Received: 21 June 2021 | Accepted: 15 August 2021
DOI: 10.1111/jcpe.13542

ORIGINAL ARTICLE

Journal of Clinical Periodontology WILEY

External validation and comparison of the predictive performance of 10 different tooth-level prognostic systems

Muhammad H. A. Saleh[1,2] | Himabindu Dukka[2] | Giuseppe Troiano[3] | Andrea Ravidà[1] | Matthew Galli[1] | Musa Qazi[1] | Henry Greenwell[2] | Hom-Lay Wang[1]

[1]Department of Periodontics and Oral Medicine, University of Michigan School of Dentistry, Ann Arbor, Michigan, USA

[2]Department of Periodontics, University of Louisville School of Dentistry, Louisville, Kentucky, USA

[3]Department of Clinical and Experimental Medicine, University of Foggia, Foggia, Italy

Correspondence
Hom-Lay Wang, Department of Periodontics and Oral Medicine, University of Michigan School of Dentistry, 1011 North University Avenue, Ann Arbor, MI 48109-1078, USA.
Email: homlay@umich.edu

Abstract

Aim: Tooth-level prognostic systems can be used for treatment planning and risk assessment. This retrospective longitudinal study aimed to evaluate the prognostic performance of 10 different tooth-level risk assessment systems in terms of their ability to predict periodontal-related tooth loss (TLP).

Materials and methods: Data were retrieved retrospectively from patients who received surgical and non-surgical periodontal treatment. Data on medical history and smoking status at baseline and the last maintenance visit were collected. Ten tooth-level prognostic systems were compared using both univariate and multivariate Cox proportional hazard regression models to analyse the prognostic capability of each system for predicting TLP risk.

Results: One-hundred and forty-eight patients with 3787 teeth, followed-up for a mean period of 26.5 ± 7.4 years, were evaluated according to 10 different tooth-level prognostic systems, making up a total of 37,870 individual measurements. All compared prognostic systems were able to stratify the risk of TLP at baseline when different classes of association were compared. After controlling for maintenance, age, and gender, all systems exhibited excellent predictive capacity for TLP with no system scoring a Harrell's C-index less than 0.925.

Conclusions: All tooth-level prognostic systems displayed excellent predictive capability for TLP. Overall, the Miller and McEntire system may have shown the best discrimination and model fit, followed by the Nunn et al. system.

KEYWORDS
attachment loss, periodontitis, risk factor assessment, tooth loss, validation study

Clinical Relevance

Scientific rationale for study: The generalizability of tooth-level periodontal prognostic systems remains to be validated to determine whether tooth-based risk assessment systems assign a meaningful prognosis predictive of long-term tooth survival. The aim of this study was to externally validate 10 commonly used tooth-level prognostic systems in terms of their predictive capacity for risk of periodontal tooth loss (TLP). The secondary aim was to compare the prognostic accuracy of the systems.

Principal findings: All tested tooth-level prognostic systems displayed excellent predictive capability for TLP. Overall, the Miller and McEntire system showed the best discrimination and

model fit, followed by the Nunn et al. system. The Faggion et al. prognostic system showed the greatest inter-class discriminatory capacity when the highest and lowest severity classes were compared.

Practical implications: All 10 tooth-level prognostic systems displayed excellent predictive capability for tooth loss due to periodontitis.

1 | INTRODUCTION

"A coin toss would be an easier and more accurate way for the clinician to assign a prognosis under traditional guidelines, if the initial prognosis is less than good" stated McGuire and Nunn, emphasizing the low sensitivity of their prognostic model at 5 and 8 years (McGuire, 1991). Predicting whether periodontally involved teeth can be retained over the long term is crucial for treatment planning in periodontitis patients. Accurate risk assessment leads to less invasive plans, better therapeutic outcomes, and reduced long-term costs (Tan et al., 2016). In 2008, the American Academy of Periodontology (AAP) defined risk assessment "as the process by which qualitative or quantitative assessments are made of the likelihood for adverse events to occur as a result of exposure to specified health hazards or by the absence of beneficial influences" (AAP, 2008).

Broadly speaking, a risk factor is any characteristic, behaviour, or exposure associated with a particular disease (Schwendicke et al., 2018), whereas prognosis refers to the likely course and outcome of a disease with or without treatment, and a prognostic factor is any characteristic that relates to the likelihood of success or survival (Beck, 1998; Miller et al., 2014). It is important to note that in periodontal disease, prognostic factors can be systemic risk factors (e.g., smoking) or local in nature (e.g., furcation involvement) (Kwok & Caton, 2007). Emerging scientific evidence inspired the development of various periodontal risk assessment tools to calculate the probability of periodontitis progression (Lang & Tonetti, 1996; Lang et al., 2015) and tooth survival after therapy (Grant et al., 1979; Becker et al., 1984; McGuire, 1991; Checchi et al., 2002; Fardal et al., 2004; Faggion et al., 2007; Kwok & Caton, 2007; Avila et al., 2009; Nunn et al., 2012; Miller et al., 2014). Most of these assessment tools calculate the risk of tooth loss as either a continuous or an ordinal (categorical) variable (Schwendicke et al., 2018).

The use of descriptive terms such as good, fair, or questionable can be perceived subjectively by clinicians due to heterogeneity in how definitions are interpreted and applied (Faggion et al., 2007; Kwok & Caton, 2007; Miller et al., 2014). In addition, numerous risk assessment tools exist, and there is no single universally accepted set of criteria for establishing a case-specific periodontal prognosis (Nunn et al., 2012). Additionally, although many of these models have showed promising results in the populations they were originally developed in, external validity of prior findings remains to be tested on a new cohort (Kwok & Caton, 2007; Schwendicke et al., 2018). Such validation is important because of the lack of evidence to determine whether tooth-based risk assessment systems assign a meaningful prognosis capable of accurately predicting long-term outcomes and tooth survival.

Patient factors (smoking, diabetes) and tooth factors (furcation involvement, increased probing depth [PD]) can influence prognosis because they not only impact disease expression and progression but also interact with each other. Patient-level prognosis is paramount for delivery of optimum treatment, setting realistic expectations for the suggested therapy and determination whether subsequent involvement of other health care providers in the treatment plan is required. However, accurate tooth-level prognosis may better capture the short-term therapeutic outcomes and reduce the overall costs of treatment (Lang & Tonetti, 1996; Tan et al., 2016).

Hence, the aim of this study was to externally validate 10 commonly used tooth-level prognostic systems in regard to their predictive capacity for the risk of periodontal tooth loss (TLP). The secondary aim was to compare the prognostic accuracy of the systems.

2 | METHODOLOGY

This study was conducted in agreement with the Helsinki Declaration of 1975 (World Medical Association, 1975) as most recently revised in 2013 (World Medical Association, 2013). The study was approved by the University of Michigan Medical School Institutional Review Board (IRBMED) (study identifier: HUM00157260). The TRIPOD statement was taken as a reference for the choice of methodologies to be used in this study (Moons et al., 2015).

2.1 | Study population

The present data were retrospectively reviewed from electronic and physical charts of patients receiving periodontal treatment between January 1966 and January 2008 at the University of Michigan School of Dentistry, Ann Arbor, Michigan, USA.

Inclusion criteria were as follows:

- Patients treated for periodontal disease (either surgically or non-surgically) and maintained for ≥10 years at the University of Michigan School of Dentistry after active therapy.
- Patients receiving at least one visit of supportive periodontal therapy (SPT)/year throughout the entire follow-up period.
- Complete periodontal charts with clinical attachment level (CAL), bleeding on probing (BOP), and full-mouth radiographic series of diagnostic-quality radiographs (taken ≤12 months from the baseline periodontal examination).

- Complete medical history recorded at baseline examination.
- Patients whose teeth were extracted at the University of Michigan School of Dentistry.

Exclusion criteria were the following:

- The reason for tooth extraction not described in patient charts.
- Patients treated or maintained in any treatment facility other than the University of Michigan School of Dentistry.
- Smokers not reporting the number of cigarettes/day or time since they started smoking.
- Diabetic patients not reporting haemoglobin A1c (HbA1c) and/or plasma glucose levels at the baseline visits.
- Patients failing SPT for ≥1 year during the studied period.

2.2 | Data collection and patient classification

Physical and digital records of the patients who fitted our preset inclusion criteria were screened and evaluated by three examiners (MQ, AR, and MS). All data regarding relevant patient-level characteristics (age, gender, medical history, etc.), as well as frequency of SPT visits/year, were collected. Tooth-specific data on clinical parameters, such as PD, CAL, BOP, mobility, furcation involvement, and keratinized tissue width were collected from patient charts at T0 (time of active periodontal therapy) and T1 (last SPT visit). PD, CAL, and BOP were evaluated at six sites per tooth. Radiographic bone loss was calculated from either periapical or bitewing radiographs (Akesson et al., 1992; Pepelassi et al., 2000). Radiographic bone loss was measured as the distance from 2 mm below the cemento-enamel junction to the most apical extension of the defect. In case of a crown restoration, the most apical aspect of the restorative margin was taken as the coronal reference point for bone loss measurements. For molar teeth, the root with the most radiographic bone loss was chosen for analysis. The presence of inter-proximal restorations or crowns, periapical pathology, endodontic root fillings, endodontic posts, and vertical osseous defects were also collected from the radiographs.

A few prognostic systems incorporated diabetic status using a binary scale; only one system (Miller et al., 2014) required HbA1c levels. Patients reporting plasma glucose levels had their results converted to HbA1c percentage using estimated average glucose levels (Rohlfing et al., 2002; Rentfro et al., 2009). Cigarette consumption was also required by one study (binary scale) (Miller et al., 2014). Smoking was self-reported, and no stratification (heavy/light smoking) was attempted.

2.3 | Teeth allocation according to different prognosis systems

In total, 10 prognostic systems were included in this study (Grant et al., 1979; Becker et al., 1984; McGuire, 1991; Checchi et al., 2002; Fardal et al., 2004; Faggion et al., 2007; Kwok & Caton, 2007; Avila et al., 2009; Nunn et al., 2012; Miller et al., 2014). Information regarding the parameters involved in each prognostic system is provided in Table S1. Two authors (MS and HD) independently analysed the allocation of each tooth to a specific class for each prognostic system. In case of any event of ambiguity, a final decision was made in a joint session with a third author (HG).

2.4 | Statistical analysis

The following patient- and tooth-level variables were analysed: age, gender, tooth identifier, position (anterior/posterior), jaw (maxilla/mandible), time from baseline to the last follow-up, tooth status at the last follow-up visit (periodontal-related loss, loss for a non-periodontal cause, present), number of annual maintenance sessions from baseline to the last follow-up, and tooth-level membership class for each prognostic system. (Grant et al., 1979; Becker et al., 1984; McGuire, 1991; Checchi et al., 2002; Fardal et al., 2004; Faggion et al., 2007; Kwok & Caton, 2007; Avila et al., 2009; Nunn et al., 2012; Miller et al., 2014).

Survival analyses were performed, after checking for the presence of proportional hazard assumption (estat phtest in STATA), for TLP using both univariate and multivariate Cox regression frailty models that were built for each classifier. In the multivariate model, the tooth classifier memberships were included with potential confounding factors (age, gender, and the number of maintenance sessions underwent by the tooth during the whole follow-up). In order to analyse the prognostic performance of the different tooth-level prognostic systems, overall performance (Harrell's C-index and Royston's D-index) and model fit (Akaike information criterion [AIC] and Bayesian information criterion [BIC]) were measured for each Cox regression model (Rahman et al., 2017). In addition, a post hoc comparison was conducted using the Bonferroni test in order to evaluate intra-class stratification within the different prognostic systems. Ratios of restricted mean survival time (RMST) in the univariate analysis were also assessed for comparison with the reference class.

3 | RESULTS

3.1 | Patient cohort description and demographics

Totally, 148 patients (80 females; 68 males; 3787 teeth) with a mean age of 46.49 ± 11.53 years were included in the analysis. At baseline, 1886 maxillary and 1901 mandibular teeth were present. Of these, 2073 teeth were located posteriorly (molar and premolar regions) and 1714 were located anteriorly. The average follow-up of teeth included in the analysis was 297.7 ± 91.3 months (24.7 ± 7.6 years). A total of 425 teeth (11.2%) were lost throughout the follow-up period; of these, 179 (4.72%) were TLP. Characteristics of the patient sample are given in Table 1.

TABLE 1 Patient characteristics and patient-related parameters of the included sample

Demographic data and patient-related parameters	
N (subjects)	148
N (female)	80
N (male)	68
Age (years)	46.49 ± 11.53
Follow-up (months)	318.7 ± 89.5
Number of teeth at T0	3787
Number of maxillary teeth at T0	1886
Number of mandibular teeth at T0	1901
Number of posterior teeth	2073
Number of anterior teeth	1714

3.2 | Prognostic stratification of the different prognostic systems

All the models accurately stratified teeth based on the risk of TLP as shown in Table 2 and Figure 1. Of note, only the Fardal, Faggion, Nunn, and Miller and McEntire (Fardal et al., 2004; Faggion et al., 2007; Nunn et al., 2012; Miller et al., 2014) prognostic systems showed a lack of statistical significance between severity categories. These findings suggest that grouping multiple classes into a single class may simplify a model in many cases without strongly affecting its predictive capacity.

Noteworthy, the Faggion et al. score "7" had a hazard ratio (HR) = 1587 (95% confidence interval [CI]: 270–9313) in the multi-level univariate analysis and an HR = 40 (95% CI: 7.38–219) in the multi-level multivariate analysis (Faggion et al., 2007) (Table 2). It is important to note that the hazard ratios in Table 2 are reported only to show the results of the models, and should not be used to draw conclusion regarding the magnitude of effects since they are characterized by low power (Rulli et al., 2018).

3.3 | Comparison of the prognostic performance of different systems

Prognostic performance was evaluated by analysing the overall performance and model fit of univariate and multivariate multi-level Cox regression frailty models (Table 3). A higher Harrell's C-index and Royston's D-index in conjunction with lower AIC and BIC was indicative of a better prognostic performance.

As shown in Table 3, all prognostic systems had excellent model performance, with a Harrell's C-index score ranging from 0.925 to 0.949. The Miller and McEntire model seemed to show the best values for model performance and model fit after both univariate and multivariate analysis, followed by the Nunn and Avila models. Multivariate analysis at the tooth level indicated that the number of SPT visits influenced the survival of periodontally affected teeth, as the multivariate analysis showed better values for both overall performance and model fit for all prognostic systems (Table 3).

The same prognostic performance analysis was done using overall tooth loss (OTL) instead of TLP to determine whether OTL could be used as an alternative measure in similar studies. The results showed that the prognostic performance of all systems was significantly impaired when analyses were conducted with OTL instead of TLP in terms of overall performance and model fit for both univariate and multivariate multi-level Cox regression frailty models (Table S2).

However, as previously reported, the Miller and McEntire model (11 classes) showed a lack of statistical significance between different disease severity classes; for this reason, we tried collapsing the number of classes in the Miller and McEntire model from 11 to 3 to evaluate the prognostic performance of a simplified version. The simplified model showed good values for overall prognostic performance (Table S3). To overcome the contrast between different classes in each predictor system by comparing HR, an RMST analysis was added for the comparison with the reference class in the univariate analysis (Table S4).

4 | DISCUSSION

Although the compared prognostic systems were originally designed to predict periodontal tooth condition/survival, these systems have seldom been externally validated in the literature (Steyerberg & Harrell, 2016). The present study evaluated the accuracy of 10 different tooth-level prognostic systems for the prediction of TLP in a sample of 3787 teeth in 148 periodontitis patients, with a 26.5 ± 7.4-year followed-up period. Our results showed that all included systems accurately stratified teeth based on the risk of TLP at baseline when different classes of association were compared. Moreover, most of the classes of all prognostic systems showed statistically significant inter-class differences. That is, in most prognostic systems, the more severe classes were associated with a higher risk of TLP (Table 2). When the prognostic performance of all systems was compared, the Miller and McEntire system (Miller et al., 2014) seemed to show the best values for model performance and fit after both univariate and multivariate analysis, followed by the Nunn et al. model (Nunn et al., 2012).

The Miller & McEntire system was created retrospectively by learning patterns from the authors' own private practice data. Miller et al. initially assigned their scores to prognostic factors specific to a tooth type (molar teeth) to derive a quantitative scoring system (Miller et al., 2014). The authors constructed a tooth survival model which showed a 38% increase in the risk of tooth loss with each unit increase in the Miller and McEntire prognostic index (Miller et al., 2014). The present study used this system for both molar and non-molar teeth; the validity of this approach was endorsed through personal communication with the author himself. The present analysis externally validated that the Miller and McEntire system may have the highest predictive capacity, which showed the best values in terms of model performance and fit after both univariate and multivariate

TABLE 2 Univariate and multivariate risk stratification performed for periodontal-related tooth loss using multi-level Cox regression frailty models

Variables			Multi-level univariate analysis		Multi-level multivariate analysis	
			HR 95% (CI)	*p*-Value	HR 95% (CI)	*p*-Value
Fardal et al. (2004)	1 (Ref)	Good	1.00	–	1.00	–
	2	Uncertain	3.24 (2.02–5.18)	.001*	1.35 (0.81–2.24)	.249
	3	Poor	7.82 (3.85–15.90)	.000*	4.07 (1.80–9.19)	.000*
	4	Hopeless	26.41 (4.62–151.0)	.000*	20.05 (3.17–126.9)	.000*
Becker et al. (1984)	1 (Ref)	Good	1.00	–	1.00	–
	2	Questionable	4.90 (3.04–7.89)	.001*	3.90 (2.36–6.45)	.000*
	3	Hopeless	117.46 (29.6–462)	.001*	6.62 (2.19–20.1)	.001*
Grant et al. (1979)	1 (Ref)	Good	1.00	–	1.00	–
	2	Fair	1.32 (1.07–1.86)	.000*	1.50 (1.12–2.18)	.000*
	3	Questionable	3.88 (1.46–8.59)	.000*	3.65 (1.68–6.42)	.000*
	4	poor	14.9 (2.66–69.24)	.000*	16.1 (4.28–87.4)	.000*
Checchi et al. (2002)	1 (Ref)	Good	1.00	–	1.00	–
	2	Questionable	5.49 (3.54–8.51)	.000*	2.77 (1.73–4.43)	.000*
	3	Hopeless	38.66 (11.2–133.9)	.000*	8.16 (2.35–28.4)	.001*
Faggion et al. (2007)[a]	1 (Ref)	≥90%	1.00	–	1.00	–
	2	80%–90%	1.84 (0.97–3.50)	.061	1.44 (0.72–2.89)	.306
	3	70%–79%	5.05 (2.15–11.86)	.000*	2.69 (1.10–6.56)	.029*
	4	60%–69%	9.68 (3.84–24.36)	.000*	9.16 (3.21–26.17)	.000*
	5	50%–59%	25.95 (9.94–67.7)	.000*	10.9 (3.76–31.85)	.000*
	6	40%–49%	151.9 (14.3–1608)	.000*	23.8 (2.14–265)	.010*
	7	30%–39%	1587 (270–9313)	.000*	40.27 (7.38–219)	.000
McGuire (1991)	1 (Ref)	Good	1.00	–	1.00	–
	2	Fair	2.50 (1.61–3.89)	.000	1.58 (0.70–4.28)	.069*
	3	Poor	5.90 (3.30–10.55)	.000*	2.74 (1.42–5.28)	.003*
	4	Questionable	27.3 (14.4–52.1)	.000*	12.1 (5.78–25.5)	.000*
	5	Hopeless	66.6 (15.3–290)	.000*	17.9 (4.2–76.5)	.000*
Kwok and Caton (2007)	1 (Ref)	Favourable	1.00	–	1.00	–
	2	Questionable	3.58 (2.15–5.94)	.000*	2.07 (1.20–3.55)	.009*
	3	Unfavourable	16.59 (9.3–26.7)	.000*	8.81 (4.58–16.97)	.000*
	4	Hopeless	43.86 (10.3–187)	.000*	16.6 (4.05–68.5)	.000*
Avila et al. (2007)	1 (Ref)	Long-term maintenance favourable	1.00	–	1.00	–
	2	Proceed with caution	4.85 (3.07–7.67)	.000*	2.69 (1.61–4.47)	.000*
	3	Long-term survival unfavourable	18.4 (8.53–39.7)	.000*	8.83 (3.78–20.6)	.000*
Nunn et al. (2012)	1 (Ref)	Good	1.00	–	1.00	–
	2	Fair	2.52 (1.62–3.92)	.000*	1.66 (0.99–2.78)	.051
	3	Poor	6.44 (3.56–11.65)	.000*	2.73 (1.41–5.32)	.000*
	4	Questionable	23.3 (11.8–45.9)	.000*	10.7 (4.94–23.1)	.000*
	5	Hopeless	56.3 (18.3–173)	.000*	19.5 (6.1–63.2)	.000*
Miller et al. (2014)[b]	1 (Ref)	98%	1.00	–	1.00	–
	2	97%	1.61 (0.50–5.14)	.425	0.86 (0.23–3.19)	.822
	3	96%	3.47 (1.09–11.02)	.035*	1.34 (0.36–4.95)	.662
	4	95%	9.24 (2.97–28.7)	.000*	4.35 (1.21–15.6)	.024*
	5	93%	8.37 (2.56–27.3)	.000*	3.19 (0.84–12.2)	.088
	6	90%	20.3 (6.19–66.5)	.000*	8.21 (2.18–3.85)	.002*

(Continues)

TABLE 2 (Continued)

			Multi-level univariate analysis		Multi-level multivariate analysis	
Variables			HR 95% (CI)	*p*-Value	HR 95% (CI)	*p*-Value
	7	86%	43.5 (12.9–147.6)	.000*	9.49 (2.34–38.44)	.002*
	8	81%	68.4 (17.6–266)	.000*	11.7 (2.51–207)	.000*
	9	75%	34.9 (6.6–184.0)	.000*	34.2 (6.2–222)	.000*
	10	67%	110.6 (20.5–598)	.000*	37.2 (6.21–222)	.000*
	11	53%	37.1 (3.53–388)	.003*	25.1 (2.1–303)	.011*

[a]The score denotes a projected 11-year tooth survival rate. The system also has a score of 8 (20%–29%), and 9 (≤20%). The present patient cohort included no teeth with such disease severity.
[b]The score of each class in % denotes the projected 15-year tooth survival rate.
*Statistically significant difference.

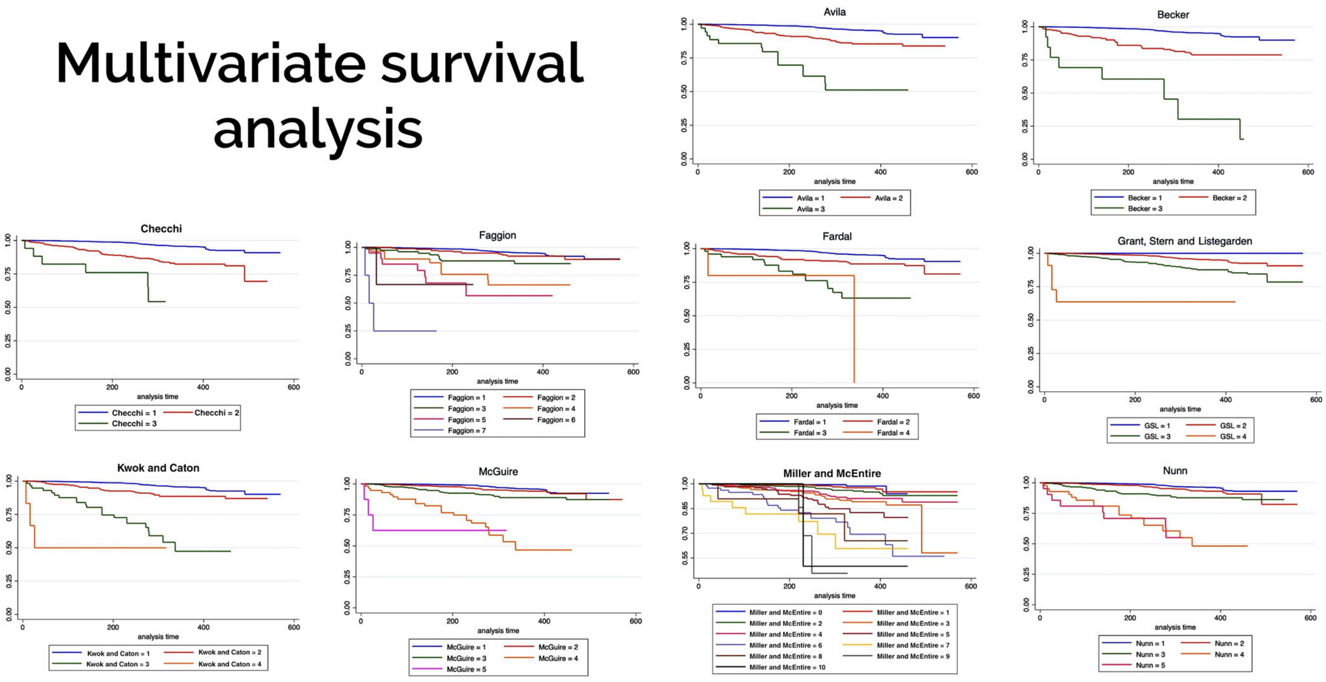

FIGURE 1 Survival curves based on the multi-level multivariate Cox regression analysis

analysis. A key feature of the Miller and McEntire prognostic system that may have contributed to its superior predictive capacity was the high emphasis placed on systemic factors. The highest score that a tooth could get based on this system was 20 points, and almost half of these (9 points) were dedicated to systemic factors (age, smoking, and diabetic status) (Levine & Miller, 2020). As stated before, with the exception of a single class in the Fardal, Faggion, and Nunn articles (Nunn, 2003; Fardal et al., 2004; Faggion et al., 2007), and 3/11 classes for the Miller and McEntire system (Miller et al., 2014), all prognostic systems showed statistically significant inter-class differences. Since the Miller and McEntire model involves 11 classes, we also modified the model by condensing the number of classes from 11 to 3. Interestingly, the abridged model still showed better overall prognostic performance than the other systems (except for the original Miller and McEntire system) (Table S2).

Faggion et al. (2007) used diabetes mellitus as one of three main parameters (in addition to tooth mobility and percentage of remaining alveolar bone height) to evaluate the risk of TLP. It is worth mentioning that this system showed the greatest difference between its lowest and highest prognostic classes (Table 2). The third included system that took systemic conditions into consideration was the Kwok and Caton system (Nguyen et al., 2020). This system proposed four prognostic categories based on the likelihood of controlling contributing local and systemic factors during treatment. The chief merit of this system was its ease of use, involving clinical judgement over evaluation of surrogate end points. The present data demonstrated that this ease of use can act like a double-edged sword. Although identifying the role of a particular risk factor/indicator is crucial, if the magnitude of that particular factor is not quantified and is instead left to

TABLE 3 Periodontal-related tooth loss

	Multi-level univariate Cox regression frailty models				Multi-level multivariate Cox regression frailty models			
Prognosis system	Harrell's C-index	Royston's D-index	Akaike's information criterion	Bayesian information criterion	Harrell's C-index	Royston's D-index	Akaike's information criterion	Bayesian information criterion
Fardal et al. (2004)	0.671	1.165	2586	2592	0.925	6.303	1913	1938
Becker et al. (1984)	0.615	1.404	2560	2566	0.927	6.333	1899	1924
Grant et al. (1979)	0.641	1.454	2567	2573	0.928	6.481	1902	1927
Checchi et al. (2002)	0.639	1.486	2504	2510	0.926	6.218	1864	1889
Faggion et al. (2007)	0.613	1.336	2483	2490	0.929	6.526	1844	1869
McGuire (1991)	0.696	1.587	2444	2450	0.931	6.617	1824	1849
Kwok and Caton (2007)	0.696	1.587	2444	2450	0.931	6.617	1824	1849
Avila et al. (2009)	0.640	1.464	2492	2498	0.928	6.353	1848	1873
Nunn et al. (2012)	0.695	1.600	2456	2462	0.932	6.584	1835	1860
Miller et al. (2014)	0.770	1.853	2266	2272	0.949	6.581	1690	1714

Note: Comparison of model risk stratification performance using measurements of model fit (Akaike's information criterion and Bayesian information criterion); and prognostic discrimination (Harrell's C-index and Royston's D-index). The higher the Harrell's C-index and Royston's D-index and the lower the AIC and BIC, the better the performance of the prognostic model.

subjective assessment, this may result in inaccurate risk allocation. It was demonstrated that even for expert clinicians, subjectively generated risk assessment may result in categorical mis-arrangement compared to relatively more objective risk assessment approaches (Persson et al., 2003).

McGuire's prognostic system has previously been validated; the authors found the system classes to be accurate for the good and hopeless categories, with low predictive accuracy for the poor and questionable categories (McGuire & Nunn, 1996). Numerous previous tooth-level prognostic validation studies achieved less than ideal results (Schwendicke et al., 2018). This can be attributed to three main reasons regarding the study design. The first is using OTL to validate prognostic systems that were intended to predict TLP. The key criteria for TLP (i.e., bone loss severity, residual PD ≥5 mm, smoking, and lack of compliance) have been found to be inconsistent with risk of OTL (Chambrone et al., 2010; Helal et al., 2019). OTL was found to include TLP plus 35%–80% TL from other reasons (i.e., fractures, caries, endodontic failure, and strategic extractions) (Al-Shammari et al., 2005; Chambrone et al., 2010). Ravida et al. (2020) demonstrated in a long-term retrospective study that patient-level risk assessment was actually prognostic for TLP but not OTL. Hence, the current study only included TLP, as indicated in patient charts. Teeth that did not have a stated reason for tooth extraction were excluded from the analysis. Indeed, a separate analysis considering OTL done in the present study showed that the prognostic performance of all systems was significantly impaired when OTL was used instead of TLP in the analysis (Table S3). The second reason was the lack of stringent inclusion criteria for patient cohorts in addition to using TLO (e.g., including both gingivitis and periodontitis patients) to validate prognostic systems tailored specifically for periodontitis patients (Nguyen et al., 2020). On the contrary, the present investigation included only treated periodontitis patients of different levels of disease severity. The third reason is that previous validation studies had put major emphasis on scrutinizing whether the initial assigned prognosis matched that of the follow-up (McGuire & Nunn, 1996; Nguyen et al., 2020). While this may be desirable in the short term, it is irrational and unexpected in the long term. This is especially true when the most categorical shifts occur towards a more positive prognosis (McGuire & Nunn, 1996). Also, longitudinal studies have shown that successful regenerative therapies can improve the prognosis of hopeless teeth which would otherwise be extracted (Cortellini et al., 2020).

Our approach was certainly not free from limitations. Indeed, the decision for tooth extraction (especially due to periodontal reasons) varies from one clinician to another, depending on their background, clinical judgement, and, in some instances, patient-related factors. Additionally, though our sample included 3787 teeth, some of the classes in particular systems were not at all examined. For example, the Faggion et al. system also had a score of 8 and 9, and our study sample did not have any teeth with such disease severity (Faggion et al., 2007). An additional limitation of this study is the absence of an a priori calculation of the sample size. This is due to the fact that this study was not carried out using a prospective cohort of participants; the available sample size was pre-determined. Inclusion of data over a long time period could have also led to systematic bias caused by alternating views on saving a tooth versus implant placement. A 10-year follow-up threshold also may have led to selection bias, though this study set the limit of ≥10 years of regular SPT as a criterion for inclusion to ensure that an effect from TLP could be demonstrated, given the very slow pattern of periodontitis progression.

It was presumed that an ideal periodontal prognostic system was one that had clear stratifications between each category and could accurately predict the outcome of a tooth after short (<5 years) and long (≥5 years) time periods (Kwok & Caton, 2007). The present study has followed the involved teeth for a mean of 26.5 years and up to 48 years and found the rate of TLP in this maintained population to be 4.72%. Though only one system, the Miller and McEntire (Miller et al., 2014), provided a long-term periodontal prediction of 15 and 30 years based on the score each tooth gained after being assessed, a better approach would be signifying short-term TLP as <10 years and long-term TLP as >10 years. This is especially true with dental implants demonstrating a high success rate of 97% and 75% over 10 and 20 years (Chappuis et al., 2013). Eventually, these systems are meant to be used for providing patients realistic and evidence-based expectations (AAP, 2008).

The frequency of SPT visits and strong compliance with oral hygiene recommendations have a significant impact on tooth mortality from a periodontal perspective (Becker et al., 1984). The present patient population had different levels of periodontitis severity. Patients were thus given individualized SPT schedules based on their case-specific disease severity. Subsequently, patients with mild forms of periodontitis or with stable periodontal conditions likely attended only one or two times per year, and others with more advanced forms of disease likely attended more frequently. As a result, the association between TLP and the prescribed maintenance protocol could not be investigated for each prognosis system in the current investigation, as the maintenance protocol was likely heavily influenced by the perceived prognosis.

5 | CONCLUSION

All tooth-level prognostic systems displayed very good predictive capability for TLP. Controlling for maintenance demonstrated an increase in the predictive ability, with no system scoring less than 0.925 accuracy. Overall, the Miller and McEntire system (Miller et al., 2014) may be considered to show the best discrimination between classes and model fit, followed by the Nunn et al. system (Nunn et al., 2012).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Study conception and design: Muhammad H. A. Saleh and Giuseppe Troiano. *Analysis and interpretation of the data*: Muhammad H. A. Saleh, Himabindu Dukka, and Giuseppe Troiano. *Data collection*: Muhammad H. A. Saleh, Himabindu Dukka, Andrea Ravidà, Musa Qazi, and Matthew Galli. *Drafting of the manuscript*: Muhammad H. A. Saleh, Himabindu Dukka, Andrea Ravidà, Matthew Galli, Henry Greenwell, and Hom-Lay Wang. All authors gave their final approval and agreed to be accountable for all aspects of the work.

ETHICS STATEMENT

The study was approved by the University of Michigan Medical School Institutional Review Board (IRBMED) (study identifier: HUM00157260).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Muhammad H. A. Saleh https://orcid.org/0000-0001-5067-7317
Giuseppe Troiano https://orcid.org/0000-0001-5647-4414
Andrea Ravidà https://orcid.org/0000-0002-3029-8130
Hom-Lay Wang https://orcid.org/0000-0003-4238-1799

REFERENCES

Akesson, L., Hakansson, J., & Rohlin, M. (1992). Comparison of panoramic and intraoral radiography and pocket probing for the measurement of the marginal bone level. *Journal of Clinical Periodontology*, *19*(5), 326–332. https://doi.org/10.1111/j.1600-051x.1992.tb00654.x

Al-Shammari, K. F., Al-Khabbaz, A. K., Al-Ansari, J. M., Neiva, R., & Wang, H. L. (2005). Risk indicators for tooth loss due to periodontal disease. *Journal of Periodontology*, *76*(11), 1910–1918. https://doi.org/10.1902/jop.2005.76.11.1910

American Academy of Periodontology. (2008). American Academy of Periodontology statement on risk assessment. *Journal of Periodontology*, *79*(2), 202. https://doi.org/10.1902/jop.2008.082001

Avila, G., Galindo-Moreno, P., Soehren, S., Misch, C. E., Morelli, T., & Wang, H. L. (2009). A novel decision-making process for tooth retention or extraction. *Journal of Periodontology*, *80*(3), 476–491. https://doi.org/10.1902/jop.2009.080454

Beck, J. D. (1998). Risk revisited. *Community Dentistry and Oral Epidemiology*, *26*(4), 220–225.

Becker, W., Becker, B. E., & Berg, L. E. (1984). Periodontal treatment without maintenance. A retrospective study in 44 patients. *Journal of Periodontology*, *55*(9), 505–509. https://doi.org/10.1902/jop.1984.55.9.505

Chambrone, L., Chambrone, D., Lima, L. A., & Chambrone, L. A. (2010). Predictors of tooth loss during long-term periodontal maintenance: A systematic review of observational studies. *Journal of Clinical Periodontology*, *37*(7), 675–684. https://doi.org/10.1111/j.1600-051X.2010.01587.x

Chappuis, V., Buser, R., Bragger, U., Bornstein, M. M., Salvi, G. E., & Buser, D. (2013). Long-term outcomes of dental implants with a titanium plasma-sprayed surface: A 20-year prospective case series study in partially edentulous patients. *Clinical Implant Dentistry and Related Research*, *15*(6), 780–790. https://doi.org/10.1111/cid.12056

Checchi, L., Montevecchi, M., Gatto, M. R., & Trombelli, L. (2002). Retrospective study of tooth loss in 92 treated periodontal patients. *Journal of Clinical Periodontology*, *29*(7), 651–656. https://doi.org/10.1034/j.1600-051x.2002.290710.x

Cortellini, P., Stalpers, G., Mollo, A., & Tonetti, M. S. (2020). Periodontal regeneration versus extraction and dental implant or prosthetic replacement of teeth severely compromised by attachment loss to the apex: A randomized controlled clinical trial reporting 10-year outcomes, survival analysis and mean cumulative cost of recurrence. *Journal of Clinical Periodontology*, *47*(6), 768–776. https://doi.org/10.1111/jcpe.13289

Faggion, C. M., Jr., Petersilka, G., Lange, D. E., Gerss, J., & Flemmig, T. F. (2007). Prognostic model for tooth survival in patients treated for periodontitis. *Journal of Clinical Periodontology*, *34*(3), 226–231. https://doi.org/10.1111/j.1600-051X.2006.01045.x

Fardal, O., Johannessen, A. C., & Linden, G. J. (2004). Tooth loss during maintenance following periodontal treatment in a periodontal practice in Norway. *Journal of Clinical Periodontology*, *31*(7), 550–555. https://doi.org/10.1111/j.1600-051X.2004.00519.x

Grant, D., Stern, I., & Listgarten, M. (1979). *Periodontics in the tradition of Gottlieb and Orban* (5th ed.). Mosby.

Helal, O., Gostemeyer, G., Krois, J., Fawzy El Sayed, K., Graetz, C., & Schwendicke, F. (2019). Predictors for tooth loss in periodontitis patients: Systematic review and meta-analysis. *Journal of Clinical Periodontology*, *46*(7), 699–712. https://doi.org/10.1111/jcpe.13118

Kwok, V., & Caton, J. G. (2007). Commentary: Prognosis revisited: A system for assigning periodontal prognosis. *Journal of Periodontology*, *78*(11), 2063–2071. https://doi.org/10.1902/jop.2007.070210

Lang, N. P., Suvan, J. E., & Tonetti, M. S. (2015). Risk factor assessment tools for the prevention of periodontitis progression a systematic review. *Journal of Clinical Periodontology*, *42*(Suppl 16), S59–S70. https://doi.org/10.1111/jcpe.12350

Lang, N. P., & Tonetti, M. S. (1996). Periodontal diagnosis in treated periodontitis. Why, when and how to use clinical parameters. *Journal of Clinical Periodontology*, *23*(3 Pt 2), 240–250. https://doi.org/10.1111/j.1600-051x.1996.tb02083.x

Levine, R. A., & Miller, P. D. (2020). *The Miller McEntire periodontal prognostic index (i.e., "the perio report card") usage in practice*. Springer.

McGuire, M. K. (1991). Prognosis versus actual outcome: A long-term survey of 100 treated periodontal patients under maintenance care. *Journal of Periodontology*, *62*(1), 51–58. https://doi.org/10.1902/jop.1991.62.1.51

McGuire, M. K., & Nunn, M. E. (1996). Prognosis versus actual outcome. II. The effectiveness of clinical parameters in developing an accurate prognosis. *Journal of Periodontology*, *67*(7), 658–665. https://doi.org/10.1902/jop.1996.67.7.658

Miller, P. D., Jr., McEntire, M. L., Marlow, N. M., & Gellin, R. G. (2014). An evidenced-based scoring index to determine the periodontal prognosis on molars. *Journal of Periodontology*, *85*(2), 214–225. https://doi.org/10.1902/jop.2013.120675

Moons, K. G., Altman, D. G., Reitsma, J. B., Collins, G. S., & Transparent Reporting of a Multivariate Prediction Model for Individual Prognosis or Development Initiative (2015). New guideline for the reporting of studies developing, validating, or updating a multivariable clinical prediction model: The TRIPOD statement. *Advances in anatomic pathology*, *22*(5), 303–305.

Nguyen, L., Krish, G., Alsaleh, A., Mailoa, J., Kapila, Y., Kao, R. T., & Lin, G. H. (2020). Analyzing the predictability of the kwok and caton periodontal prognosis system: A retrospective study. *Journal of Periodontology*, *92*, 662–669. https://doi.org/10.1002/JPER.20-0411

Nunn, M. E. (2003). Understanding the etiology of periodontitis: An overview of periodontal risk factors. *Periodontology 2000*, *32*, 11–23. https://doi.org/10.1046/j.0906-6713.2002.03202.x

Nunn, M. E., Fan, J., Su, X., Levine, R. A., Lee, H. J., & McGuire, M. K. (2012). Development of prognostic indicators using classification and regression trees for survival. *Periodontology 2000*, *58*(1), 134–142. https://doi.org/10.1111/j.1600-0757.2011.00421.x

Pepelassi, E. A., Tsiklakis, K., & Diamanti-Kipioti, A. (2000). Radiographic detection and assessment of the periodontal endosseous defects. *Journal of Clinical Periodontology*, *27*(4), 224–230. https://doi.org/10.1034/j.1600-051x.2000.027004224.x

Persson, G. R., Mancl, L. A., Martin, J., & Page, R. C. (2003). Assessing periodontal disease risk: A comparison of clinicians' assessment versus a computerized tool. *Journal of the American Dental Association*, *134*(5), 575–582. https://doi.org/10.14219/jada.archive.2003.0224

Rahman, M. S., Ambler, G., Choodari-Oskooei, B., & Omar, R. Z. (2017). Review and evaluation of performance measures for survival prediction models in external validation settings. *BMC Medical Research Methodology*, *17*(1), 60. https://doi.org/10.1186/s12874-017-0336-2

Ravida, A., Qazi, M., Troiano, G., Saleh, M. H. A., Greenwell, H., Kornman, K., & Wang, H. L. (2020). Using periodontal staging and grading system as a prognostic factor for future tooth loss: A long-term retrospective study. *Journal of Periodontology*, *91*(4), 454–461. https://doi.org/10.1002/JPER.19-0390

Rentfro, A. R., McEwen, M., & Ritter, L. (2009). Perspectives for practice: Translating estimated average glucose (eAG) to promote diabetes self-management capacity. *The Diabetes Educator*, *35*(4), 581, 585–5866, 588–590 passim. https://doi.org/10.1177/0145721709338222

Rohlfing, C. L., Wiedmeyer, H. M., Little, R. R., England, J. D., Tennill, A., & Goldstein, D. E. (2002). Defining the relationship between plasma glucose and HbA(1c): Analysis of glucose profiles and HbA(1c) in the Diabetes Control and Complications Trial. *Diabetes Care*, *25*, 275–279.

Rulli, E., Ghilotti, F., Biagioli, E., Porcu, L., Marabese, M., D'Incalci, M., Bellocco, R., & Torri, V. (2018). Assessment of proportional hazard assumption in aggregate data: a systematic review on statistical methodology in clinical trials using time-to-event endpoint. *British journal of cancer*, *119*(12), 1456–1463.

Schwendicke, F., Schmietendorf, E., Plaumann, A., Salzer, S., Dorfer, C. E., & Graetz, C. (2018). Validation of multivariable models for predicting tooth loss in periodontitis patients. *Journal of Clinical Periodontology*, *45*(6), 701–710. https://doi.org/10.1111/jcpe.12900

Steyerberg, E. W., & Harrell, F. E., Jr. (2016). Prediction models need appropriate internal, internal–external, and external validation. *Journal of Clinical Epidemiology*, *69*, 245–247. https://doi.org/10.1016/j.jclinepi.2015.04.005

Tan, H., Peres, K. G., & Peres, M. A. (2016). Retention of teeth and oral health-related quality of life. *Journal of Dental Research*, *95*(12), 1350–1357. https://doi.org/10.1177/0022034516657992

World Medical Association. (1975). *Declaration of Helsinki: Recommendations guiding medical doctors in biomedical research involving human subjects; Adopted by the 18th World Medical Assembly, Helsinki, Finland, June 1964 and as revised by the 29th World Medical Assembly*. (p. 4). Tokyo, Japan. https://www.wma.net/wp-content/uploads/2018/07/DoH-Oct1975.pdf

World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, *310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Saleh, M. H. A., Dukka, H., Troiano, G., Ravidà, A., Galli, M., Qazi, M., Greenwell, H., & Wang, H.-L. (2021). External validation and comparison of the predictive performance of 10 different tooth-level prognostic systems. *Journal of Clinical Periodontology*, *48*(11), 1421–1429. https://doi.org/10.1111/jcpe.13542