


Systematic Review

Clinical Prediction Models for Recurrence in Patients with Resectable Grade 1 and 2 Sporadic Non-Functional Pancreatic Neuroendocrine Tumors: A Systematic Review

Jeffrey W. Chen ^{1,2,3,*}, Charlotte M. Heidsma ^{1,2,3}, Anton F. Engelsman ^{1,2,3}, Ertunç Kabaktepe ^{1,3}, Susan van Dieren ^{1,3}, Massimo Falconi ⁴ , Marc G. Besselink ^{1,3} and Els J. M. Nieveen van Dijkum ^{1,2,3}

¹ Department of Surgery, Amsterdam UMC, Location University of Amsterdam, 1081 HV Amsterdam, The Netherlands

² Amsterdam Center for Endocrine and Neuroendocrine Tumors (ACCENT), 1081 HV Amsterdam, The Netherlands

³ Cancer Center Amsterdam, 1081 HV Amsterdam, The Netherlands

⁴ Pancreatic Surgery, IRCCS Ospedale San Raffaele, Università Vita-Salute, 20132 Milan, Italy

* Correspondence: j.w.chen@amsterdamumc.nl

Simple Summary: The risk prediction for tumor recurrence after surgery for non-functional pancreatic neuroendocrine tumors is a major unmet clinical need. Accurate recurrence risk prediction could pave the way for tailor-made follow-up protocols in the future, as well as help select suitable patients for cancer treatment trials. Multiple prediction models have been developed; however, none are currently incorporated into international guidelines. This systematic review found 13 original models, of which 3 were validated outside the patient group in which they were developed. The effectiveness of a prediction model is not proven in the wider population without this validation; thus, the lack of it hinders the progress toward clinical use. We propose to test the included models in a large, multinational database to compare their performance. We recommend all authors developing a prediction model to perform the minimally required tests of model development and to implement it by creating an online calculator.

Abstract: Recurrence after resection in patients with non-functional pancreatic neuroendocrine tumors (NF-pNET) has a considerable impact on overall survival. Accurate risk stratification will tailor optimal follow-up strategies. This systematic review assessed available prediction models, including their quality. This systematic review followed PRISMA and CHARMS guidelines. PubMed, Embase, and the Cochrane Library were searched up to December 2022 for studies that developed, updated, or validated prediction models for recurrence in resectable grade 1 or 2 NF-pNET. Studies were critically appraised. After screening 1883 studies, 14 studies with 3583 patients were included: 13 original prediction models and 1 prediction model validation. Four models were developed for preoperative and nine for postoperative use. Six models were presented as scoring systems, five as nomograms, and two as staging systems. The *c* statistic ranged from 0.67 to 0.94. The most frequently included predictors were tumor grade, tumor size, and lymph node positivity. Critical appraisal deemed all development studies as having a high risk of bias and the validation study as having a low risk of bias. This systematic review identified 13 prediction models for recurrence in resectable NF-pNET with external validations for 3 of them. External validation of prediction models improves their reliability and stimulates use in daily practice.

Keywords: pancreatic neuroendocrine tumor; non-functional; resectable; recurrence; prediction model; grade 1; grade 2



Citation: Chen, J.W.; Heidsma, C.M.; Engelsman, A.F.; Kabaktepe, E.; van Dieren, S.; Falconi, M.; Besselink, M.G.; Nieveen van Dijkum, E.J.M. Clinical Prediction Models for Recurrence in Patients with Resectable Grade 1 and 2 Sporadic Non-Functional Pancreatic Neuroendocrine Tumors: A Systematic Review. *Cancers* **2023**, *15*, 1525. <https://doi.org/10.3390/cancers15051525>

Academic Editor: Adam E. Frampton

Received: 5 January 2023

Revised: 20 February 2023

Accepted: 22 February 2023

Published: 28 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Survival of patients with pancreatic neuroendocrine tumors (pNET) after resection is largely dependent on the development of tumor recurrence [1]. In contrast to other pancreatic neoplasms, pNET displays heterogeneous behavior with survival depending on tumor size, tumor grade, and treatment outcome. Patients with small (<2 cm) nonmetastatic grade 1 pNET have 5-year survival outcomes approaching 100%. Meanwhile, patients with larger metastasized grade 2 pNET had a 5-year survival rate of less than 40% [2,3]. Most patients are diagnosed incidentally with a non-functional pNET (NF-pNET), and only a subset of these patients eventually develops recurrence after tumor resection [4]. Surgical resection is the only curative intent treatment for localized pNET [5,6]. Improvements in the therapeutic armamentarium for advanced-stage disease in pNET have broadened the treatment options for patients with recurrence after surgery. Systemic therapies such as somatostatin analogs and chemotherapy appear to provide favorable outcomes in patients with metastases, while PRRT can be offered to a select group of patients [6–9].

Identifying patients at high risk for recurrence after resection with curative intent is a challenge faced by clinicians treating patients with a pNET. Currently, no tool is used to provide patients with standardized postoperative follow-up regimens [10]. Follow-up programs could be tailored according to the risk of recurrence, reducing the follow-up intensity of low-risk patients while those at high risk could undergo intensive surveillance. Furthermore, adjuvant therapy could also be tested in high-risk patients to prevent disease spread, but no such therapy is offered yet. Various studies have explored the risk factors associated with recurrence in patients with pNET [11–13]. Combining them into a prediction model is needed to make these tailor-made treatments possible. Multiple prediction models have been developed, but none of these models have been incorporated either into European Neuroendocrine Tumor Society (ENETS) guidelines or into clinical practice yet [7,14,15].

Accurate patient selection for the intensity of follow-up and (neo)adjuvant treatment is both an unmet need according to the ENETS guidelines, therefore warranting the need for a clinical prediction model [16]. The aim of this systematic review was to identify and evaluate currently available prediction models for recurrence in resected grade 1 and 2 NF-pNET and to select models for future use in clinical practice and clinical trials.

2. Materials and Methods

The systematic review followed the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) guideline [17]. The study was reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [18]. The study was registered at PROSPERO (CRD42022380671).

2.1. Search Strategy

A systematic search was performed in MEDLINE via PubMed, Embase, and the Cochrane Library for studies published from inception until December 2022. The search included synonyms for [pancreatic neuroendocrine tumors] combined with [prognostic/predictive/prediction models] and [recurrence] (Appendix A). After the removal of duplicates, titles/abstracts and full-text articles were screened independently by two authors (J.W.C. and C.M.H.). Differences in opinion were resolved through discussion; if necessary, a third author (E.J.M.N.v.D.) was consulted. References of included articles were checked for other potentially eligible studies.

2.2. Eligibility

Studies developing, updating, or validating a clinical prediction model for recurrence in patients with grade 1 or 2 NF-pNET undergoing resection were included. Models designed for use in either the preoperative or postoperative setting were included. Studies providing an absolute probability estimate and those stratifying patients into risk categories were included, as both strategies might serve the clinical aim of providing more tailored

treatment. To allow evaluation and comparison of model performance, at least one of the following performance outcomes had to be reported: *c* statistic, area under the curve (AUC), R2, Brier score, sensitivity, specificity, calibration plots, or calibration statistics.

Review papers, abstracts, case studies, studies with patients aged less than 18 years, and non-human studies were excluded. Articles including patients with diseases other than pancreatic neuroendocrine tumors (such as pancreatic ductal adenocarcinoma or intraductal papillary mucinous neoplasm) or other neuroendocrine tumor localizations were excluded. Studies were excluded if more than 20% of its population were patients with WHO grade 3, genetic background, metastatic or functional pNET. If not reported, the studies were included but marked as a risk of bias. Studies validating a staging system (i.e., TNM, WHO, AJCC) without any modification were not included, because the limited discriminative strength of these systems was one of the reasons for developing better models [19–21].

2.3. Data Extraction and Analysis

Data extraction and critical appraisal were performed independently by two authors (J.W.C. and C.M.H.). Articles were categorized into development/model update studies and validation studies. A data extraction sheet was used to extract data for all studies and included: the first author, year of publication, pre- or postoperative setting, sample size, study interval, source of data, countries of inclusion, number of centers, predicted outcome, included predictors, model performance, information on validation and number of citations. For development/update studies, model development method, number of prognostic factors screened, and final model presentation were additionally extracted. Predictors included in regression analyses were collected and scored for statistical significance. If multiple models were developed or tested, the prediction model with the highest *c* statistic or the model proposed by the authors was included. Meta-analysis was not possible due to heterogeneity in prediction model variables and outcomes. Critical appraisal was performed following the CHARMS guidelines [22]. The PROBAST critical appraisal tool was used for the assessment of methodological quality [23]. A clinical epidemiologist was consulted for this review [S.v.D.].

2.4. Definitions and Terminology

A prediction model was defined as “a formal combination of multiple predictive factors from which risks of a specific endpoint can be calculated for individual patients” [24]. Clinical prediction models should “discriminate between individual patients who do and do not experience a specific event (discrimination), should make accurate predictions (calibration), perform well across different patient populations (generalizability) and be readily interpretable” [25,26].

Model performance can be assessed by its ability to discriminate and calibrate. Discrimination (i.e., do patients who have the outcome also have a higher predicted risk than those who do not have the outcome) can be quantified through measures such as sensitivity, specificity, area under the receiver operating curve, or by the concordance statistic (*c* statistic) [27,28]. A *c* statistic of 1.0 represents a perfect model, while a score of 0.5 indicates that the model is not better at prediction than random selection [29]. In binary outcomes, the AUC is equal to the *c* statistic. In turn, calibration is important for model performance since it compares the predicted probability of the outcome (i.e., recurrence) with the actual outcome. It is most often visualized using calibration plots or assessed as goodness of fit, which can be quantified using the Hosmer–Lemeshow test ($p < 0.050$ indicates poor calibration).

Model validation is imperative and can be performed using several different techniques. It is possible to internally validate (reproducibility) followed by external validation (generalizability). Internal validation can be performed through split-sample validation (training and test set), cross-validation (development in random segments of the population, tested in the remaining segment and repeating this process), and bootstrapping (random

samples of the same size are drawn with replacement) [24,29]. Ideally, all prediction models should undergo not only internal but also external validation. Geographical (using a new cohort from a different center), temporal (same center, but patients at a different time interval), and fully independent (new research group at a different center) are the three most important external validation options [25,29].

The term tumor grade was used for both the World Health Organization (WHO) definition of tumor grade as well as the Ki-67 index, unless specified otherwise. Tumor grade was defined according to the WHO 2017 classification: grade 1 (Ki-67 index of <3%), grade 2 (Ki-67 index of 3–20%), and grade 3 (Ki-67 index of >20%), unless otherwise specified [20]. Histological grade was defined as the grading of tumor cell differentiation into well, intermediate, or poorly differentiated.

3. Results

3.1. Baseline Characteristics

The search yielded 1883 studies (Figure 1), of which 14 could be included in this review containing 3583 patients, of which 1668 (46.6%) were female. The median age was 58 years (IQR 53–60). The median follow-up duration was 48 months (IQR 40–56). Model development data from 3241 patients were available for this systematic review. Models were developed with a median number of 140 patients (IQR 87–235) consisting of 83.0% non-functioning pNET cases, 6.7% metastatic, and 7.9% genetic. The population consisted of 1527 patients (62.7%) with a WHO grade 1 tumor, 836 patients (34.3%) with a grade 2 tumor, and a small percentage of patients had grade 3 ($n = 73$, 3.0%). No patients received (neo)adjuvant therapy. The most commonly performed procedure was distal pancreatectomy ($n = 1179$, 46.7%), followed by pancreatoduodenectomy ($n = 696$, 27.6%), enucleation ($n = 233$, 9.2%), central pancreatectomy ($n = 95$, 3.8%), and total pancreatectomy ($n = 53$, 2.1%). The 5-year RFS was reported by seven studies, reporting a median survival of 74% (IQR 72.0–84.6) [30,31]. The 5-year overall survival was reported by six studies with a median survival of 90.6% (IQR 80.1–91.0) [32–37]. Full details of other characteristics can be found in Table 1.

Table 1. Prediction model characteristics.

Author (Year)	n	Setting	Country	Single or Multicenter	Study Interval	Model Development Method	Study Type	Follow-Up (IQR)	Outcome Predicted	Definition of Outcome
Ballian et al. (2009)	43	Postop	USA	Single center	1991–2007	Univariable Cox regression analysis	DEV	68 months	RFS (5-year) DSS (5-year)	Radiographic evidence of new tumors
Dong et al. (2021)	416	Postop	China	8 centers	1997–2016	Uni- and multivariable logistic regression analysis	DEV	31 months (11.3–56.6)	RFS (5-, 10-year)	Time from surgery to the time of identification of suspicious imaging findings or biopsy-proven tumor
Fisher et al. (2019)	224	Preop	USA	8 centers	2000–2016	Univariable Cox regression analysis	DEV	37 months (15–62)	RFS	Time of resection until the time of radiographic or pathological evidence of tumor recurrence
Genc et al. (2018)	211	Postop	Netherlands, Italy	3 centers	1992–2015	Uni- and multivariable Cox regression analysis	DEV	51 months (29–72)	RFS (5-year) DSS (5-year)	RFS: Percentage of patients without recurrence in the pancreas, new positive lymph nodes, or distant metastasis after resection DSS: Percentage of patients who have not died due to pNET

Table 1. Cont.

Author (Year)	n	Setting	Country	Single or Multicenter	Study Interval	Model Development Method	Study Type	Follow-Up (IQR)	Outcome Predicted	Definition of Outcome
Heidsma et al. (2021)	342	Postop	Australia, Austria, France, Germany, Netherlands, Sweden, USA	7 centers	1991–2018	Uni- and multivariable Cox regression analysis	VAL	50.5 months (22.3–103)	RFS (5-year)	Date of the first cross-sectional imaging on which a new local or metastatic lesion was detected
Liu et al. (2013)	75	Postop	USA	Single center	1993–2009	Uni- and multivariable Cox regression analysis	DEV	69 months (range 1–212)	RFS	Time from surgery to death due to disease or to disease recurrence at local, regional, or distant sites, whichever occurred first
Primavesi et al. (2020)	D: 160 V: 204	Preop	Austria, Germany, Italy, Netherlands	D: 6 centers V: 4 centers	D: 1998–2017 V: 1990–2018	Uni- and multivariable Cox regression analysis	DEV	57.5 months (28.3–83.3)	RFS (5-, 10-year) DSS (5-, 10-year) OS (5-, 10-year)	RFS: Time from initial curative intent to recurrence/last follow-up DSS: Time to pNET related-death/last follow-up OS: Time from first surgery to death/last follow-up
Pulvirenti et al. (2021)	D: 632 V: 328	Postop	Australia, UK, USA		2000–2016	Univariable Cox regression analysis	DEV	51 months	RFS (5-year)	Date of curative surgery until date of first recurrence identified through routine postoperative CT-scans
Sho et al. (2019)	140	Postop	USA	Single center	1989–2015	Uni- and multivariable Cox regression analysis	DEV	56 months	RFS (5-year)	Time to the last known date the patient was disease free
Sun et al. (2019)	81	Preop	China	Single center	2009–2017	Uni- and multivariable Cox regression analysis	DEV	16 months (range 6–108)	RFS (1-, 2-, 3-year)	Day of surgery to the time of local recurrence or distant metastatic disease on radiological images, last clinical follow-up, or death
Viúdez et al. (2016)	92	Postop	USA	Single center	1998–2010	Uni- and multivariable Cox regression analysis	DEV	n.r.	RFS OS	Time of surgery to date of relapse, death, or last follow-up
Wei et al. (2021)	D: 125 V: 77	Postop	China	Single center	2012–2018	Uni- and multivariable Cox regression analysis	DEV	41 months (27–59.8)	RFS (3-, 5-year)	n.r.
Zhou et al. (2017)	125	Preop	China	Single center	2003–2016	Uni- and multivariable Cox regression analysis	DEV	45.8 months (SD 37.01)	RFS OS	RFS: Time from date of surgery to date of recurrence OS: Time from date of initial diagnosis until date of death from any cause or date of last known contact
Zou et al. (2020)	245	Postop	China	Single center	2002–2018	Uni- and multivariable Cox regression analysis	DEV	40 months	RFS (3-, 5-year)	RFS: Date of recurrence in any forms

IQR, interquartile range; Postop, postoperative; Preop, preoperative; USA, United States of America; UK, United Kingdom; D, development cohort; V, validation cohort; DEV, development study; VAL, validation study; OS, overall survival; DSS, disease-specific survival; RFS, recurrence-free survival; SD, standard deviation; n.r., not reported.

3.2. Preoperative Prediction Models

Four out of thirteen prediction models were designed for preoperative use [31,33,36,37], with *c* statistics ranging from 0.78 to 0.94. Sun et al. used preoperative MRI-imaging variables to calculate the recurrence risk [31]. Tumors exhibiting hypoenhancement and low apparent diffusion coefficient values were associated with worse RFS after curative resection. Fisher et al. included serum chromogranin A > 5 times the upper limit and the presence of a recurrent tumor as recurrence predictors [33]. Their sensitivity analysis

suggested that chromogranin A might have a low predictive value in low-grade tumors. Primavesi et al. included C-reactive protein ≥ 0.2 mg/dL and the presence of distant metastasis as predictors [36]. However, in their study, C-reactive protein was not associated with inferior 5-year RFS (72.2% vs. 70.4%, $p = 0.89$) or increased recurrence rate (25.6% vs. 21.2%, $p = 0.34$). The model by Zhou et al. was designed for preoperative prediction of both recurrence and survival, but only the AUC of overall survival was reported (0.83) [37]. Their study analyzed the predictive strength of “gamma-glutamyl transferase/lymph-node-ratio” as a preoperative predictor, but the predictive strength of adding this predictor to the AJCC staging system (AUC 0.81 [95% CI 0.73–0.90]) or WHO classification system (AUC 0.81 [0.73–0.88]) was inferior to the combination of the AJCC staging system and WHO classification (AUC 0.83 [0.75–0.91]). Full details on prediction model performance can be found in Table 2.

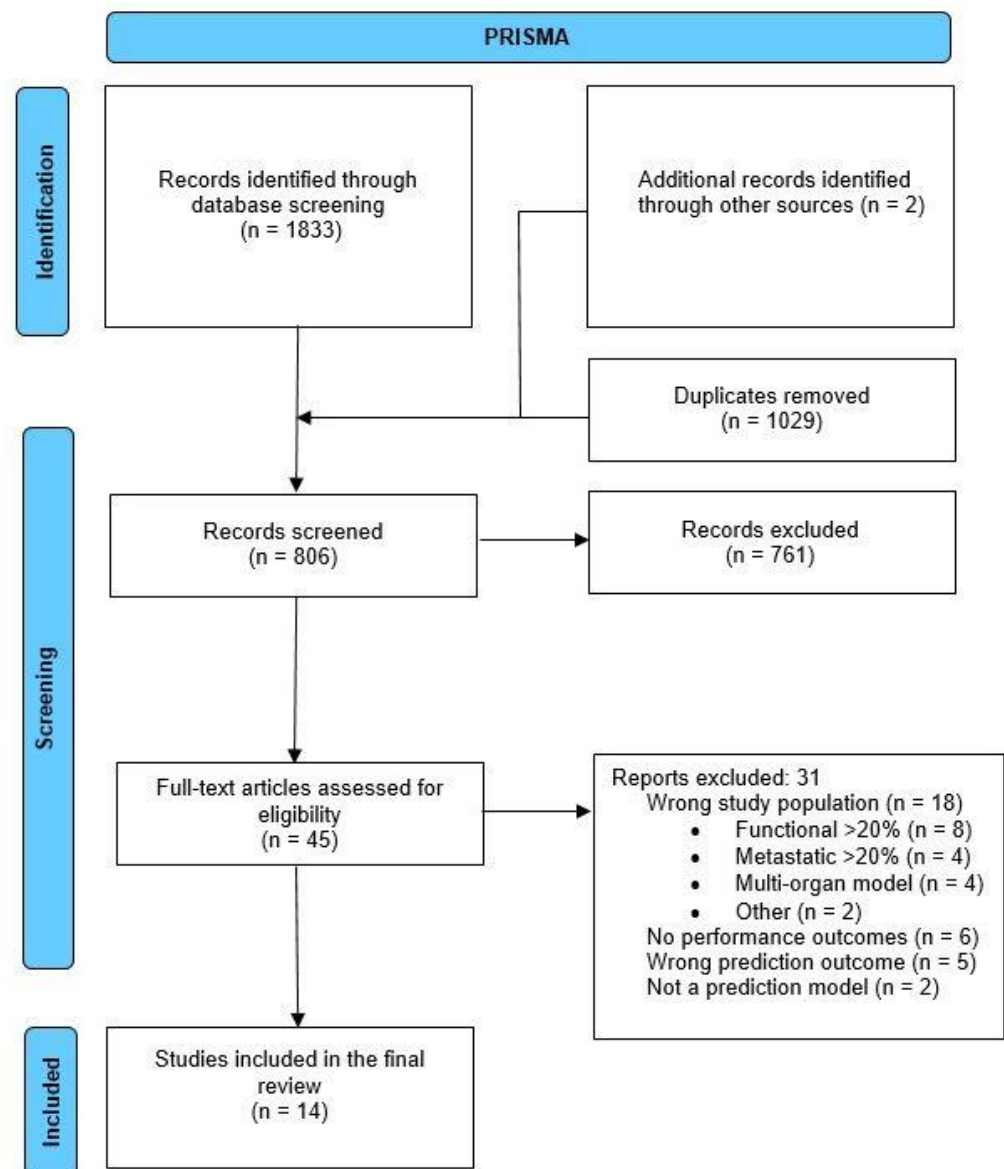


Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flow diagram reporting the number of studies identified, screened, and selected for inclusion in this systematic review.

Table 2. Prediction model performance.

Author (Year)	Model Type	No. of Variables Screened	Predictors in Final Model	Discrimination, c Statistic (95% CI)	Calibration	Internal Validation	External Validation
Ballian et al. (2009)	Scoring system	9	Tumor size ≥ 5 cm, Hochwald grading system (<i>mitotic index + necrosis</i>), positive lymph node, R1 resection	RFS: 0.80 (0.68–0.91) DSS: 0.81 (0.73–0.90)	n.r.	n.r.	n.r.
Dong et al. (2021)	Nomogram	10	Tumor grade (2010), tumor burden score ($(\text{max. tumor diameter})^2 + (\text{number of tumors})^2$), positive lymph node	0.75 (0.66–0.79)	Good performance	Bootstrapping: 5000 iterations 0.71 (0.65–0.75)	n.r.
Fisher et al. (2019)	Scoring system	6	Tumor grade 2 or 3, chromogranin A $> 5 \times$ upper limit, surgery for tumor recurrence, tumor size ≥ 4 cm	See internal validation	n.r.	Split-sample validation AUC: 0.78	n.r.
Genç et al. (2018)	Nomogram	11	Tumor grade (2010), positive lymph node, perineural invasion	0.81 (0.75–0.87)	GOF; Hosmer Lemeshow Chi-square 11.25, $p = 0.258$	Performed	n.r.
Heidsma et al. (2021)	Nomogram	N/A	Tumor grade (2017), positive lymph node, perineural invasion	0.77 (0.71–0.83)	Calibration slope 0.74	N/A	N/A
Liu et al. (2013)	Staging system	9	Ki-67 index, Hochwald grading system (<i>mitotic index + necrosis</i>)	0.79	n.r.	n.r.	n.r.
Primavesi et al. (2020)	Scoring system	7	CRP > 0.2 mg/dL, tumor size > 3 cm, metastasis	RFS: AUC 66.5 (58.9–74.2) DSS: AUC 77.3 (67.2–87.5) OS: AUC 68.9 (61.5–76.4)	n.r.	n.r.	n.r.
Pulvirenti et al. (2021)	Nomogram	11	Number of positive lymph nodes, Ki-67 index, tumor size, vascular and/or perineural invasion	0.85	Performed	Bootstrapping: 100 iterations	0.84 (0.79–0.88)
Sho et al. (2019)	Scoring system	13	Tumor size ≥ 5 cm, positive lymph node, tumor grade (2010)	0.82 (0.72–0.92)	n.r.	Bootstrapping: 1000 iterations	n.r.
Sun et al. (2019)	Nomogram	22	Tumor size > 2 cm, hypoenhancement, apparent diffusion coefficient	0.91 (0.84–0.98)	Performed	n.r.	n.r.
Viúdez et al. (2016)	Scoring system	14	RFS: AJCC, tumor size, tumor grade, Immunohistochemistry Prognostic Score (<i>MGMT, PHLDA-3, NDRG-1 expressions</i>) OS: AJCC, age > 60 , Immunohistochemistry Prognostic Score (<i>MGMT, PHLDA-3, NDRG-1 expressions</i>)	RFS: 0.80 OS: 0.79	n.r.	n.r.	n.r.
Wei et al. (2021)	Nomogram	20	Metastasis, tumor grade, Immunoscore ($0.261 \times$ the status of <i>CCL19</i>) + ($0.490 \times$ the status of <i>IL-16</i>) + ($0.123 \times$ the status of <i>CD163</i>) + ($0.044 \times$ the status of <i>CD8PT</i>) – ($0.011 \times$ the status of <i>CD8IT</i>) – ($0.493 \times$ the status of <i>IRF4</i>)	0.92 (0.88–0.95)	Good agreement	Bootstrapping	0.86 (0.80–0.93)
Zhou et al. (2017)	Staging system	27	AJCC, tumor grade (2010)	OS: AUC 0.83 (0.75–0.91)	n.r.	n.r.	Not performed
Zou et al. (2020)	Scoring system	11	Positive lymph node, tumor size, tumor grade (2017)	3 year: 0.91 5 year: 0.94 8 year: 0.93	Good performance	n.r.	n.r.

GOF, Goodness-of-Fit; N/A, not applicable; CRP, C-reactive protein; DSS, disease-specific survival; RFS, recurrence-free survival; OS, overall survival; n.r., not reported; AJCC, American Joint Committee on Cancer.

3.3. Postoperative Prediction Models

Nine prediction models were developed for postoperative use [11,30,32,34,35,38–41]. The *c* statistics ranged from 0.75 to 0.94. Four studies created a model based on a combination of tumor grade, tumor size, and lymph node positivity (Table 2) [11,30,35,40]. The most frequently included predictor was tumor grade (*n* = 11) [11,30,34,35,38–41]. All but two studies followed the WHO 2017 cut-off values for tumor grade [20]. Pulvirenti et al. incorporated the Ki-67 index as a continuous variable in their nomogram [11]. Sho et al. used a Ki-67 index <8% as a low tumor grade cut-off value for its scoring system [35]. Tumor size was the second most included predictor (*n* = 11) [11,30,32,35,40,41]. Two models incorporated tumor size as a continuous variable [11,40], two models used a cut-off value of ≥ 5 cm [32,35], one used a cut-off value of ≥ 2.5 cm [41], and one model incorporated tumor size into a tumor burden score [30]. Viúdez et al. developed the Immunohistochemistry Prognostic Score as a predictor (high Immunohistochemistry score; HR 2.68 [95% CI 1.60–4.49], *p* = 0.03), where they quantified the immunohistochemistry patterns of DNA-methylation [42]. Wei et al. also developed their own predictor, i.e., the Immunoscore (low Immunoscore; HR 0.13 [0.06–0.39], *p* < 0.001), where the immune response in the tumor microenvironment was quantified [39]. Furthermore, Ballian et al. and Liu et al. incorporated the Hochwald grading system as a predictor, where a mitotic rate of >2 per 50 high-power field and signs of histopathological necrosis were associated with a higher recurrence rate [32,38,42]. However, association through multivariable analyses was not reported in both studies.

3.4. Predictor Selection

Predictor selection based on uni- and multivariable Cox regression analysis was performed in nine studies (Table 1) [31,34,35,37–41,43]. Three studies selected predictors through univariable Cox regression analysis [11,32,33] and Dong et al. through logistic regression analysis [30]. Although stating that the model was built on the beta coefficients of multivariable analysis, only the results of their univariable analysis were reported by Dong et al. [30,44]. The included studies screened a median of 11 variables (range 9–17) and included a median of 3 variables (range 2–7) in their final prediction model. Tumor grade was most frequently associated with RFS in 10 out of 11 included analyses after regression analysis (91%) [11,30,33–35,37–41] (Table 3), followed by tumor size (73%) [11,31–33,35,38,40,41]. The patients' sex was analyzed by seven studies, but only one study showed an association with RFS (Male; HR 2.22 [95% CI 1.14–4.31], *p* = 0.018) [37].

Table 3. Predictor association with recurrence-free survival.

Study	Tumor Grade/ Ki-67 Index	Tumor Size	Positive Lymph Node	Sex	Age	Perineural Invasion	R1 Resection	Vascular Invasion	Histological Grade	Functional Tumor	Symptomatic	Distant Metastasis	TNM-Stage	Minimally Invasive	Predictors Screened
Pulvirenti (2021) ^b	✓	✓	✓	X	✓	✓	✓	✓		✓				X	10
Dong (2021) ^b	✓		✓	X	X	X	X		✓		X			✓	9
Liu (2013) ^a	✓	✓	✓	X	X		X	✓		X					8
Ballian (2009) ^b		✓	✓	X	X	✓	✓	X							7
Zhou (2017) ^a	✓	X		✓		X					X	X	✓		7
Primavesi (2020) ^a		X		X	✓						X	✓			5
Zou (2020) ^a	✓	✓		X		X		X							5

Table 3. Cont.

Study	Tumor Grade/ Ki-67 Index	Tumor Size	Positive Lymph Node	Sex	Age	Perineural Invasion	RI Resection	Vascular Invasion	Histological Grade	Functional Tumor	Symptomatic	Distant Metastasis	TNM-Stage	Minimally Invasive	Predictors Screened
Fisher (2019) ^b	✓	✓			X					X					4
Sho (2019) ^a	✓	✓	✓						✓						4
Sun (2019) ^a	X	✓	X						X						4
Genç (2018) ^a	✓		✓			✓									3
Viúdez (2016) ^a	✓	✓											X		3
Wei (2021) ^a	✓	X	X												3
Predictor analyzed by study (out of 13), n (%)	11 (85)	11 (85)	8 (62)	7 (54)	6 (46)	6 (46)	4 (31)	4 (31)	3 (23)	3 (23)	3 (23)	2 (15)	2 (15)	2 (15)	
Predictor significance, n (%)	10/11 (91)	8/11 (72)	6/8 (75)	1/7 (14)	2/6 (33)	3/6 (50)	2/4 (50)	2/4 (50)	2/3 (67)	1/3 (33)	0/3 (0)	1/2 (50)	1/2 (50)	1/2 (50)	

^a Studies that performed both uni- and multivariable analysis; ^b Studies that performed only univariable analysis; ✓ = statistically significant association; X = no statistically significant association; AJCC American Joint Committee on Cancer.

3.5. Grading

Heidsma et al., Pulvirenti et al., and Zou et al. all reported grading following the 2017 WHO guidelines [11,40,43]. Two studies failed to report which grading guideline was used [33,39]. Two used the Hochwald criteria for histologic grading [32,38]. The other studies followed the 2010 WHO tumor grading system.

3.6. Discrimination

Discrimination of the prediction models, expressed in *c* statistic or AUC, was described in all studies (Table 2). However, Zhou et al. only provided the *c* statistic for OS, despite presenting their model as a predictor for recurrence also [37]. The *c* statistic for predicting RFS ranged between 0.67 and 0.94. The *c* statistic for predicting OS ranged from 0.69 to 0.83 [36,37,41]. Primavesi et al. tested their model development for three different outcomes (RFS, DSS, OS), of which it worked best for the prediction of DSS (77.3 [67.2–87.5]). However, it had the lowest discrimination score of the included models, with an AUC of 66.5 [36]. The highest discrimination was reported by the prediction model of Zou et al., with a *c* statistic of 0.94 for 5-year RFS [40].

3.7. Calibration

Calibration was reported by seven studies [11,30,31,34,39,40,43]. Although all studies concluded that their model had good agreement of calibration, differences were present. The calibration curves presented by Pulvirenti et al. and Wei et al. showed a relatively high tendency to underpredict the risk of recurrence compared to the other models [11,39]. Genç et al. did not present their calibration curves but did provide a Hosmer–Lemeshow test with a Chi-square of 11.25, *p* = 0.258, indicating good calibration [34].

Four studies performed an internal validation using bootstrapping [11,30,35,39]. One performed a split-sample validation and had a median *c* statistic of 0.78 (range 0.74–0.84) [33]. Two studies reported the *c* statistic both before and after internal validation [30,41]. Dong et al. performed a bootstrapping method with 5000 iterations and went from a *c* statistic for RFS of 0.75 (95% CI 0.66–0.79) to 0.71 (0.65–0.75). Viúdez et al. reported a *c* statistic of 0.80 before and 0.78 after internal validation. The *c* statistic for OS was 0.79 before and 0.76 after internal validation.

Three out of thirteen original prediction models were externally validated [11,34,39]. The prediction model by Genç et al. was externally validated by Heidsma et al. [34,43]. The *c* statistic of the development model was 0.81 (95% CI 0.75–0.87) and 0.77 (0.71–0.83) in the validation model. Heidsma et al. performed a sub-analysis in patients with a tumor > 2 cm, in which the model performed better with a *c* statistic of 0.79 (0.73–0.85). Pulvirenti et al. developed a model with a *c* statistic of 0.85, which remained comparable after externally validating it in a separate international cohort (*c* statistic 0.84 [0.79–0.88]) [11]. The development model by Wei et al. had a *c* statistic of 0.92 (0.88–0.95) and a *c* statistic of 0.86 (0.80–0.93) in the external validation [39].

3.8. Critical Appraisal

Methodological assessment of the included studies showed poor overall quality, primarily in the Analysis domain (Table S1). All developed models were scored positive for risk of bias. Most studies did not have the minimally required number of patients with an event (recurrence) in the development cohorts for the prevention of overfitting. Dong et al. and Fisher et al. were the only two models with enough events per variable, 66 events and 60 events, respectively [30,33]. However, Fisher et al. developed their preoperative prediction model using postoperatively determined tumor grade [33]. Only the models reported by Genç et al., Pulvirenti et al., and Wei et al. were externally validated [11,34,39]. The models developed by Wei et al. and Viúdez et al. were flagged for concerns of applicability due to the need for highly specific predictors (Immunoscore and Immunohistochemistry Prognostic Score, respectively) in order to use these models in daily clinical practice [39,41].

4. Discussion

This first systematic review to evaluate existing prediction models for recurrence in patients with resectable grade 1 and 2 NF-pNET found 13 model development studies and one validation study. Most models were presented as scoring systems, followed by nomograms and modified staging systems. The most frequently incorporated risk factor was tumor grade, which had the highest rate of significant association with recurrence after regression analyses (91%). However, 10/13 (76.9%) models were not externally validated, thereby hindering the progress towards clinical implementation. The studies that did perform an external validation also performed the minimally required performance tests of prediction model development, i.e., discrimination, calibration, internal validation, and external validation.

Taking the outcomes of the critical appraisal into account, the results of the development models by Genç et al. (*c* statistic 0.77), Wei et al. (0.86), and Pulvirenti et al. (0.84) appear to be the most reliable since these are the only studies that performed the minimally required tests [11,34,39]. The predictors used by these three models were tumor grade/Ki-67 index, positive lymph nodes, tumor size, vascular and/or perineural invasion, metastasis, and 'Immunoscore'. Tumor grade was incorporated in all three models. The incorporated predictors are in line with the significant predictors described in the literature [45].

Genç et al. and Wei et al. both dichotomized their continuous variables, including the Ki-67 index [34,39]. Dichotomizing a continuous variable leads to a loss of predictive ability due to an assumption of a constant level of risk above and below the threshold [46]. A separate study by Genç et al. showed that variations within the margins of WHO grade 2 (Ki-67 index 3–20%) led to significant differences in recurrence rate [12]. Lopez-Aguilar et al. reported that these significant changes in prognosis even occur within the margins of grade 1 (0–2.99%) [47]. In turn, dichotomized variables are user-friendlier than continuous variables and will likely result in the more frequent use of the prediction model. For instance, frequent internationally used models, such as the CHA2DS2VASc-model for calculating stroke risk for patients with atrial fibrillation [48] and the Wells' criteria model for predicting the risk of a pulmonary embolism [49], also use dichotomized variables. As such, the loss of the predictive ability of the variable does not outweigh the benefit of

dichotomizing it since one of the major pitfalls of prediction models is that they are seldom used in clinical practice.

Wei et al. developed a unique predictor where they quantified the immune response in the tumor microenvironment into the Immunoscore [39]. A low Immunoscore, i.e., a pattern of low peritumoral inflammatory activity and high intratumoral CD8+ activity, was associated with a better RFS. Several studies have reported significant associations between immune response patterns in the tumor microenvironment and NET prognosis [50–52]. Takahashi et al. found that certain immune patterns were complementary to the WHO 2017 grading system, providing an argument for a possible augmenting effect if combined [53]. Immunoscores could be a powerful addition in the future; however, the scarcity of literary evidence speaks against suggesting this model for immediate implementation. Moreover, the highly specific tests required to determine the immune response patterns and the risk of interobserver variations make it less likely that the model would be internationally used in daily clinical practice.

Contrary to the other models, Pulvirenti et al. preserved the continuous function of its numeric predictors in their nomogram, i.e., Ki-67 index, tumor size, and positive lymph nodes [11]. A possible confounding factor is that they combined vascular invasion and perineural invasion into a single predictor, while the variables were associated with different hazard ratios (8.55 [95% CI 5.14–14.21] and 5.91 [3.72–9.40], respectively). Thereby risking underprediction of RFS when only vascular invasion is present, overprediction when only perineural invasion is present, and losing predictive strength when both predictors are present. Additionally, calibration of the nomogram by Pulvirenti et al. showed a tendency to underpredict if the 5-year RFS was between 55 and 80%. However, in clinical practice, their model could be used to identify patients with low-risk tumors (>80% chance of 5-year RFS) that would benefit from a low-frequency postoperative follow-up.

Selecting the correct population is particularly important in models predicting outcomes in pNET. They exhibit heterogeneous behavior, which challenges accurate risk stratification [54]. Current guidelines now recommend watchful waiting for asymptomatic NF-pNET ≤ 2 cm due to its favorable prognosis [7,14,55]. Yet, a total of 39.8% of patients from 8 of the development cohorts in this review had a tumor of ≤ 2 cm [31,34,35,37–41]. The model by Genç et al. was additionally externally validated by Heidsma et al. for pNET > 2 cm and resulted in better discrimination [34,43]. The risk of recurrence is likely to be underestimated by a prediction model if a large proportion of indolent tumors are included. Future studies should exclude these small, low-grade tumors to increase the relevance of its population when studying resectable cases or perform a separate analysis for this group.

The main practical advantage of using a prediction model is its ability to discriminate patients with a low risk for recurrence from high-risk patients. The recurrence risk threshold up to which pNET can be considered low-risk tumors differed among the studies included in this review from 3.1% to 19.9% [11,31,33,35,37–40,43]. Prospective studies using recurrence prediction models are needed to determine the optimal cut-off value, as well as the optimal timing for postoperative imaging.

The results of this review should be interpreted considering several limitations. First, the studies included in this systematic review are prone to publication bias since poorly performing prediction models are rarely publicized. Second, a frequent risk of bias was the low number of events per variable due to small sample sizes. For the screening of a single predictor, the rule of thumb states that at least 10 events are required to eliminate bias and preferably more [56–58]. However, the studies developing a prediction model for recurrence screened a median of 11 predictors, while the median number allowed was less than 2. This means that the predictors that were selected are susceptible to overfitting. Third, predictor selection based on univariable analysis could result in omitting potentially relevant predictors or inclusion on the basis of accidental association. For example, as seen in Table 3, tumor grade shows a strong association with RFS. The omission of this predictor by Sun et al., due to it not reaching statistical significance after univariable analysis, probably resulted in a less effective prediction model [31]. Preferably, predictors

should be chosen based on their relationship to the predicted outcome and not solely on the basis of their statistical significance [59]. However, the use of univariable analysis to choose predictors for the multivariable analysis is common practice and, if handled correctly, should not pose a problem for the reliability of a model. Fourth, the most important bias-introducing factor was the lack of external validation. The stated accuracy of a prediction model cannot be made certain without it, thus limiting the applicability of the results of this systematic review. It provides a strong argument for multicenter collaboration when developing and validating prediction models for patients with an NF-pNET. It is otherwise nearly impossible to obtain sufficient events (i.e., recurrence) to create a reliable prediction model.

The main strength of the current systematic review is the strict list of eligibility criteria that was used to create a homogeneous study population, thereby limiting confounding factors that influence the risk of recurrence. This study also provides an overview of the current landscape of developed prediction models as well as an overview of the predictive strength of the analyzed and included predictors. This information could be used for the selection of predictors for multivariable regression analysis when developing a prediction model.

5. Conclusions and Future Directions

Despite the development of several prediction models for the recurrence of NF-pNET after surgery, they are rarely evaluated on their performance in other participant data. To judge the true performance of a prediction model, it must be externally validated to evaluate model overfitting or deficiencies in the statistical modeling [29,60]. None of the models have undergone impact analysis in prospective studies. This might be a reason why none of the models are incorporated into (international) guidelines. Preferably, a model should be investigated in multiple external validation studies and later, ideally, also in a (randomized) controlled trial [61]. We propose to test the included prediction models in a large, multinational cohort to compare their performance. The best-performing prediction model could thereafter be applied in both clinical and trial settings to determine risk outcomes in pNET. In turn, models presented as an online calculator may be the most suitable for regular use since they are easier to access for clinicians and insightful for patients. The only model currently available online is the model by Genç et al. [34]. We highly recommend all authors developing a prediction model to calibrate, discriminate, and validate their results and to implement the model by creating an online calculator.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cancers15051525/s1>, Table S1: Critical appraisal table.

Author Contributions: Conceptualization, J.W.C., C.M.H. and E.K.; methodology, J.W.C. and C.M.H.; data curation, J.W.C. and C.M.H.; formal analysis, J.W.C. and C.M.H.; writing—original draft preparation, J.W.C. and C.M.H.; writing—review and editing, A.F.E., S.v.D., M.F., M.G.B. and E.J.M.N.v.D.; supervision, A.F.E. and E.J.M.N.v.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: All data are provided within this article or supplementary material.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Search Term (1883 Hits)

(pancreas neuroendocrine[tiab] OR pancreatic neuroendocrine[tiab] OR pancreas neuroendocrine[tiab] OR pancreatic neuro-endocrine[tiab] OR ("Pancreatic Neoplasms"[Mesh] AND (neuroendocrine[tiab] OR neuro-endocrine[tiab])) OR pnet[tiab] OR pnen[tiab] OR pancreatic NET[tiab] OR panNEN[tiab] OR panNENs[tiab]) AND ("Predictive Value of Tests"[Mesh] OR "Nomograms"[Mesh] OR predict*[tiab] OR score[tiab] OR scores[tiab] OR scoring system[tiab] OR scoring systems[tiab] OR nomogram*[tiab] OR prognostic[tiab]).

References

1. Yao, J.C.; Hassan, M.; Phan, A.; Dagohoy, C.; Leary, C.; Mares, J.E.; Abdalla, E.K.; Fleming, J.B.; Vauthey, J.; Rashid, A.; et al. One hundred years after “carcinoid”: Epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. *J. Clin. Oncol.* **2008**, *26*, 3063–3072. [[CrossRef](#)] [[PubMed](#)]
2. Fathi, A.H.R.J.; Barati, M.; Choudhury, U.; Chen, A.; Sosa, J.A. Predicting aggressive behavior in nonfunctional pancreatic neuroendocrine tumors with emphasis on tumor size significance and survival trends: A population-based analysis of 1187 patients. *Am. Surg.* **2020**, *86*, 458–466. [[CrossRef](#)] [[PubMed](#)]
3. Genç, C.G.; Klumpen, H.J.; van Oijen, M.G.H.; van Eijck, C.H.J.; Nieveen van Dijkum, E.J.M. A Nationwide Population-Based Study on the Survival of Patients with Pancreatic Neuroendocrine Tumors in The Netherlands. *World J. Surg.* **2018**, *42*, 490–497. [[CrossRef](#)] [[PubMed](#)]
4. Zerbi, A.; Falconi, M.; Rindi, G.; Delle Fave, G.; Tomassetti, P.; Pasquali, C.; Capitanio, C.; Boninsegna, L.; Di Carlo, V.; AISP-Network Study Group. Clinicopathological features of pancreatic endocrine tumors: A prospective multicenter study in Italy of 297 sporadic cases. *Am. J. Gastroenterol.* **2010**, *105*, 1421–1429. [[CrossRef](#)] [[PubMed](#)]
5. Strosberg, J.R.; Cheema, A.; Weber, J.M.; Ghayouri, M.; Han, G.; Hodul, P.J.; Kvolis, L.K. Relapse-free survival in patients with nonmetastatic, surgically resected pancreatic neuroendocrine tumors: An analysis of the AJCC and ENETS staging classifications. *Ann. Surg.* **2012**, *256*, 321–325. [[CrossRef](#)]
6. Akirov, A.; Larouche, V.; Alshehri, S.; Asa, S.L.; Ezzat, S. Treatment Options for Pancreatic Neuroendocrine Tumors. *Cancers* **2019**, *11*, 828. [[CrossRef](#)]
7. Falconi, M.; Eriksson, B.; Kaltsas, G.; Bartsch, D.K.; Capdevila, J.; Caplin, M.; Koskudla, B.; Kwekkeboom, D.; Rindi, G.; Klöppel, G.; et al. ENETS consensus guidelines update for the management of patients with functional pancreatic neuroendocrine tumors and non-functional pancreatic neuroendocrine tumors. *Neuroendocrinology* **2016**, *103*, 153–171. [[CrossRef](#)]
8. Pavel, M.; O’Toole, D.; Costa, F.; Capdevila, J.; Gross, D.; Kianmanesh, R.; Krenning, E.; Knigge, U.; Salazar, R.; Pape, U.-F.; et al. ENETS Consensus Guidelines Update for the Management of Distant Metastatic Disease of Intestinal, Pancreatic, Bronchial Neuroendocrine Neoplasms (NEN) and NEN of Unknown Primary Site. *Neuroendocrinology* **2016**, *103*, 172–185. [[CrossRef](#)]
9. Ambrosini, V.; Kunikowska, J.; Baudin, E.; Bodei, L.; Bouvier, C.; Capdevila, J.; Cremonesi, M.; de Herder, W.; Dromain, C.; Falconi, M.; et al. Consensus on molecular imaging and theranostics in neuroendocrine neoplasms. *Eur. J. Cancer* **2021**, *146*, 56–73. [[CrossRef](#)]
10. Knigge, U.; Capdevila, J.; Bartsch, D.K.; Baudin, E.; Falkerby, J.; Kianmanesh, R.; Kos-Kudla, B.; Niederle, B.; Nieveen van Dijkum, E.J.M.; O’Toole, D.; et al. ENETS Consensus Recommendations for the Standards of Care in Neuroendocrine Neoplasms: Follow-Up and Documentation. *Neuroendocrinology* **2017**, *105*, 310–319. [[CrossRef](#)]
11. Pulvirenti, A.; Javed, A.A.; Landoni, L.; Jamieson, N.B.; Chou, J.F.; Miotto, M.; He, J.; Pea, A.; Tang, L.H.; Nessi, C.; et al. Multi-institutional Development and External Validation of a Nomogram to Predict Recurrence after Curative Resection of Pancreatic Neuroendocrine Tumors. *Ann. Surg.* **2021**, *274*, 1051–1057. [[CrossRef](#)]
12. Genç, C.G.; Falconi, M.; Partelli, S.; Muffatti, F.; van Eeden, S.; Doglioni, C.; Klumpen, H.J.; van Eijck, C.H.J.; Nieveen van Dijkum, E.J.M. Recurrence of Pancreatic Neuroendocrine Tumors and Survival Predicted by Ki67. *Ann. Surg. Oncol.* **2018**, *25*, 2467–2474. [[CrossRef](#)]
13. Zaidi, M.Y.; Lopez-Aguiar, A.G.; Switchenko, J.M.; Lipscomb, J.; Andreasi, V.; Partelli, S.; Gamboa, A.C.; Lee, R.M.; Poultsides, G.A.; Dillhoff, M.; et al. A Novel Validated Recurrence Risk Score to Guide a Pragmatic Surveillance Strategy After Resection of Pancreatic Neuroendocrine Tumors. *Ann. Surg.* **2019**, *270*, 422–433. [[CrossRef](#)]
14. Kulke, M.H.; Anthony, L.B.; Bushnell, D.L.; De Herder, W.W.; Goldsmith, S.J.; Klimstra, D.S.; Marx, S.J.; Pasieka, J.L.; Pommier, R.F.; Yao, J.C.; et al. NANETS Treatment Guidelines: Well-Differentiated Neuroendocrine Tumors of the Stomach and Pancreas. *Pancreas* **2010**, *39*, 735–752. [[CrossRef](#)]
15. Kulke, M.H.; Shah, M.H.; Benson, A.B., 3rd; Bergsland, E.; Berlin, J.D.; Blaszkowsky, L.S.; Emerson, L.; Engstrom, P.F.; Fanta, P.; Giordano, T.; et al. National comprehensive cancer network. Neuroendocrine tumors, version 1.2015. *J. Natl. Compr. Cancer Netw.* **2015**, *13*, 78–108. [[CrossRef](#)]
16. Jensen, R.T.; Bodei, L.; Capdevila, J.; Couvelard, A.; Falconi, M.; Glasberg, S.; Kloppel, G.; Lamberts, S.; Peeters, M.; Rindi, G.; et al. Unmet Needs in Functional and Nonfunctional Pancreatic Neuroendocrine Neoplasms. *Neuroendocrinology* **2019**, *108*, 26–36. [[CrossRef](#)]
17. Moons, K.G.; Altman, D.G.; Reitsma, J.B.; Collins, G.S. New Guideline for the Reporting of Studies Developing, Validating, or Updating a Multivariable Clinical Prediction Model: The TRIPOD Statement. *Adv. Anat. Pathol.* **2015**, *22*, 303–305. [[CrossRef](#)]
18. Liberati, A.; Altman, D.G.; Tetzlaff, J.; Mulrow, C.; Gøtzsche, P.C.; Ioannidis, J.P.A.; Clarke, M.; Devereaux, P.J.; Kleijnen, J.; Moher, D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ* **2009**, *339*, b2700. [[CrossRef](#)]
19. Amin, M.B.; Greene, F.L.; Edge, S.B.; Compton, C.C.; Gershenwald, J.E.; Brookland, R.K.; Meyer, L.; Gress, D.M.; Byrd, D.R.; Winchester, D.P. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. *CA Cancer J. Clin.* **2017**, *67*, 93–99. [[CrossRef](#)]
20. Lloyd, R.V.O.R.; Klöppel, G.; Rosai, J. *WHO Classification of Tumours of Endocrine Organs*, 4th ed.; International Agency for Research on Cancer: Lyon, France, 2017; p. 355.

21. Luo, G.; Javed, A.; Strosberg, J.R.; Jin, K.; Zhang, Y.; Liu, C.; Xu, J.; Soares, K.; Weiss, M.J.; Zheng, L.; et al. Modified Staging Classification for Pancreatic Neuroendocrine Tumors on the Basis of the American Joint Committee on Cancer and European Neuroendocrine Tumor Society Systems. *J. Clin. Oncol.* **2017**, *35*, 274–280. [[CrossRef](#)]
22. Moons, K.G.; de Groot, J.A.; Bouwmeester, W.; Vergouwe, Y.; Mallett, S.; Altman, D.G.; Reitsma, J.B.; Collins, G.S. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: The CHARMS checklist. *PLoS Med.* **2014**, *11*, e1001744. [[CrossRef](#)] [[PubMed](#)]
23. Wolff, R.F.; Moons, K.G.M.; Riley, R.D.; Whiting, P.F.; Westwood, M.; Collins, G.S.; Reitsma, J.B.; Kleijnen, J.; Mallett, S. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann. Intern. Med.* **2019**, *170*, 51–58. [[CrossRef](#)] [[PubMed](#)]
24. Steyerberg, E.W.; Moons, K.G.M.; van der Windt, D.A.; Hayden, J.A.; Perel, P.; Schroter, S.; Riley, R.D.; Hemingway, H.; Altman, D.G. Prognosis Research Strategy (PROGRESS) 3: Prognostic model research. *PLoS Med.* **2013**, *10*, e1001381. [[CrossRef](#)]
25. Altman, D.G.; Vergouwe, Y.; Royston, P.; Moons, K.G.M. Prognosis and prognostic research: Validating a prognostic model. *BMJ* **2009**, *338*, b605. [[CrossRef](#)] [[PubMed](#)]
26. Royston, P.; Moons, K.G.M.; Altman, D.G.; Vergouwe, Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ* **2009**, *338*, b604. [[CrossRef](#)]
27. Harrell, F.E. *Regression Modeling Strategies: With Applications to Linear Models Ir, and Survival Analysis*; Springer: New York, NY, USA, 2001.
28. Moons, K.G.; Kengne, A.P.; Grobbee, D.E.; Royston, P.; Vergouwe, Y.; Altman, D.G.; Woodward, M. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* **2012**, *98*, 691–698. [[CrossRef](#)]
29. Steyerberg, E.W. *Clinical Prediction Models: A Practical Approach to Development V, and Updating*; Springer: New York, NY, USA, 2009.
30. Dong, D.H.; Zhang, X.F.; Lopez-Aguilar, A.G.; Poultides, G.; Rocha, F.; Weber, S.; Fields, R.I.; Idrees, K.; Cho, C.; Maithel, S.K.; et al. Recurrence of Non-functional Pancreatic Neuroendocrine Tumors After Curative Resection: A Tumor Burden-Based Prediction Model. *World J. Surg.* **2021**, *45*, 2134–2141. [[CrossRef](#)]
31. Sun, H.T.; Zhang, S.L.; Liu, K.; Zhou, J.J.; Wang, X.X.; Shen, T.T.; Song, X.H.; Guo, Y.L.; Wang, X.L. MRI-based nomogram estimates the risk of recurrence of primary nonmetastatic pancreatic neuroendocrine tumors after curative resection. *J. Magn. Reson. Imaging* **2019**, *50*, 397–409. [[CrossRef](#)]
32. Ballian, N.; Loeffler, A.G.; Rajamanickam, V.; Norstedt, P.A.; Weber, S.M.; Cho, C.S. A simplified prognostic system for resected pancreatic neuroendocrine neoplasms. *HPB* **2009**, *11*, 422–428. [[CrossRef](#)]
33. Fisher, A.V.; Lopez-Aguilar, A.G.; Rendell, V.R.; Pokrzywa, C.; Rocha, F.G.; Kanji, Z.S.; Poultides, G.A.; Eleftherios, A.M.; Dillhoff, M.E.; Beal, E.W.; et al. Predictive Value of Chromogranin A and a Pre-Operative Risk Score to Predict Recurrence after Resection of Pancreatic Neuroendocrine Tumors. *J. Gastrointest. Surg.* **2019**, *23*, 651–658. [[CrossRef](#)]
34. Genç, C.G.; Jilesen, A.P.; Partelli, S.; Falconi, M.; Muffatti, F.; van Kemenade, F.J.; van Eeden, S.; Verheij, J.; van Dieren, S.; van Eijck, C.H.J.; et al. A New Scoring System to Predict Recurrent Disease in Grade 1 and 2 Nonfunctional Pancreatic Neuroendocrine Tumors. *Ann. Surg.* **2018**, *267*, 1148–1154. [[CrossRef](#)]
35. Sho, S.; Court, C.M.; Winograd, P.; Toste, P.A.; Pisegna, J.R.; Lewis, M.; Donahue, T.R.; Hines, O.J.; Reber, H.A.; Dawson, D.W.; et al. A Prognostic Scoring System for the Prediction of Metastatic Recurrence Following Curative Resection of Pancreatic Neuroendocrine Tumors. *J. Gastrointest. Surg.* **2019**, *23*, 1392–1400. [[CrossRef](#)]
36. Primavesi, F.; Andreasi, V.; Hoogwater, F.J.H.; Partelli, S.; Wiese, D.; Heidsma, C.; Cardini, B.; Klieser, E.; Marsoner, K.; Fröschl, U.; et al. A Preoperative Clinical Risk Score Including C-Reactive Protein Predicts Histological Tumor Characteristics and Patient Survival after Surgery for Sporadic Non-Functional Pancreatic Neuroendocrine Neoplasms: An International Multicenter Cohort Study. *Cancers* **2020**, *12*, 1235. [[CrossRef](#)]
37. Zhou, B.; Zhan, C.; Wu, J.; Liu, J.; Zhou, J.; Zheng, S. Prognostic significance of preoperative gamma-glutamyltransferase to lymphocyte ratio index in nonfunctional pancreatic neuroendocrine tumors after curative resection. *Sci. Rep.* **2017**, *7*, 13372. [[CrossRef](#)]
38. Liu, T.; Hamilton, N.; Hawkins, W.; Gao, F.; Cao, D. Comparison of WHO Classifications (2004, 2010), the Hochwald Grading System, and AJCC and ENETS Staging Systems in Predicting Prognosis in Locoregional Well-differentiated Pancreatic Neuroendocrine Tumors. *Am. J. Surg. Pathol.* **2013**, *37*, 853–859. [[CrossRef](#)]
39. Wei, M.; Xu, J.; Hua, J.; Meng, Q.; Liang, C.; Liu, J.; Zhang, B.; Wang, W.; Yu, X.; Shi, S. From the Immune Profile to the Immunoscore: Signatures for Improving Postsurgical Prognostic Prediction of Pancreatic Neuroendocrine Tumors. *Front. Immunol.* **2021**, *12*, 654660. [[CrossRef](#)]
40. Zou, S.; Jiang, Y.; Wang, W.; Zhan, Q.; Deng, X.; Shen, B. Novel scoring system for recurrence risk classification of surgically resected G1/2 pancreatic neuroendocrine tumors—Retrospective cohort study. *Int. J. Surg.* **2020**, *74*, 86–91. [[CrossRef](#)]
41. Viudez, A.; Carvalho, F.L.; Maleki, Z.; Zuharak, M.; Laheru, L.; Stark, A.; Azad, N.Z.; Wolfgang, C.L.; Baylin, S.; Herman, J.G.; et al. A new immunohistochemistry prognostic score (IPS) for recurrence and survival in resected pancreatic neuroendocrine tumors (PanNET). *Oncotarget* **2016**, *7*, 24950–24961. [[CrossRef](#)]
42. Hochwald, S.N.; Zee, S.; Conlon, K.C.; Colleoni, R.; Louie, O.; Brennan, M.F.; Klimstra, D.S. Prognostic factors in pancreatic endocrine neoplasms: An analysis of 136 cases with a proposal for low-grade and intermediate-grade groups. *J. Clin. Oncol.* **2002**, *20*, 2633–2642. [[CrossRef](#)]

43. Heidsma, C.M.; van Roessel, S.; van Dieren, S.; Engelsman, A.F.; Strobel, O.; Buechler, M.W.; Schimmack, S.; Perinel, J.; Adham, M.; Deshpande, V.; et al. International Validation of a Nomogram to Predict Recurrence after Resection of Grade 1 and 2 Non-Functioning Pancreatic Neuroendocrine Tumors. *Neuroendocrinology* **2021**, *112*, 571–579. [[CrossRef](#)]
44. Dong, D.H.; Zhang, X.F.; Lopez-Aguiar, A.G.; Poultides, G.; Makris, E.; Rocha, F.; Kanji, Z.; Weber, S.; Fisher, A.; Fields, R.; et al. Tumor burden score predicts tumor recurrence of non-functional pancreatic neuroendocrine tumors after curative resection. *HPB* **2020**, *22*, 1149–1157. [[CrossRef](#)] [[PubMed](#)]
45. Andreasi, V.; Ricci, C.; Partelli, S.; Guarneri, G.; Ingaldi, C.; Muffatti, F.; Crippa, S.; Casadei, R.; Falconi, M. Predictors of disease recurrence after curative surgery for nonfunctioning pancreatic neuroendocrine neoplasms (NF-PanNENs): A systematic review and meta-analysis. *J. Endocrinol. Investig.* **2022**, *45*, 705–718. [[CrossRef](#)] [[PubMed](#)]
46. Steyerberg, E. *Clinical Prediction Models—A Practical Approach to Development, Validation and Updating*; Springer Nature Switzerland AG: Cham, Switzerland, 2019.
47. Lopez-Aguiar, A.G.; Ethun, C.G.; Postlewait, L.M.; Zhelnin, K.; Krasinskas, A.; El-Rayes, B.F.; Russell, M.C.; Sarmiento, J.M.; Kooby, D.A.; Staley, C.A.; et al. Redefining the Ki-67 Index Stratification for Low-Grade Pancreatic Neuroendocrine Tumors: Improving Its Prognostic Value for Recurrence of Disease. *Ann. Surg. Oncol.* **2018**, *25*, 290–298. [[CrossRef](#)] [[PubMed](#)]
48. Lip, G.Y.; Nieuwlaat, R.; Pisters, R.; Lane, D.A.; Crijns, H.J. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: The euro heart survey on atrial fibrillation. *Chest* **2010**, *137*, 263–272. [[CrossRef](#)]
49. Wells, P.S.G.J.; Anderson, D.R.; Kearon, C.; Gent, M.; Turpie, A.G.; Bormanis, J.; Weitz, J.; Chamberlain, M.; Bowie, D.; Barnes, D.; et al. Use of clinical model for safe management of patients with suspected pulmonary embolism. *Ann. Intern. Med.* **1998**, *129*, 997–1005. [[CrossRef](#)]
50. Cai, L.; Michelakos, T.; Deshpande, V.; Arora, K.S.; Yamada, T.; Ting, D.T.; Taylor, M.S.; Fernancez-Del Castillo, C.; Warshaw, A.L.; Lillemoe, K.T.; et al. Role of Tumor-Associated Macrophages in the Clinical Course of Pancreatic Neuroendocrine Tumors (PanNETs). *Clin. Cancer Res.* **2019**, *25*, 2644–2655. [[CrossRef](#)]
51. Takkenkamp, T.J.; Jalving, M.; Hoogwater, F.J.H.; Walenkamp, A.M.E. The immune tumour microenvironment of neuroendocrine tumours and its implications for immune checkpoint inhibitors. *Endocr. Relat. Cancer* **2020**, *27*, R329–R343. [[CrossRef](#)]
52. Bosch, F.; Bruwer, K.; Altendorf-Hofmann, A.; Auernhammer, C.J.; Spitzweg, C.; Westphalen, C.B.; Boeck, S.; Schubert-Fritschle, G.; Werner, J.; Heinemann, V.; et al. Immune checkpoint markers in gastroenteropancreatic neuroendocrine neoplasia. *Endocr. Relat. Cancer* **2019**, *26*, 293–301. [[CrossRef](#)]
53. Takahashi, D.; Kojima, M.; Suzuki, T.; Sugimoto, M.; Kobayashi, S.; Takahashi, S.; Konishi, M.; Gotohda, N.; Ikeda, M.; Nakatsura, T.; et al. Profiling the Tumour Immune Microenvironment in Pancreatic Neuroendocrine Neoplasms with Multispectral Imaging Indicates Distinct Subpopulation Characteristics Concordant with WHO 2017 Classification. *Sci. Rep.* **2018**, *8*, 13166. [[CrossRef](#)]
54. Dasari, A.; Shen, C.; Halperin, D.; Zhao, B.; Zhou, S.; Xu, Y.; Shih, T.; Yao, J.C. Trends in the Incidence, Prevalence, and Survival Outcomes in Patients with Neuroendocrine Tumors in the United States. *JAMA Oncol.* **2017**, *3*, 1335–1342. [[CrossRef](#)]
55. Heidsma, C.M.; Engelsman, A.F.; van Dieren, S.; Stommel, M.W.J.; de Hingh, I.; Vriens, M.; Hol, L.; Festen, S.; Hoogwater, F.J.H.; Daams, F.; et al. Watchful waiting for small non-functional pancreatic neuroendocrine tumours: Nationwide prospective cohort study (PANDORA). *Br. J. Surg.* **2021**, *108*, 888–891. [[CrossRef](#)]
56. Peduzzi, P.C.J.; Feinstein, A.R.; Holford, T.R. Importance of events per independent variable in proportional hazards regression analysis. *J. Clin. Epidemiol.* **1995**, *48*, 1503–1510. [[CrossRef](#)]
57. Ogundimu, E.O.; Altman, D.G.; Collins, G.S. Adequate sample size for developing prediction models is not simply related to events per variable. *J. Clin. Epidemiol.* **2016**, *76*, 175–182. [[CrossRef](#)]
58. Riley, R.D.; Ensor, J.; Snell, K.I.E.; Harrell, F.E., Jr.; Martin, G.P.; Reitsma, J.B.; Moons, K.G.M.; Collins, G.; van Smeden, M. Calculating the sample size required for developing a clinical prediction model. *BMJ* **2020**, *368*, m441. [[CrossRef](#)]
59. Moons, K.G.M.; Wolff, R.F.; Riley, R.D.; Whiting, P.F.; Westwood, M.; Collins, G.S.; Reitsma, J.B.; Kleijnen, J.; Mallett, S. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann. Intern. Med.* **2019**, *170*, W1–W33. [[CrossRef](#)]
60. Siontis, G.C.; Tzoulaki, I.; Castaldi, P.J.; Ioannidis, J.P. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J. Clin. Epidemiol.* **2015**, *68*, 25–34. [[CrossRef](#)]
61. Reilly, B.M.; Evans, A.T. Translating clinical research into clinical practice: Impact of using prediction rules to make decisions. *Ann. Intern. Med.* **2006**, *144*, 201–209. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.