[b] and Servicio de Hematología,[d] Hospital Universitario de Salamanca; and Laboratory 12[c] and Laboratory 7,[e] IBMCC-CSIC, Salamanca.
Drs Puebla-Tornero and Corchete-Sánchez contributed equally to this article.
Funding sources: Dr Cañueto is partially supported by grant PI18/000587 (Instituto de Salud Carlos III, cofinanced by FEDER funds).
IRB approval status: The study was approved by the University Hospital of Salamanca's Institutional Ethics Review Board.

Accepted for publication December 4, 2020.
Reprints not available from the authors.
Correspondence to: Javier Cañueto, MD, PhD, Department of Dermatology, Complejo Asistencial Universitario de Salamanca, Paseo San Vicente 58-182, 37007 Salamanca, Spain. E-mail: jcanueto@usal.es.
Published online January 9, 2021.
0190-9622/$36.00

https://doi.org/10.1016/j.jaad.2020.12.020

improvements essential to meet clinical practice needs. Alternative staging systems have been developed but require continual refinement as new prognostic evidence emerges. Although high-risk cutaneous SCCs represent a minority of cases, identifying them is the main challenge in effective risk stratification, and the predominant focus of research. Cutaneous SCC is considered high risk when staged as T3/T4 stage according to the American Joint Committee on Cancer eighth edition's classification (T3/T4-AJCC8). Most high-risk cutaneous SCCs are staged as T3-AJCC8[4] because T4-AJCC8 accounts for only 0.3% of cutaneous SCCs in some series.[5]

The University of Tübingen group stratified cutaneous SCCs into prognostic groups based on tumor thickness,[6] width,[7] and several associated risk factors,[7] but it is unclear how these characteristics relate to T categories.[8] The same group recently proposed a risk-stratification system based on the accumulation of risk factors.[9] Although originally designed for cutaneous SCC–specific death, it may be valid for predicting other poor outcome events. In 2013, the Brigham and Women's Hospital alternative staging system emerged,[10,11] quickly demonstrating its superiority over the American Joint Committee on Cancer, seventh edition.[12] It considers the combinations of 4 risk factors to classify cutaneous SCCs into T stages.[10,11] Its use in clinical practice is straightforward and popular, allowing stratification of tumors beyond the head and neck. The American Joint Committee on Cancer, eighth edition (AJCC8), based on additional studies[6,12] and featuring significant changes concerning cutaneous SCC,[13] was released in 2017. AJCC8 has limitations, including significant heterogeneity within the T3 stage, to which category most high-risk cutaneous SCCs belong.[14,15] Combining AJCC8 risk factors allows various prognostic subgroups to be identified in T3-AJCC8 (Table I).[4] In this tentative subclassification, cases featuring 3 or more risk factors were at particularly high risk of recurrence.[4]

Several comparisons of staging systems have emerged since AJCC8 was released,[5,8,14,16] but none involving these 3 alternative systems (Brigham and Women's Hospital, Tübingen, and T3-AJCC8) and focused on high-risk cutaneous SCC. Our study compared their performance in predicting poor prognosis and death from disease in a series of head and neck high-risk cutaneous SCCs. We identify opportunities for improving staging systems and implement a useful methodology for comparing staging systems for a broader range of entities than cutaneous SCC alone.

CAPSULE SUMMARY

- A more precise stratification of high-risk cutaneous squamous cell carcinoma may help clinicians in daily practice.
- Current alternative staging systems are an improvement on the official American Joint Committee on Cancer, eighth edition but have low accuracy for high-risk cutaneous squamous cell carcinomas. They all have high negative predictive value, but their specificity and positive predictive value are low overall. Considering the combination of risk factors can help better identify which cutaneous squamous cell carcinomas are at greatest risk of recurrence.

PATIENTS AND METHODS

Patients and variables

We selected a retrospectively collected cohort of 196 primary head and neck cutaneous SCCs, staged as T3 by the AJCC8, from the University Hospital of Salamanca, Spain.[4] All 2391 pathology reports between January 1, 2005, and December 31, 2016, were evaluated; those meeting AJCC8 criteria for classifying cutaneous SCCs as T3 were selected. Of the 216 tumors retrieved, 9 were excluded because of missing clinical data or unavailability of slides, and 11 were excluded because they were recurrent, leaving a final cohort of 196 primary T3 tumors. The study was approved by the University Hospital of Salamanca's Institutional Ethics Review Board and complied with Strengthening the Reporting of Observational Studies in Epidemiology recommendations.

Patients' age, sex, and immune status, and the tumor traits involved in prognosis and staging, were recorded. Tumors were classified by the 3 alterative staging systems of the Salamanca University Hospital, Brigham and Women's Hospital, and Tübingen University Hospital. Disease-specific poor outcome was evaluated as local recurrence, metastasis, and disease-specific death. Metastasis and disease-specific death were considered major events of poor outcome. Thus, hereafter, disease-specific poor outcome includes local recurrence, metastasis, or death, whereas major events refer only to metastasis or death. Local recurrences were assessed by considering the pathology and medical records. Irrespective of how the recurrence was managed, a biopsy was always performed.

Statistical analysis

Prognosis was analyzed by Cox proportional hazards regression using competing risk analysis. Two types of model were fitted to account for competing risks: cause-specific hazard models,[17]

Abbreviations used:

AJCC8:	American Joint Committee on Cancer, eighth edition
SCC:	squamous cell carcinoma
T3/T4-AJCC8:	T3/T4 stage according to the American Joint Committee on Cancer eighth edition's classification

which consider events of interest separately from competing risk events; and Fine and Gray proportional subdistribution hazards regression models,[18] which consider risk of events of interest over time in the presence of competing risk events. Disease-specific poor outcome, local recurrence, major events, and disease-specific death were analyzed (Supplemental Material available via Mendeley at https://doi.org/10.17632/c9bjyxbk2r.1). For tumors without any poor outcome, survival time was censored on the date of death or, if the patient was alive at data collection, on the date of their last medical record review. Quality of the survival models was compared by the Akaike and Bayesian information criteria, in which lower indices indicate a better-quality model.

McNemar test was used for paired categoric data to analyze differences in T-stage assignment of tumors between classification systems. Distinctiveness, homogeneity, and monotonicity were assessed to evaluate the performance of all 3 systems. Distinctiveness was assessed by evaluating prognostic discrimination among stages. Odds ratios of outcome events within each system were estimated with binary logistic regression in R (R Foundation). Homogeneity was evaluated by comparing the prognosis of low T stages (T3a, T1/T2a, and T1/T2 in the Salamanca refinement of T3-AJCC8, the Brigham and Women's Hospital, and Tübingen staging systems, respectively). Monotonicity was assessed by comparing the prognosis of high T stages (T3b-T3c, T2b-T3, and T3-T4), following a similar approach to those of other groups[12,16] and ours.[14]

Concordance analysis estimates the probability of selecting a pair of patients for which the one with the greatest risk is also characterized by the shortest disease-free survival. Concordance among staging systems was compared by 2 approaches: the Harrell C-index[19] and the Gönen and Heller concordance probability estimate,[20] estimated in R with the dynpred (version 0.1.2) and concordance probability estimate (version 1.5.1) packages, respectively. Higher indices indicate greater accuracy. Concordance values were calculated by considering all T-staging groups within each system and by grouping staging categories as low or high risk (Supplemental Material).

Sensitivity, specificity, positive and negative predictive value, and accuracy were calculated considering the aforementioned dichotomous grouping categories (see Supplemental Material for details of statistical methods). Analyses were performed with R's boot (version 1.3-22) package.

RESULTS

Data from our high-risk cutaneous SCC cohort are summarized in Table II. The median age was 84.4 years, 116 of 196 patients were men, and almost 20% were immunosuppressed. Overall, 52 patients (26.5%) developed disease-specific poor outcomes during follow-up, and 32 events (16.3%) were major. Ten patients (5.1%) died from the disease. The AJCC8 T3b/T3c represented 51.5% of cases, the Brigham and Women's Hospital T2b/T3 47.9%, and the Tübingen system (T3-T4 points) 34.7%.

Outcomes by T stage

Homogeneity was assessed by comparing outcomes in low T categories in the 3 staging systems (Table III). The lowest proportion of tumors evolving poor outcome events in low T stages was observed in T3a-AJCC8 (28.8% and 21.9% for disease-specific poor outcome and major events, respectively); the lowest proportion of disease-related death in low T stages occurred in the T1/T2a Brigham and Women's Hospital system. Conversely, the highest proportion of poor prognoses occurred in T3b/T3c AJCC8; the highest proportion of disease-related death was for T2b/T3 Brigham and Women's Hospital system (where all deaths belonged). The Salamanca and Brigham and Women's Hospital systems had similar homogeneity and monotonicity, producing better results than the Tübingen system. The McNemar tests demonstrated that high T stages were similarly captured by the Salamanca and Brigham and Women's Hospital systems. However, significant differences ($P < .05$) were observed between the Tübingen and the other 2 systems (Supplemental Table I).

Distinctiveness was assessed by evaluating prognostic discrimination between stages (Table III). Distinctiveness was similar for the Salamanca refinement of the T3-AJCC8 and the Brigham and Women's Hospital system with respect to disease-specific poor outcome and major events. The 5-year cumulative incidence for the evaluated outcomes gradually increased with the Brigham and Women's Hospital and Salamanca systems, but did not increase as clearly with the Tübingen system,

Table I. Features of alternative staging systems

		T3-AJCC8 Salamanca staging system	BWH staging system, high-risk factors, no.[1*]		Tübingen staging system, points[2†]	
Low risk	T3a	Thickness >6 mm (with no invasion beyond subcutaneous fat) ± Width ≥ 4 cm	T1	None	T1	0–1
			T2a	1	T2	2
High risk	T3b	Invasion beyond the subcutaneous fat (according to AJCC8)[‡] or Perineural invasion (according to AJCC8)[‡]	T2b	2–3	T3	3
	T3c	Combination of both T3b risk factors (invasion beyond the subcutaneous fat + perineural invasion) or ≥3 risk factors combined[§]	T3	≥4	T4	4

BWH, Brigham and Women's Hospital; *OR*, odds ratio.

*High-risk factors: tumor diameter greater than or equal to 2 cm, invasion beyond subcutaneous fat, poorly differentiated, and perineural invasion greater than or equal to 0.1 mm.

†A tumor thickness of greater than 2 to 4 mm scores 1 point; a thickness of greater than or equal to 4 mm scores 2 points. Immunosuppression and the presence of desmoplasia each score an additional point.

‡Perineural invasion according to the AJCC8 consists of the involvement of nerves with a diameter of greater than or equal to 0.1 mm, with the involvement of nerves located deeper than the dermis, or by clinical or radiologically evident perineural invasion.

§The combination of 3 or more risk factors may include any of the risk factors that define a tumor as stage T3.

which was therefore considered less distinctive (Table III). Odds ratios for the different outcomes gradually increased with the Salamanca and Brigham and Women's Hospital systems (Supplemental Table II). The Tübingen system showed an improvement when cases were classified as low or high risk, meaning that it was at least as distinctive as the Brigham and Women's Hospital and Salamanca systems.

Prognostic accuracy

Supplemental Figs 1 and 2, respectively, show the cumulative incidence of events considering all T stages for the different systems for all outcome events and for low- and high-risk subgroups. The goodness of fit of these survival models was approximately the same for almost all events (Supplemental Table III).

We used the Akaike and Bayesian information criteria to measure the relative quality of the staging models (Supplemental Table IV). The Tübingen staging system was the most balanced in terms of quality for local recurrence because its Akaike and Bayesian information criteria values were the lowest of all the scenarios. Conversely, the Salamanca and Brigham and Women's Hospital systems showed higher quality for major events and disease-specific death.

C-indices were similar for the Salamanca and Brigham and Women's Hospital staging systems, especially for major events and disease-specific death, with respective values of 69.35% and 83.55% for Salamanca, and 68.94% and 82.94% for Brigham and Women's Hospital, considering cause-specific hazard. The Tübingen system performed slightly worse, with C-indices of 61.03% and 75.08% for major events and disease-specific death. Considering the low- and high-risk groups within each staging system, the Brigham and Women's Hospital system had the greatest concordance for major events and disease-specific death (C-index 65.72% and 77.07%). The C-index cannot discriminate deaths from other causes because staging systems are designed to identify only tumor-related deaths and yielded values close to 50% for the competing risk (Supplemental Table V).

The Salamanca and Brigham and Women's Hospital staging systems were the most sensitive, with major events and disease-specific death values of 78.3% and 90.2% in the former, and 75.8% and 100% in the latter. The Tübingen system had greater specificity than the other 2, with values of 71.2%, 68.8%, and 67.4% for local recurrence, major events, and disease-specific death, respectively. The negative predictive value was high in all 3 systems, with values of greater than 80% for almost all

Table II. Summary of characteristics of the cohort

	Frequency	%
Patient history		
Mean age (SD), y	84.48 (10.50)	
Sex, men/women	116/80	
Immunosuppression	38	19.4
Organ transplant recipients	4	2
Hematologic malignancies*	20	10.2
Chronic lymphocytic leukemia	16	8.2
Immunosuppressive treatment for chronic inflammatory disease	6	3.1
Solid tumors†	8	4.1
Tumor traits		
Width (SD), mm	22.32 (11.76)	
Thickness (SD), mm	9.27 (4.21)	
Grade of differentiation		
Good to moderate	138	70.4
Poor	58	29.6
Perineural invasion	70	35.7
Perineural invasion of nerves ≥0.1 mm	34	17.3
Deep nerve invasion	18	9.2
Invasion beyond the subcutaneous fat	95	48.5
Muscle invasion	86	43.9
Bone invasion	5	2.6
Tumor location		
Lower lip	23	11.7
Ear	15	7.7
Other location	158	80.6
Growth pattern		
Noninfiltrative	118	60.2
Infiltrative	78	39.8
Desmoplasia	37	18.9
Poor prognosis		
Disease-specific poor outcome	52	26.5
Local recurrence	36	18.4
Major events	32	16.3
Disease-specific death	10	5.1
Follow-up, mean and SD, mo	44.54 (29.37)	

SD, Standard deviation.
*Hematologic malignancies include chronic lymphocytic leukemia (n = 16) and other hematologic dyscrasias (n = 4).
†Excluding nonmelanoma skin cancer.

evaluated outcomes, but the positive predictive value was very low for all 3. Accuracy was low for all 3 staging systems; the Tübingen system was the best in this respect (Table IV).

DISCUSSION

Prognostic stratification of high-risk cutaneous SCC remains a huge problem for clinicians. Different staging systems are available for cutaneous SCC stratification, but their performance in high-risk cutaneous SCCs needs to be specifically explored. Most high-risk cutaneous SCCs are staged as T3-AJCC8, but a substantial proportion of these behave well.[14] Our group recently demonstrated heterogeneity within T3-AJCC8 for cutaneous SCC, which prompted us to propose a refinement of this stage, which includes most high-risk cutaneous SCCs.[4] Alternative staging systems are available, the Brigham and Women's Hospital and Tübingen systems being the most popular. Here, we compared all 3 systems and provided evidence of their low prognostic accuracy for high-risk cutaneous SCC, the heterogeneity of the T3-AJCC8 staging category, and the value of considering combinations of risk factors for cutaneous SCC stratification in future staging editions.

The Brigham and Women's Hospital and Salamanca systems partially overlapped, probably because the AJCC8 is based on previous work by the Brigham and Women's Hospital group.[10,12] They were more distinctive than the Tübingen system for disease-specific poor outcome and major events; the Brigham and Women's Hospital system was the best for disease-specific death. Indeed, all cutaneous SCCs causing death in this cohort were staged as T2b/T3 in the Brigham and Women's Hospital system. Only 1 of the 9 T3a Salamanca patients died from the disease, so the Salamanca system's performance is more than fair relative to disease-specific death. The increase in odds ratio of poor outcome with T stage was clearer for the Salamanca and Brigham and Women's Hospital systems than for the Tübingen system, especially for major events and disease-specific death.

The Akaike and Bayesian information criteria indices showed that the Tübingen system had higher prognostic accuracy for local recurrence, but the Brigham and Women's Hospital and Salamanca methods were more accurate for major events and disease-specific death. C-indices were higher for the Brigham and Women's Hospital and Salamanca systems than for the Tübingen system. Cutaneous SCC-specific death yielded the biggest differences, in which the first 2 systems had a C-index of greater than 80%, implying excellent concordance for disease-specific death. They were also the most sensitive, whereas the Tübingen system was the most specific. The negative predictive value was high for all 3 staging systems but the positive predictive value was very low, meaning that an advanced T-stage cutaneous SCC with any of these systems may behave well yet still reflect the heterogeneity within these "high-risk" groups. Conversely, an early T stage is highly predictive of good prognosis. All 3 systems had low accuracy, the Tübingen system performing the best.

Sensitivity and specificity were lower than in other comparative studies,[8,16,21] possibly because we studied a cohort of high-risk cutaneous SCCs, whereas others considered high- and low-risk

Table III. Homogeneity, monotonicity, and distinctiveness of staging systems

	System	Group	n	DSPO, no. of events (%)	LR, no. of events (%)	MEs, no. of events (%)	DSD, no. of events (%)
Homogeneity and monotonicity							
Homogeneity	Salamanca	T3a	95	15/52 (28.8)	9/36 (25.0)	7/32 (21.9)	1/10 (10.0)
	BWH	T1/T2a	102	20/52 (38.5)	14/36 (38.9)	8/32 (25.0)	0/10
	Tübingen	T1/T2	128	25/52 (48.1)	14/36 (38.9)	15/32 (46.9)	3/10 (30.0)
Monotonicity	Salamanca	T3b/T3c	101	37/52 (71.2)	27/36 (75.0)	25/32 (78.1)	9/10 (90.0)
	BWH	T2b/T3	94	32/52 (61.5)	22/36 (61.1)	24/32 (75.0)	10/10 (100)
	Tübingen	T3/T4	68	27/52 (51.9)	22/36 (61.1)	17/32 (53.1)	7/10 (70.0)

5-y CIF Classification	System	Group	n	No. of events (%)	5-y CIF (95% CI)	No. of events (%)	5-y CIF (95% CI)	No. of events (%)	5-y CIF (95% CI)	No. of events (%)	5-y CIF (95% CI)
Distinctiveness											
All groups	Salamanca	T3a	95	15 (7.65)	15 (8.6–23.1)	9 (4.59)	9.8 (4.8–16.9)	7 (3.57)	6.3 (2.6–12.5)	1 (0.51)	1.1 (0.1–5.3)
		T3b	59	19 (9.69)	32.5 (20.3–45.3)	14 (7.14)	24.2 (13.5–36.6)	10 (5.1)	18 (9.1–29.4)	0	NA
		T3c	42	18 (9.18)	43.8 (27.8–58.8)	13 (6.63)	31.6 (18.0–46.1)	15 (7.65)	36.1 (21.7–50.7)	9 (4.59)	20.4 (9.4–34.5)
	BWH	T1	26	3 (1.53)	7.7 (1.3–22.1)	2 (1.02)	7.7 (1.3–22.1)	1 (0.51)	NA	0	NA
		T2a	76	17 (8.67)	23.6 (14.0–34.7)	12 (6.12)	16.1 (8.4–26.1)	7 (3.57)	9.9 (4.3–18.3)	0	NA
		T2b	85	26 (13.27)	31 (21.3–41.1)	17 (8.67)	20.4 (12.4–29.6)	19 (9.69)	22.6 (14.3–32.0)	7 (3.57)	7.2 (2.9–14.2)
		T3	9	6 (3.06)	66.7 (20.5–90.1)	5 (2.55)	63 (15.5–89.2)	5 (2.55)	55.6 (16.2–82.7)	3 (1.53)	38.1 (6.8–70.7)
	Tübingen	T1	8	2 (1.02)	25 (3.0–57.9)	2 (1.02)	25 (3.0–57.9)	0	NA	0	NA
		T2	120	23 (11.73)	18.8 (12.3–26.4)	12 (6.12)	10.4 (5.6–16.8)	15 (7.65)	11.9 (6.8–18.5)	3 (1.53)	1.8 (0.3–5.8)
		T3	62	25 (12.76)	40.3 (27.5–52.7)	19 (9.69)	30.7 (19.2–42.9)	16 (8.16)	26.6 (16.1–38.3)	6 (3.06)	10.2 (4.1–19.6)
		T4	6	2 (1.02)	NA	3 (1.53)	NA	1 (0.51)	16.7 (0.5–54.9)	1 (0.51)	33.3 (0.1–84.5)
Low- and high-risk groups	Salamanca	T3a	95	15 (7.65)	15 (8.6–23.1)	9 (4.59)	9.8 (4.8–16.9)	7 (3.57)	6.3 (2.6–12.5)	1 (0.51)	1.1 (0.1–5.3)
		T3b/T3c	101	37 (18.88)	37.3 (27.5–47.1)	27 (13.78)	27.3 (18.6–36.6)	25 (12.76)	25.5 (17.3–34.6)	9 (4.59)	8.5 (3.9–15.3)
	BWH	T1/T2a	102	20 (10.20)	19.2 (11.8–28.0)	14 (7.14)	13.8 (7.7–21.8)	8 (4.08)	7.3 (3.2–13.6)	0	NA
		T2b/T3	94	32 (16.33)	34.4 (24.8–44.1)	22 (11.22)	23.9 (15.7–33.1)	24 (12.24)	25.8 (17.3–35)	10 (5.1)	9.9 (4.8–17.2)
	Tübingen	T1/T2	128	25 (12.76)	19.2 (12.8–26.7)	14 (7.14)	11.3 (6.5–17.7)	15 (7.65)	11.1 (6.4–17.4)	3 (1.53)	1.7 (0.3–5.5)
		T3/T4	68	27 (13.78)	40.4 (28.1–52.4)	22 (11.22)	32.9 (21.5–44.8)	17 (8.67)	26.2 (16.1–37.4)	7 (3.57)	11.1 (4.8–20.4)
		T3/T4	68	27 (13.78)	2.7 (1.4–5.2)	22 (11.22)	3.9 (1.9–8.4)	17 (8.67)	2.5 (1.2–5.5)	7 (3.57)	4.8 (1.3–22.8)

"All groups" stands for the classification that considers all staging categories; "dichotomous" classification stands for the classification that splits the cohort into high- and low-stage categories within each staging system. For clarity, data are shown as the percentage and 5-year cumulative incidence of the outcome of interest. *BWH*, Brigham and Women's Hospital; *CI*, confidence interval; *CIF*, cumulative incidence function; *DSD*, disease-specific death; *DSPO*, disease-specific poor outcome; *LR*, local recurrence; *MEs*, major events; *NA*, not applicable.

Table IV. Sensitivity, specificity, negative predictive value, positive predictive value, and accuracy

Staging system	Tumors, n	Group	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Accuracy (95% CI)
Salamanca	196	DSPO	71.3 (70.2–72.3)	55.4 (54.7–56.0)	36.6 (35.9–37.4)	84.2 (83.6–84.8)	60 (59.4–60.5)
		LR	74.3 (73.1–75.5)	53.9 (53.3–54.5)	26.6 (26.0–27.3)	90.4 (90.0–90.9)	57.2 (56.7–57.8)
		MEs	78.3 (77.1–79.4)	53.5 (52.8–54.1)	24.5 (23.8–25.2)	92.7 (92.3–93.2)	58.5 (57.9–59.1)
		DSD	90.2 (88.6–91.7)	50 (49.4–50.6)	8.5 (8.0–8.9)	98.9 (98.8–99.1)	52.4 (51.8–52.9)
BWH	196	DSPO	61.9 (60.8–63.1)	56.8 (56.2–57.5)	33.9 (33.2–34.6)	80.4 (79.8–81.0)	58.4 (57.8–59.0)
		LR	61.8 (60.5–63.2)	55.2 (54.6–55.8)	22.9 (22.2–23.6)	86.3 (85.8–86.9)	56.8 (56.2–57.3)
		MEs	75.8 (74.5–77.0)	57.6 (56.9–58.2)	25.6 (24.8–26.4)	92.4 (91.9–92.8)	60.6 (60.0–61.2)
		DSD	100 (100–100)	54.7 (54.1–55.4)	10.8 (10.3–11.3)	100 (100–100)	57.2 (56.6–57.7)
Tübingen	196	DSPO	51.4 (50.3–52.6)	71.2 (70.6–71.8)	40.2 (39.2–41.1)	80.5 (80.0–81.1)	66.2 (65.6–66.7)
		LR	61.8 (60.6–63.1)	71.2 (70.6–71.8)	32.4 (31.6–33.2)	89.1 (88.6–89.5)	69.6 (69.1–70.2)
		MEs	52.3 (51.1–53.6)	68.8 (68.2–69.4)	25.9 (24.9–26.8)	88.1 (87.6–88.6)	66.2 (65.7–66.7)
		DSD	70.9 (68.6–73.3)	67.4 (66.9–68.0)	10.6 (10.0–11.2)	97.6 (97.4–97.8)	67.1 (66.6–67.6)

Data were generated with bootstrap resampling with 150 replications to estimate the mean and the 95% confidence interval. Results are represented as percentages. *BWH*, Brigham and Women's Hospital; *CI*, confidence interval; *DSD*, disease-specific death; *DSPO*, disease-specific poor outcome; LR, local recurrence; *MEs*, major events; *NPV*, negative predictive value; *PPV*, positive predictive value.

cutaneous SCCs. Our current work provides additional information about the subgroup of cutaneous SCCs at highest risk of metastasis and death. However, given the cases on which it is based, data about sensitivity, specificity, negative predictive value, and positive predictive value may be not completely comparable with those of other studies. Our data reflect the significantly lower sensitivity, positive predictive value, and C-index for the high-risk cutaneous SCC subgroup, in which improvement is desirable.

From a broader perspective, the 3 alternative staging systems had greater prognostic accuracy than the official AJCC8. Indeed, if all the tumors had been classified the same, as would have occurred with the AJCC8 (because all tumors were T3-AJCC8), the C-index and concordance probability estimate would both have been 50% for all outcome events, implying that the system was no better at predicting outcomes in these tumors than if they had been randomly assigned. The Salamanca and Brigham and Women's Hospital systems slightly outperformed the Tübingen system, but all 3 would be more effective for classifying high-risk cutaneous SCCs than the AJCC8. The Tübingen system performed better when low- and high-risk categories were combined (Supplemental Fig 1). Thus, considering the prognostic effect of risk-factor accumulation, as with our T3-AJCC8 and the Brigham and Women's Hospital and Tübingen staging systems, would probably increase prognostic accuracy and improve risk stratification of high-risk cutaneous SCC in AJCC9.

All 3 alternative staging systems rely on prognoses that are based on knowledge of accumulated risk factors. The presence of multiple risk factors is associated with poorer prognosis than that arising from a single one[14,16] and implies a higher risk of a positive sentinel lymph node biopsy result.[22] These alternative staging systems therefore highlight the existence of subgroups of high-risk cutaneous SCC in which sentinel lymph node biopsies, adjuvant therapy, or both would be more useful. Nevertheless, prognostic accuracy is relatively low, meaning that risk stratification of high-risk cutaneous SCCs still needs improvement.

This retrospective study is limited by differential loss to follow-up, and possibly lower-quality data, because records were not collected specifically for the purpose of this study. Standardized pathology records are needed to stratify cutaneous SCCs. Indeed, several high-risk features of proven prognostic significance (eg, desmoplasia) are not regularly recorded, and other high-risk features (eg, perineural invasion, tumor thickness) are not recorded in a standardized, comprehensive manner. Although this work focuses on high-risk cutaneous SCC, there were no T4-AJCC8 cases in this cohort. Such cases are rare and their evaluation requires a multicenter approach. Another limitation is that this was a single-center study from a University Hospital, involving a cohort that was older, on average, than others,[6,8,23,24] although competing risk analysis accounts for age-related bias. Finally, our study did not consider AJCC8 low-risk tumors because clinical decisions about them are usually less controversial, and because no proposals for refining low AJCC8 T stages have yet been published, to our knowledge.

Until the AJCC staging system is upgraded, these alternative systems offer a more accurate way of classifying most high-risk cutaneous SCCs and

identifying subgroups of higher-risk patients within this staging category, for whom specific clinical approaches may be considered. However, the prognostic accuracy of all 3 alternative high-risk cutaneous SCC staging systems is low, so better stratification of high-risk cutaneous SCC is desirable. By considering the accumulation of risk factors and incorporating new risk factors, including genomic data, better stratification of cutaneous SCC patients at high risk of recurrence is highly feasible.

Conflicts of interest

None disclosed.

REFERENCES

1. Stratigos A, Garbe C, Lebbe C, et al. Diagnosis and treatment of invasive squamous cell carcinoma of the skin: European consensus-based interdisciplinary guideline. *Eur J Cancer*. 2015;51:1989-2007.
2. National Comprehensive Cancer Network. Squamous Cell Skin Cancer. NCCN Clinical Practice Guidelines in Oncology. 2019.
3. Cancer Council Australia Keratinocyte Cancers Guideline Working Party. Clinical practice guidelines for keratinocyte cancer. Sydney: Cancer Council Australia. [Version URL: https://wiki.cancer.org.au/australiawiki/index.php?oldid=213931, cited 2021 Jan 4]. https://wiki.cancer.org.au/australia/Guidelines:Keratinocyte_carcinoma. Accessed January 4, 2021.
4. Conde-Ferreiros A, Corchete LA, Puebla-Tornero L, et al. Definition of prognostic subgroups in the T3-AJCC8 stage for cutaneous squamous cell carcinoma: tentative T3 stage subclassification. *J Am Acad Dermatol*. 2020. https://doi.org/10.1016/j.jaad.2020.03.088
5. Ruiz ES, Karia PS, Besaw R, Schmults CD. Performance of the American Joint Committee on Cancer Staging Manual, 8th Edition vs the Brigham and Women's Hospital tumor classification system for cutaneous squamous cell carcinoma. *JAMA Dermatol*. 2019;155:819-825.
6. Brantsch KD, Meisner C, Schonfisch B, et al. Analysis of risk factors determining prognosis of cutaneous squamous-cell carcinoma: a prospective study. *Lancet Oncol*. 2008;9:713-720.
7. Breuninger H, Brantsch K, Eigentler T, Hafner HM. Comparison and evaluation of the current staging of cutaneous carcinomas. *J Dtsch Dermatol Ges*. 2012;10:579-586.
8. Roscher I, Falk RS, Vos L, et al. Validating 4 staging systems for cutaneous squamous cell carcinoma using population-based data: a nested case-control study. *JAMA Dermatol*. 2018;154:428-434.
9. Eigentler TK, Leiter U, Hafner HM, Garbe C, Rocken M, Breuninger H. Survival of patients with cutaneous squamous cell carcinoma: results of a prospective cohort study. *J Invest Dermatol*. 2017;137:2309-2315.
10. Jambusaria-Pahlajani A, Kanetsky PA, Karia PS, et al. Evaluation of AJCC tumor staging for cutaneous squamous cell carcinoma and a proposed alternative tumor staging system. *JAMA Dermatol*. 2013;149:402-410.
11. Karia PS, Han J, Schmults CD. Cutaneous squamous cell carcinoma: estimated incidence of disease, nodal metastasis, and deaths from disease in the United States, 2012. *J Am Acad Dermatol*. 2013;68:957-966.
12. Karia PS, Jambusaria-Pahlajani A, Harrington DP, Murphy GF, Qureshi AA, Schmults CD. Evaluation of American Joint Committee on Cancer, International Union Against Cancer, and Brigham and Women's Hospital tumor staging for cutaneous squamous cell carcinoma. *J Clin Oncol*. 2014;32:327-334.
13. Amin MB, Edge S, Greene F, et al., eds. *AJCC Cancer Staging Manual*. 8th ed. Springer; 2017.
14. Canueto J, Burguillo J, Moyano-Bueno D, et al. Comparing the eighth and the seventh editions of the American Joint Committee on Cancer staging system and the Brigham and Women's Hospital alternative staging system for cutaneous squamous cell carcinoma: implications for clinical practice. *J Am Acad Dermatol*. 2019;80:106-113.e2.
15. Canueto J, Roman-Curto C. Novel additions to the AJCC's new staging systems for skin cancer. *Actas Dermosifiliogr*. 2017;108:818-826.
16. Karia PS, Morgan FC, Califano JA, Schmults CD. Comparison of tumor classifications for cutaneous squamous cell carcinoma of the head and neck in the 7th vs 8th edition of the AJCC cancer staging manual. *JAMA Dermatol*. 2018;154:175-181.
17. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data. 1980.
18. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94:496-509.
19. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA*. 1982;247:2543-2546.
20. Gönen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika*. 2005;92:965-970.
21. Blechman AB, Carucci JA, Stevenson ML. Stratification of poor outcomes for cutaneous squamous cell carcinoma in immunosuppressed patients using the American Joint Committee on Cancer Eighth Edition and Brigham and Women's Hospital staging systems. *Dermatol Surg*. 2019;45:1117-1124.
22. Mooney CP, Martin RCW, Dirven R, et al. Sentinel node biopsy in 105 high-risk cutaneous SCCs of the head and neck: results of a multicenter prospective study. *Ann Surg Oncol*. 2019;26:4481-4488.
23. Schmults CD, Karia PS, Carter JB, Han J, Qureshi AA. Factors predictive of recurrence and death from cutaneous squamous cell carcinoma: a 10-year, single-institution cohort study. *JAMA Dermatol*. 2013;149:541-547.
24. Mourouzis C, Boynton A, Grant J, et al. Cutaneous head and neck SCCs and risk of nodal metastasis - UK experience. *J Craniomaxillofac Surg*. 2009;37:443-447.