BJD
British Journal of Dermatology

Validation of four cutaneous squamous cell carcinoma staging systems using nationwide data*

Zoe Claire Venables,[1,2] Selin Tokez,[3] Loes M. Hollestein,[3,4] Antien L. Mooyaart,[5] Renate Ruthvan den Bos,[3] Brian Rous,[2] Irene M. Leigh,[6] Tamar Nijsten[3] and Marlies Wakkee[3]

[1]Department of Dermatology, Norfolk and Norwich University Hospital, Norwich,NR4 7UY, UK
[2]Public Health England, West Wing, Victoria House, Capital Park, Fulbourn, Cambridge, CB21 5XA, UK
[3]Departments of, Dermatology, Erasmus MC Cancer Institute
[4]Department of Research & Development, Netherlands Comprehensive Cancer Organization (IKNL), PO Box 19079, Utrecht, DB, 3501, the Netherlands
[5]Pathology, Erasmus University Medical Center, PO Box 2040, Rotterdam, CA, 3000, the Netherlands
[6]Barts and the London School of Medicine and Dentistry, London, UK

Linked Comment: P. Gjersvik. Br J Dermatol 2022; **186**:763.

Summary

Background Cutaneous squamous cell carcinoma (cSCC) is the second most common cancer worldwide with relatively low metastatic potential (2–5%). Developments in therapeutic options have highlighted the need to better identify high-risk patients who could benefit from closer surveillance, adjuvant therapies and baseline/follow-up imaging, while at the same time safely omitting low-risk patients from further follow-up. Controversy remains regarding the predictive performance of current cSCC staging systems and which methodology to adopt.

Objectives To validate the performance of four cSCC staging systems [American Joint Committee on Cancer 8th edition (AJCC8), Brigham and Women's Hospital (BWH), Tübingen and Salamanca T3 refinement] in predicting metastasis using a nationwide cohort.

Methods A nested case–control study using data from the National Disease Registration Service, England, 2013–2015 was conducted. Metastatic cSCC cases were identified using an algorithm to identify all potential cases for manual review. These were 1 : 1 matched on follow-up time to nonmetastatic controls randomly selected from 2013. Staging systems were analysed for distinctiveness, homogeneity, monotonicity, specificity, positive predictive value (PPV), negative predictive value (NPV) and c-index.

Results We included 887 metastatic cSCC cases and 887 nonmetastatic cSCC controls. The BWH system showed the highest specificity [92.8%, 95% confidence interval (CI) 90.8–94.3%, PPV (13.2%, 95% CI 10.6–16.2) and c-index (0.84, 95% CI 0.82–0.86). The AJCC8 showed superior NPV (99.2%, 95% CI 99.2–99.3), homogeneity and monotonicity compared with the BWH and Tübingen diameter and thickness classifications ($P < 0.001$). Salamanca refinement did not show any improvement in AJCC8 T3 cSCC staging.

Conclusions We validated four cSCC staging systems using the largest nationwide dataset of metastatic cSCC so far. Although the BWH system showed the highest overall discriminative ability, PPV was low for all staging systems, which shows the need for further improvement and refining of current cSCC staging systems.

Correspondence
Marlies Wakkee.
Email: m.wakkee@erasmusmc.nl

Accepted for publication
25 November 2021

Funding sources
No external funding.

Conflicts of interest
M.W. participated as an advisory board member on advanced cutaneous squamous cell carcinoma for Sanofi Genzyme for which she received a financial reimbursement. All other authors have no conflicts of interest to disclose.

Data availability statement
The data that support the findings of this study are available from the National Cancer Registration and Analysis Service (NCRAS), run by Public Health England. Restrictions apply to the availability of these data, which were used under licence for this study. Data are available from the authors with the permission of NCRAS.

Z.C.V. and S.T. contributed equally.

*Plain language summary available online

DOI 10.1111/bjd.20909

Cutaneous squamous cell carcinoma (cSCC) is the second commonest cancer worldwide.[1–3] While the majority has an excellent post-surgical prognosis, a small subset (2–5%) of tumours metastasizes.[4–8] A low-risk cSCC may require no clinical follow-up or further investigations whereas a high-risk cSCC may be considered for intense surveillance, imaging,

sentinel lymph node biopsy or even adjuvant therapy. With the development of targeted immunotherapies, accurate identification of high-risk patients becomes even more important for treatment, clinical trials and healthcare planning.[9]

The most widely used staging system is the American Joint Committee on Cancer 8th edition (AJCC8).[10] Due to the suboptimal performance of its previous 7th edition,[11] the Brigham and Women's Hospital (BWH)[12] and Tübingen University[13] (i.e. Breuninger) staging systems were developed. Comparative studies on their predictive performances have mostly been limited to single academic centre data.[14,15] Only one study has validated AJCC8, BWH and Tübingen staging systems using population-based data but the number of metastatic cSCCs were relatively small ($n = 103$).[16] Recently, aimed at further refining the AJCC8 T3 stage to reduce prognostic heterogeneity, the Salamanca T3 classification was developed.[17] In the present study, we aimed to validate the predictive performances of the AJCC8, BWH, Tübingen and Salamanca staging systems in nationwide cancer registry data from England, comprising the largest sample of metastatic cSCC so far.

Patients and methods

Patient population

Nationwide data from all patients with a histopathologically confirmed primary cSCC diagnosed between 1 January 2013 and 31 December 2015 were retrieved from the National Disease Registration Service (NDRS), England.[4] Metastatic cSCC cases were identified from NDRS data using a verified algorithm based on identifying free-text key words, treatment codes and mortality data which has been published previously.[4] All pathology reports identified by the algorithm were reviewed by a dermatologist (Z.C.V.) with a second opinion from a pathologist (B.R.) when required. We included all patients who developed metastatic cSCC for whom an excision biopsy pathology report was available for the identified primary tumour source of metastasis, where the primary cSCC occurred between 2013 and 2015, and which included at least three of the following variables: diameter, thickness, depth of invasion or Clark level. SCCs from genital or oral mucosa were excluded. Pathology reports for diagnostic biopsies or shave/curettage only were excluded due to lack of required reported variables and potentially unreliable representation of the tumour. For multiple potential primary cSCCs, the primary site was chosen based on clinical judgement using the information from the pathology report (topography, lymphatic drainage, lack of presence of other potential primary sources) and the time from primary tumour diagnosis to metastasis.

To obtain a nested case–control design, we randomly selected control patients with a cSCC diagnosis in 2013 from the same dataset that did not develop a metastasis during the study period and matched cases and controls on follow-up time. Ethical approval and informed consent were not required for this study as per Section 251 of the National Health Service Act 2006.[18]

Staging systems

cSCCs of cases and controls were classified according to the AJCC8, BWH, Tübingen and Salamanca refinement staging systems (Table S1; see Supporting Information). All required tumour characteristics were extracted from pathology reports and comprised a localization at the head and neck (with eyelid, ear and lip, in particular) and other body sites, tumour diameter in centimetres, tumour thickness in millimetres, differentiation grade (good/moderate vs. poor/undifferentiated), Clark level (II–III vs. IV–V), invasion beyond subcutaneous fat (yes/no), bone invasion (yes/no), perineural invasion (yes/no), desmoplastic tumour morphology (yes/no) and immunosuppression (yes/no). If Clark level was not stated, it was assumed as not invading beyond subcutaneous fat. The method used for measuring thickness was not routinely reported in the pathology reports and was therefore assumed to be standardized to the Royal College of Pathologists (RCPath) guidance (i.e. measured from the adjacent normal granular layer).[19] Diameter of nerve invasion was not routinely reported and therefore if perineural invasion was reported as present, it was assumed to meet the diameter criterion of ≥ 0.1 mm as per the AJCC8 and BWH systems. To assess for immunosuppression, registry data and Hospital Episode Statistics were analysed for diagnosis/operation codes associated with haematological malignancies and solid organ transplantations before the date of primary tumour diagnosis or within 6 months.

Statistical analysis

The AJCC8 staging system includes only head and neck cSCCs (excluding eyelid) and the Salamanca refinement only AJCC8 T3 cSCCs. To have fair comparisons, all analyses were performed in: (i) cSCCs of all body sites; (ii) head and neck cSCCs excluding eyelid; and (iii) AJCC8 T3 cSCCs. Each cSCC was scored according to the T-stage criteria of all four staging systems to evaluate their predictive performances in terms of distinctiveness (outcome differences between categories within a staging system), homogeneity (outcome similarity within categories between staging systems) and monotonicity (outcome worsening with increasing categories between staging systems, equal to the sensitivity). Conditional logistic regression analyses were performed to obtain odds ratios (ORs) with 95% confidence intervals (CIs) on the metastasis outcome per T stage. Missing values were imputed 20 times using multivariate imputation by chained equations with predicted mean matching. The imputation model included all covariates, the outcome and ethnicity and deprivation as auxiliary variables. To evaluate homogeneity, the proportion of metastases occurring in the low T stages (T1/T2 for AJCC8, T1/T2a for BWH, no- and low-risk groups for Tübingen, and T3a for Salamanca refinement) were compared between the staging systems using the McNemar test. To evaluate monotonicity, this has been done for the proportion of metastases occurring in the high T

stages. A false discovery rate of 5% was used for homogeneity and monotonicity to correct for multiple comparisons.[20] Specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated to assess classification performances, where PPV and NPV have been adjusted for a 2% metastasis prevalence in the general population.[21] Harrell's concordance index (c-index) was calculated, which is equal to the area under the receiver operating characteristic curve for binary outcomes in logistic regression models, and represents the probability that a patient who will develop metastasis will be assigned a higher risk score of this event by the model/staging system compared with a patient who will remain event free.[22] An erratum of the AJCC8 classification was published in February 2018 downgrading cSCCs of 2 cm without other high-risk criteria to T1 and 4 cm to T2.[23] Because all major studies on cSCC staging systems have used the initial AJCC8 criteria without including the erratum version, we did so as well in our study to enable comparability across the published literature. However, we additionally computed the c-index and number of tumours per T stage for the AJCC8 using the erratum criteria to assess whether this would change the results considerably. All analyses were performed using SPSS 25.0 statistical software (SPSS Inc., Chicago, IL, USA). Tests were two-sided at a 5% statistical significance level.

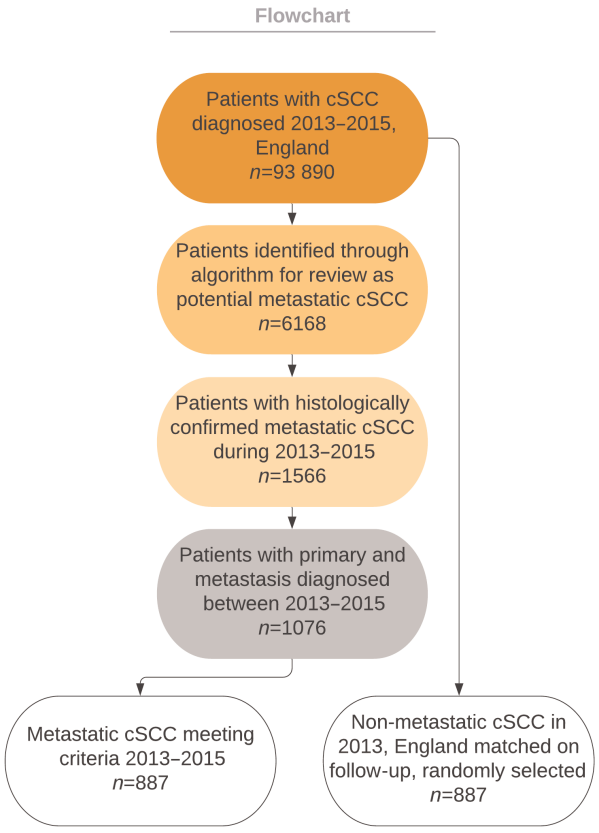

Figure 1 Flowchart describing the retrieval of the cases and controls in our study population. cSCC, cutaneous squamous cell carcinoma

Results

In total, 887 metastatic cases and 887 nonmetastatic controls met the inclusion criteria (Figure 1). The median follow-up time to metastasis was 183 days [interquartile range (IQR), 78–336] for cases. Cases had more unfavourable tumour characteristics than controls (Table 1). The majority of controls fell in the T1 category of the AJCC8 (70%) or BWH (67%) classification (Table 2). Following the Tübingen classification, most controls were present in the low-risk group for diameter (85%), no- and low-risk groups for thickness (together 86%) and low-risk group for the co-risk factors (71%). Cases were most commonly AJCC8 T3 tumours (68%, n = 604) and secondly T1 tumours (19%, n = 172). Only 14 (2%) cases and one (0.1%) control were T4a/b tumours.

Using the BWH classification, about half of all metastatic cases (46%, n = 411) belonged to the T2b group with a much smaller proportion of 8% (n = 71) in the T3 category. Following the Tübingen classification, the highest percentage of cases was found in the high-risk group for co-risk factors (68%) with a less distinct metastatic proportion among the high-risk diameter and thickness groups (60% and 58%, respectively).

Predictive performance – cutaneous squamous cell carcinomas of all body sites

The risk of metastasis increased with increasing T stage for the AJCC8 and BWH staging systems: OR of 3.9 (95% CI 2.6–5.8) for T2 and OR of 11.6 (95% CI 8.3–16) for T3 cSCCs compared with T1 cSCCs using AJCC8 (Table 2). Using the BWH classification, the risk of metastasis was seven-fold for T2a (OR 6.8, 95% CI 4.6–10.1) and 33-fold for T2b (OR 33.3, 95% CI 20.8–53.2) compared with T1 cSCCs. Following the Tübingen classification, the highest metastasis risk was captured by the high-risk group for tumour thickness: OR of 36 (95% CI 20.8–62.3).

The AJCC8 staging system showed superior homogeneity and monotonicity compared with both BWH and Tübingen tumour diameter and tumour thickness classifications whereas these outcome measures were not different compared with the Tübingen co-risk factors classification [Table 3; see Table S2 (see Supporting Information) for P values]. As second best, the Tübingen tumour diameter and co-risk factors classifications showed superior monotonicity and homogeneity compared with the BWH classification. Comparing the three Tübingen classifications between themselves produced the best homogeneity and monotonicity for the Tübingen co-risk factors classification.

The BWH system had the highest specificity (92.8%, 95% CI 90.8–94.3%), PPV (13.2%, 95% CI 10.6–16.2%) and c-index (0.84, 95% CI 0.82–0.86) of all staging systems. NPV was highest for the AJCC8 classification (99.2%, 95% CI 99.2–99.3) (Table 4).

The use of the AJCC8 erratum criteria did not change the c-index (i.e. 0.78, 95% CI 0.76–0.80) and resulted in restaging of only 3.3% of the total cohort with changes seen in

Table 1 Descriptive characteristics of the 1774 patients with a primary cSCC, stratified by metastasis outcome

Characteristic	Total group, n = 1774 (%)	Metastatic cases, n = 887 (%)	Nonmetastatic controls, n = 887 (%)	P value
Sex				
Male	1264 (71.3)	696 (78.5)	568 (64.0)	< 0.001
Female	510 (28.7)	191 (21.5)	319 (36.0)	
Age, median (IQR)	80.4 (73.1–86.5)	80.8 (73.4–86.6)	79.7 (72.7–86.4)	0.41
Immunosuppressed				
No	1601 (90.2)	791 (89.2)	810 (91.3)	0.13
Yes	173 (9.8)	96 (10.8)	77 (8.7)	
Site of primary cSCC				
Head and neck	1258 (70.9)	690 (77.8)	568 (64.0)	< 0.001
Eyelid	22 (1.7)	13 (1.9)	9 (1.6)	
Ear	292 (23.2)	190 (27.5)	102 (18.0)	
Lip	99 (7.9)	61 (8.8)	38 (6.7)	
Other sites	513 (28.9)	197 (22.2)	316 (35.6)	
Unknown	3 (0.2)	0 (0.0)	3 (0.3)	
Tumour diameter				
< 2 cm	976 (55.0)	287 (32.4)	689 (77.7)	< 0.001
2–4 cm	533 (30.0)	386 (43.5)	147 (16.6)	
≥ 4 cm	211 (11.9)	188 (21.2)	23 (2.6)	
Unknown	54 (3.0)	26 (2.9)	28 (3.2)	
Tumour thickness				
≤ 2 mm	273 (15.4)	24 (2.7)	249 (28.1)	< 0.001
2–6 mm	797 (44.9)	332 (37.4)	465 (52.4)	
> 6 mm	553 (31.2)	465 (52.4)	88 (9.9)	
Unknown	151 (8.5)	66 (7.4)	85 (9.6)	
Differentiation grade				
Good/moderate	1159 (65.3)	402 (45.3)	757 (85.3)	< 0.001
Poor	588 (33.1)	474 (53.4)	114 (12.9)	
Unknown	27 (1.5)	11 (1.2)	16 (1.8)	
Clark level				
II/III	134 (7.6)	15 (1.7)	119 (13.4)	< 0.001
IV/V	1243 (70.1)	755 (85.1)	488 (55.0)	
Unknown	397 (22.4)	117 (13.2)	280 (31.6)	
Invasion beyond subcutaneous fat				
No	1533 (86.4)	668 (75.3)	865 (97.5)	< 0.001
Yes	239 (13.5)	217 (24.5)	22 (2.5)	
Unknown	2 (0.1)	2 (0.2)	0 (0.0)	
Perineural invasion				
No	1323 (74.6)	581 (65.5)	742 (83.7)	< 0.001
Yes	243 (13.7)	207 (23.3)	36 (4.1)	
Unknown	208 (11.7)	99 (11.2)	109 (12.3)	
Desmoplastic morphology				
No	1772 (99.9)	887 (100.0)	885 (99.8)	0.16
Yes	2 (0.1)	0 (0.0)	2 (0.2)	

cSCC, cutaneous squamous cell carcinoma; IQR, interquartile range.

metastatic cases T1 (+2.6%), T2 (–2.4%) and T3 (–0.1%), and nonmetastatic cases T1 (+3.7%), T2 (–3.5%) and T3 (–0.1%).

Predictive performance – head and neck cutaneous squamous cell carcinomas

Restricting to head and neck cSCCs resulted in 422 eligible cases and 422 controls. While the AJCC8 has been developed for head and neck cSCCs only, the staging system performed worse when fitted on this subgroup compared with all body sites: homogeneity 33.2%, monotonicity 66.8%, specificity 79.1% (95% CI 75.0–82.9%), PPV 6.1% (95% CI, 5.1–7.4), NPV 99.2% (95% CI, 99.0–99.3) and c-index of 0.76 (95% CI, 0.73–0.79) (Tables S3–S5; see Supporting Information). Use of the AJCC8 erratum criteria produced the same c-index (0.76, 95% CI 0.72–0.79) values. Worse results were also found for the other staging systems except for the Tübingen diameter classification, which showed a slightly higher specificity and PPV and equal c-index in head and neck cSCCs compared with cSCCs of all body sites.

Table 2 Conditional logistic regression analyses between the AJCC8, BWH and Tübingen staging systems and the metastasis outcome (distinctiveness)

	OR (95% CI) for metastasis	With metastasis, n = 887 (%)	Without metastasis, n = 887 (%)	P value ORs
AJCC8				
T1	1.0	172 (19.4)	620 (69.9)	
T2	3.9 (2.6–5.8)	97 (10.9)	94 (10.6)	< 0.001
T3	11.6 (8.3–16.0)	604 (68.1)	172 (19.4)	< 0.001
T4a/T4b	NA	14 (1.6)	1 (0.1)	
BWH				
T1	1.0	111 (12.5)	596 (67.2)	
T2a	6.8 (4.6–10.1)	294 (33.1)	227 (25.6)	< 0.001
T2b	33.3 (20.8–53.2)	411 (46.3)	61 (6.9)	< 0.001
T3	NA	71 (8.0)	3 (0.3)	
Tübingen				
Diameter (cT)				
Low risk (≤ 2 cm)	1.0	355 (40.0)	754 (85.0)	
High risk (> 2 cm)	8.0 (6.0–10.6)	532 (60.0)	133 (15.0)	< 0.001
Thickness				
No risk (≤ 2 mm)	1.0	29 (3.3)	264 (29.8)	
Low risk (2–6 mm)	6.0 (3.7–9.8)	347 (39.1)	499 (56.3)	< 0.001
High risk (> 6 mm)	36.0 (20.8–62.3)	512 (57.7)	124 (14.0)	< 0.001
Co-risk factors				
Low risk	1.0	285 (32.1)	627 (70.7)	
High risk	5.1 (4.0–6.5)	602 (67.9)	260 (29.3)	< 0.001

AJCC8, American Joint Committee on Cancer 8th edition; BWH, Brigham and Women's Hospital; CI, confidence interval; OR, odds ratio.

Predictive performance – AJCC8 T3 cutaneous squamous cell carcinomas

To validate the Salamanca T3 refinement staging, only AJCC8 T3 cSCCs were included, resulting in 37 eligible cases and 37 controls (Tables S6–S8; see Supporting Information). Although not significant, metastasis risk in the T3b stage was not higher compared with T3a stage (OR of 0.7, 95% CI 0.2–2.0, P = 0.47). The T3c stage showed a nonsignificant trend towards a higher risk of metastasis compared with T3a (OR 2.9, 95% CI 0.8–10.9, P = 0.12). In terms of homogeneity, monotonicity, PPV, NPV and c-index, the Salamanca T3 refinement performed worse than all other staging systems.

Table 3 Homogeneity and monotonicity of the AJCC8, BWH and Tübingen staging systems.

T stage per staging system	Metastasis (n = 887), n (%)
Evaluation of homogeneity: proportion of metastases occurring in low T stages between the three staging systems	
AJCC8 T1+T2	269 (30.3)
BWH T1+T2a	405 (45.7)
Tübingen tumour diameter, low risk	355 (40.0)
Tübingen tumour thickness, no and low risk	376 (42.4)
Tübingen co-risk factors, low risk	285 (32.1)
Evaluation of monotonicity: proportion of metastases occurring in high T stages between the three staging systems	
AJCC8 T3+T4	618 (69.7)
BWH T2b+T3	482 (54.3)
Tübingen tumour diameter, high risk	532 (60.0)
Tübingen tumour thickness, high risk	512 (57.7)
Tübingen co-risk factors, high risk	602 (67.9)

AJCC8, American Joint Committee on Cancer 8th edition; BWH, Brigham and Women's Hospital.

Discussion

We describe the largest reported nationwide dataset used to validate four cSCC staging systems. The BWH system had highest specificity, PPV and c-index, while the AJCC8 system performed best in terms of NPV, homogeneity and monotonicity. Thus, when determining high-risk patients who might require close follow-up or adjuvant therapy, the BWH T2b/3 would be most appropriate, whereas in correctly identifying who could be safely omitted from follow-up visits, the AJCC8 T1 would fit best. Overall, the BWH showed the highest c-index, which is singularly the best indicator of the predictive capability of a staging system. However, the PPV of all staging systems was suboptimal (ranging from 5% to 13%) and is important for selecting patients for adjuvant (immuno) therapy and/or intensive surveillance. There remains a need to identify new patient and tumour characteristics that are associated with the likelihood of cSCC progression.

A prior study showed comparable results to ours in head and neck cSCCs with a higher c-index for BWH (0.91) than AJCC8

Table 4 Specificity, positive predictive value, negative predictive value and c-index of the AJCC8, BWH and Tübingen staging systems

Parameter	Value
Specificity, % (95% CI)	
AJCC8	80.5 (77.7–83.1)
BWH	92.8 (90.8–94.3)
Tübingen tumour diameter	85.0 (82.5–87.3)
Tübingen tumour thickness	86.0 (83.6–88.2)
Tübingen co-risk factors	70.7 (67.6–73.7)
Positive predictive value,[a] % (95% CI)	
AJCC8	6.8 (6.0–7.7)
BWH	13.2 (10.6–16.2)
Tübingen tumour diameter	7.6 (6.5–8.8)
Tübingen tumour thickness	7.8 (6.6–9.1)
Tübingen co-risk factors	4.5 (4.1–5.0)
Negative predictive value,[a] % (95% CI)	
AJCC8	99.2 (99.2–99.3)
BWH	99.0 (98.9–99.1)
Tübingen tumour diameter	99.1 (99.0–99.1)
Tübingen tumour thickness	99.0 (98.9–99.1)
Tübingen co-risk factors	99.1 (99.0–99.2)
Staging system, c-index (95% CI)[b]	
AJCC8	0.78 (0.76–0.80)
BWH	0.84 (0.82–0.86)
Tübingen diameter	0.73 (0.70–0.75)
Tübingen thickness	0.77 (0.75–0.79)
Tübingen co-risk factors	0.69 (0.67–0.72)

[a] For the positive and negative predictive values, we adjusted the calculation with a metastasis prevalence of 2%. [b]P value < 0.001 in each instance. AJCC8, American Joint Committee on Cancer 8th edition; BWH, Brigham and Women's Hospital; CI, confidence interval.

(0.84).[14] Although they found an equal NPV of 99% for both staging systems, specificity and PPV were higher for BWH as was also the case in our study. A limitation of this study was that part of the same data for BWH staging development was used with the original development dataset including 256 tumours in 237 patients with 25 metastatic cases from 1998 to 2005.[12] Validation against AJCC8 used a dataset including 680 tumours in 459 patients with 23 metastatic cases from 2000 to 2009 and appears to be sourced from the same single academic centre.[14] This could have led to selection bias, poor generalizability and optimistic results. Cañueto *et al.* concluded that the BWH did not have great advantages over the AJCC8 in staging head and neck cSCCs and many overlaps were observed in terms of homogeneity and monotonicity.[15] This could be due to the relatively small sample size (n = 186) and zero to very few tumours included in the AJCC8 T4 and T2 categories, respectively, as well as the BWH T3 category. Roscher *et al.* did use a population-based external cohort and found a c-index of 0.82 for the Tübingen classification compared with 0.81 for the BWH and 0.75 for the AJCC8.[16] However, for the AJCC8 staging, only head and neck cSCC were included, whereas the BWH and Tübingen classifications comprised cSCCs of all body locations, resulting in an unequal comparison. Besides, while a c-index rounded to 0.8 can be considered a high discriminative ability, the c-index evaluates only the ranking of case–control pairs. Metastases can still occur in low-risk T stages (e.g. 46% for BWH1/2a, Table 3 homogeneity analysis) and controls can be assigned to high-risk T stages. This will not affect the c-index as long as the case is ranked to a higher risk category than the control, which means that the c-index alone is not enough in assessing the performance of a staging system and the remaining outcome measures should also be considered.

For all three staging systems, the PPVs were quite low (4.5–13.2%), as we adjusted for a 2% metastasis prevalence in the general population. Higher PPVs in studies among high-risk populations should be interpreted with caution if no adjustment for a metastasis prevalence in the general population was made.[14,24]

A remarkable finding from our study was that when restricting to head and neck sites only, AJCC8 performed worse with a lower c-index compared with including all body sites. So, while AJCC8 staging is only advised for head and neck cSCCs, it could be used for all sites.

Our study comprises the first external validation of the Salamanca refinement system.[17] Although the Salamanca refinement system has been compared with the BWH and Tübingen systems before, this was performed on the same sample as the development of this staging system.[24] We did not observe any improvement in risk stratification for the Salamanca refinement system in our study. The Salamanca study combined the outcomes of metastasis and disease-specific death as 'major events' and included only 32 patients in this category, which may have limited the performance of the refinement. This once again highlights the importance of external validation of staging systems.

Tumour thickness has the greatest predictive value for metastasis,[6,25] yet it is not robustly utilized in AJCC8 classification, limited only to identifying tumours > 6 mm, and in BWH staging only considered as invasion beyond subcutaneous fat. In contrast, Breslow thickness has multiple categories in AJCC8 melanoma staging.[26] Similarly, differentiation grade was not included in AJCC8 despite poor differentiation being an important risk factor for worse outcomes.[8,25,27,28] This is probably related to the poor reproducibility of differentiation grade. There is an urgent need to better define differentiation grade and validate this among pathologists to be able to include this in staging systems in the future. The search for a more refined staging system also reflects the requirement for less rigid categorization into limited risk-profile groups. Comparable with melanoma, a nomogram model which offers a predictive risk calculation could be an attractive alternative.[29,30]

The strengths of this study are the use of nationwide cancer registry data producing generalizable results to other populations with similar characteristics and the conduct of the analyses in three different body-site samples that the individual staging systems had been developed on. This is relevant for clinical practice, where the staging systems are often used for cSCCs from all body sites. Limitations of our study include the restriction to information provided by pathology reports available at the NDRS. Although uniform reporting standards following the

RCPath guidance are assumed, we could not assess the histological slides ourselves in case of missing parameter values.[19] Also, given that our study period dated from before the introduction of the analysed staging systems, nerve diameter for perineural invasion assessment was not routinely documented. The retrieval of data on immunosuppression has not been validated before and is likely to be an underestimation. Finally, we included all patients with the primary and metastatic tumour identified between 2013 and 2015. We previously found that 85% of the cSCC metastases occur within 2 years and 93% within 3 years.[4] Our data may therefore exclude a minority of tumours which metastasize late.

In conclusion, we showed that the BWH and AJCC8 are both superior to the Tübingen and Salamanca T3 staging systems and that each staging system has its own strengths, with the BWH showing the highest overall discriminative ability. With emerging treatments being developed to treat advanced cSCC, identifying which tumours have the highest risk profile (i.e. PPV) is essential and currently needs improvement for all staging systems.[31] Due to the high volume of cSCCs, it is equally important to correctly identify those at low risk (i.e. NPV), thus enabling streamlined patient care. Possibly, better utilizing variables such as tumour thickness and differentiation grade may enhance outcome prediction. In the future, it would be optimal to have a single staging system that performs optimally for all key outcomes.

Acknowledgments

This work uses data that have been provided by patients and collected by the National Health Service as part of their care and support. The data are collated, maintained and quality assured by the National Disease Registration Service, which is part of Public Health England.

References

1 Lomas A, Leonardi-Bee J, Bath-Hextall F. A systematic review of worldwide incidence of nonmelanoma skin cancer. *Br J Dermatol* 2012; **166**:1069–80.
2 Tokez S, Hollestein L, Louwman M *et al.* Incidence of multiple vs first cutaneous squamous cell carcinoma on a nationwide scale and estimation of future incidences of cutaneous squamous cell carcinoma. *JAMA Dermatol* 2020; **156**:1300–6.
3 Venables ZC, Nijsten T, Wong KF *et al.* Epidemiology of basal and cutaneous squamous cell carcinoma in the U.K. 2013–15: a cohort study. *Br J Dermatol* 2019; **181**:474–82.
4 Venables ZC, Autier P, Nijsten T *et al.* Nationwide incidence of metastatic cutaneous squamous cell carcinoma in England. *JAMA Dermatol* 2019; **155**:298–306.
5 Tokez S, Wakkee M, Kan W *et al.* Cumulative incidence and disease-specific survival of metastatic cutaneous squamous cell carcinoma: a nationwide cancer registry study. *J Am Acad Dermatol* 2022; **86**:331–8.
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**:713–20.
7 Nelson TG, Ashton RE. Low incidence of metastasis and recurrence from cutaneous squamous cell carcinoma found in a UK population: do we need to adjust our thinking on this rare but potentially fatal event? *J Surg Oncol* 2017; **116**:783–8.
8 Schmults CD, Karia PS, Carter JB *et al.* Factors predictive of recurrence and death from cutaneous squamous cell carcinoma: a 10-year, single-institution cohort study. *JAMA Dermatol* 2013; **149**:541–7.
9 US Food and Drug Administration. FDA approves cemiplimab-rwlc for metastatic or locally advanced cutaneous squamous cell carcinoma. 28 September 2018. Available at: https://www.fda.gov/drugs/drug-approvals-and-databases/fda-approves-cemiplimab-rwlc-metastatic-or-locally-advanced-cutaneous-squamous-cell-carcinoma. (last accessed 12 January 2022).
10 Lydiatt WM, Patel SG, O'Sullivan B *et al.* Head and neck cancers – major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin* 2017; **67**:122–37.
11 Farasat S, Yu SS, Neel VA *et al.* A new American Joint Committee on Cancer staging system for cutaneous squamous cell carcinoma: creation and rationale for inclusion of tumor (T) characteristics. *J Am Acad Dermatol* 2011; **64**:1051–9.
12 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**:402–10.
13 Breuninger H, Brantsch K, Eigentler T *et al.* Comparison and evaluation of the current staging of cutaneous carcinomas. *J Dtsch Dermatol Ges* 2012; **10**:579–86.
14 Ruiz ES, Karia PS, Besaw R *et al.* 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**:819–25.
15 Cañueto 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**:106–13 e2. https://doi.org/10.1016/j.jaad.2018.06.060.
16 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**:428–34.
17 Conde-Ferreiros A, Corchete LA, Puebla-Tornero L *et al.* Definition of prognostic subgroups in the T3 stage of the eighth edition of the American Joint Committee on Cancer staging system for cutaneous squamous cell carcinoma: tentative T3 stage subclassification. *J Am Acad Dermatol* 2021; **85**:1168–77.
18 National Archives. National Health Service Act 2006. Available at: http://www.legislation.gov.uk/ukpga/2006/41/contents.
19 Slater D, Barrett P. Standards and datasets for reporting cancers. Standards and datasets for reporting cancers: Dataset for histopathological reporting of primary invasive cutaneous squamous cell carcinoma and regional lymph nodes. Royal College of Pathologists. February 2019. Available at: https://www.rcpath.org/uploads/assets/9c1d8f71-5d3b-4508-8e6200f11e1f4a39/Dataset-for-histopathological-reporting-of-primary-invasive-cutaneous-squamous-cell-carcinoma-and-regional-lymph-nodes.pdf (last accessed 24 January 2022).
20 Glickman ME, Rao SR, Schultz MR. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *J Clin Epidemiol* 2014; **67**:850–7.
21 MedCalc. Diagnostic test evaluation calculator. Available at: https://www.medcalc.org/calc/diagnostic_test.php (last accessed 12 January 2022).

22 Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; **15**:361–87.
23 American College of Surgeons. AJCC 8th Edition Updates and Corrections: https://www.facs.org/Quality-Programs/Cancer/AJCC/cancer-staging/updates-corrections. Accessed 8 October 2021.
24 Puebla-Tornero L, Corchete-Sanchez LA, Conde-Ferreiros A *et al.* Performance of Salamanca refinement of the T3-AJCC8 versus the Brigham and Women's Hospital and Tubingen alternative staging systems for high-risk cutaneous squamous cell carcinoma. *J Am Acad Dermatol* 2021; **84**:938–45.
25 Thompson AK, Kelley BF, Prokop LJ *et al.* Risk factors for cutaneous squamous cell carcinoma recurrence, metastasis, and disease-specific death: a systematic review and meta-analysis. *JAMA Dermatol* 2016; **152**:419–28.
26 Gershenwald JE, Scolyer RA, Hess KR *et al.* Melanoma staging: evidence-based changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin* 2017; **67**:472–92.
27 Rowe DE, Carroll RJ, Day CL Jr. Prognostic factors for local recurrence, metastasis, and survival rates in squamous cell carcinoma of the skin, ear, and lip. Implications for treatment modality selection. *J Am Acad Dermatol* 1992; **26**:976–90.
28 Brougham ND, Dennett ER, Cameron R *et al.* The incidence of metastasis from cutaneous squamous cell carcinoma and the impact of its risk factors. *J Surg Oncol* 2012; **106**:811–15.
29 Wong SL, Kattan MW, McMasters KM *et al.* A nomogram that predicts the presence of sentinel node metastasis in melanoma with better discrimination than the American Joint Committee on Cancer staging system. *Ann Surg Oncol* 2005; **12**:282–8.
30 El Sharouni MA, Ahmed T, Varey AHR *et al.* Development and validation of nomograms to predict local, regional, and distant recurrence in patients with thin (T1) melanomas. *J Clin Oncol* 2021; **39**:1243–52.
31 Migden MR, Rischin D, Schmults CD *et al.* PD-1 blockade with cemiplimab in advanced cutaneous squamous-cell carcinoma. *N Engl J Med* 2018; **379**:341–51.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Table S1–S8. Overview of the AJCC8, T3 Salamanca refinement, BWH and Tübingen cSCC staging systems

Table S2 McNemar P values comparing homogeneity and monotonicity between staging systems for cSCCs of all body sites

Table S3 Conditional logistic regression analyses between the AJCC8, BWH and Tübingen staging systems and the metastasis outcome (distinctiveness) for head and neck cSCCs

Table S4 Homogeneity and monotonicity of the AJCC8, BWH and Tübingen staging system for head and neck cSCCs

Table S5 Specificity, positive predictive value, negative predictive value and c-index of the AJCC8, BWH and Tübingen staging systems for head and neck cSCCs

Table S6 Conditional logistic regression analyses between the Salamanca T3 refinement, BWH and Tübingen staging systems and the metastasis outcome (distinctiveness) for AJCC8 T3 cSCCs

Table S7 Homogeneity and monotonicity of the Salamanca T3 refinement, BWH and Tübingen staging system for AJCC8 T3 cSCCs

Table S8 Specificity, positive predictive value, negative predictive value and c-index of the Salamanca T3 refinement, BWH and Tübingen staging systems for AJCC8 T3 cSCCs.