ORIGINAL ARTICLE

ueg journal WILEY

International external validation of a stratification tool to identify branch-duct intraductal papillary mucinous neoplasms at lowest risk of progression

Kasper A. Overbeek[1] | Nikki van Leeuwen[2] | Matteo Tacelli[3] | Muhammad S. Anwar[4] | Muhammad N. Yousaf[4] | Ankit Chhoda[4] | Paolo Giorgio Arcidiacono[3] | Tamas A. Gonda[5] | Michael B. Wallace[6] | Gabriele Capurso[3,7] | James J. Farrell[4] | Djuna L. Cahen[1] | Marco J. Bruno[1]

[1]Department of Gastroenterology and Hepatology, Erasmus MC Cancer Institute, University Medical Center, Rotterdam, The Netherlands

[2]Department of Public Health, Erasmus MC University Medical Center, Rotterdam, The Netherlands

[3]Pancreas Translational and Clinical Research Centre, San Raffaele Scientific Institute IRCCS, San Raffaele Scientific Institute IRCCS, Milan, Italy

[4]Section of Digestive Diseases, Yale University School of Medicine, New Haven, Connecticut, USA

[5]Department of Medicine, NYU Langone, New York, New York, USA

[6]Department of Gastroenterology and Hepatology, Mayo Clinic, Jacksonville, Florida, USA

[7]Digestive and Liver Disease Unit, S Andrea Hospital, Rome, Italy

Correspondence
Kasper A. Overbeek, Doctor Molewaterplein 40, 3015 GD, Rotterdam, The Netherlands.
Email: k.overbeek@erasmusmc.nl

Abstract

Background: Identifying branch-duct intraductal papillary mucinous neoplasms (BD-IPMNs) at lowest risk of progression may allow for a reduced intensity of surveillance.

Objective: We aimed to externally validate the previously developed Dutch-American Risk stratification Tool (DART-1; https://rtools.mayo.edu/DART/), which identifies cysts at low risk of developing worrisome features (WFs) or high-risk stigmata (HRS).

Methods: Three prospective cohorts of individuals under surveillance for BD-IPMNs were combined, independent from the original development cohort. We assessed the performance (discrimination and calibration) of DART-1, a multivariable Cox-proportional logistic regression model with five predictors for the development of WFs or HRS.

Results: Of 832 individuals (mean age 77 years, SD 11.5) under surveillance for a median of 40 months (IQR 44), 163 (20%) developed WFs or HRS. DART-1's discriminative ability (C-statistic 0.68) was similar to that in the development cohort (0.64–0.72) and showed moderate calibration. DART-1 adequately estimated the risk for patients in the middle risk quintile, and slightly underestimated it in the lowest quintiles. Their range of predicted versus observed 3-year risk was 0%–0% versus 0%–3.7% for Q1; 0.3%–0.4% versus 3%–11% for Q2; and 2.6%–3% versus 2.4%–9.8% for Q3. The development of WFs or HRS was associated with pancreatic cancer ($p < 0.001$). Vice versa, in absence of WFs or HRS, the risk of malignancy was low (0.3%).

Conclusions: The performance of DART-1 to predict the development of WFs or HRS in BD-IPMN was validated in an external international cohort, with a

discriminative ability equal as in the development cohort. Risk estimations were most accurate for patients with BD-IPMNs in the middle risk quintile and slightly underestimated in the lowest quintiles.

KEYWORDS

intraductal papillary mucinous neoplasm, pancreatic cyst, prediction, prognosis, screening, surveillance

INTRODUCTION

Cystic lesions of the pancreas are highly prevalent in the general population and thus often encountered as an incidental finding in asymptomatic patients. Their prevalence is 25% when assessed by magnetic resonance cholangiopancreatography, and increases with age.[1] Mucinous neoplastic cysts may progress to pancreatic ductal adenocarcinoma (PDAC),[2] at a risk as high as 50% when the main pancreatic duct (PD) is involved.[3–5] However, the vast majority of neoplastic cysts are branch-duct intraductal papillary mucinous neoplasms (BD-IPMNs), with a much lower estimated risk of malignancy, of up to 4.4%.[6–12] Therefore, most BD-IPMNs will not progress.

In an attempt to diagnose pancreatic cancer at an early stage, guidelines recommend lifelong surveillance with frequent imaging of all BD-IPMNs,[13,14] resulting in a significant burden for participants and on health care resources. Surveillance is aimed at detecting signs of malignant progression, such as a solid component or mural nodule within a cyst, a thickened cyst wall, cyst growth, or dilation of the main PD.[4,12,15–20] These characteristics, either referred to as worrisome features (WFs) and high-risk stigmata (HRS) in the international Fukuoka criteria, or as relative and absolute indications for surgery in the European guidelines, prompt for intensified surveillance or surgical resection.[13,14] In contrast, BD-IPMNs without these features at diagnosis have a low risk of malignancy,[11,15,19,21–24] which decreases to less than 1% after several years of unremarkable follow-up.[15,22]

Therefore, a tool to accurately predict which BD-IPMNs are at risk to develop WFs or HRS will enable a more personalized approach, with reduced intensity of surveillance in individuals at lowest risk. In turn, this would lower the patient burden and costs associated with cyst surveillance. Recently, we developed a multivariable prediction model in an international cohort of individuals with BD-IPMNs from three university hospitals, named the Dutch-American Risk stratification Tool (DART-1, accessible at https://rtools.mayo.edu/DART/).[22] Dutch-American Risk stratification Tool identifies BD-IPMNs with the lowest risk of developing one or more WFs or HRS within three or 5 years after diagnosis, based on five easily obtainable cyst and patient characteristics. These include multifocality and size of the cyst and a patient's history of smoking, acute pancreatitis, and extrapancreatic malignancy. For the current study, our objective was to externally validate DART-1 in an independent international cohort.

Key summary

What is already known?

- Intraductal papillary mucinous neoplasms (IPMNs) are highly prevalent, most of which concern branch-duct IPMNs, the vast majority of which will not progress to pancreatic cancer because of their low estimated malignancy risk.
- In an attempt to diagnose pancreatic cancer in an early stage, current guidelines recommend lifelong surveillance for all branch-duct IPMNs, leading to a large burden of surveillance on patients and health care resources.
- Previously, we developed the Dutch-American Risk stratification Tool (DART-1), a multivariable cox-proportional logistic regression model that identifies branch-duct IPMNs at low risk of developing worrisome features or high-risk stigmata.

What are the new findings?

- DART-1 was validated in an international external cohort, demonstrating an equal discriminative ability as in the development cohort.
- Risk estimations were most accurate for patients with branch-duct IPMNs in the middle risk quintiles, slightly underestimated in the lowest quintiles, and slightly overestimated in the highest quintiles.
- Patients with the lowest-risk IPMNs as assessed by DART-1 are potential candidates for a reduced intensity of surveillance.

METHODS

Study design

We combined prospectively maintained databases on pancreatic cyst surveillance from three university hospitals: the Yale New Haven Hospital (New Haven, USA), the San Raffaele Scientific Institute (Milan, Italy), and the San Andrea Hospital (Rome, Italy). The cohorts from the two Italian centers have been described previously.[25] The cyst surveillance databases received

institutional review board approval in all centers (2000020031 Yale, 133/2016 San Raffaele, 251/2012 San Andrea). Written informed consent was obtained from all Italian participants prior to enrollment. In Yale, the institutional review board exempted the study from requiring informed consent. The study was performed according to the declaration of Helsinki and this manuscript complies with the statement for the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD).[26]

Participants

The cohorts included individuals who started pancreatic cyst surveillance between 2002 and 2018 (Yale) and 2009 and 2018 (Italy). Surveillance was executed as recommended by the clinical guidelines, with periodic MRI and/or endoscopic ultrasonography. In line with the development cohort,[22] we selected patients with a radiologically presumed BD-IPMN without WFs or HRS at diagnosis (Table 1), who were followed for at least 12 months.

Endpoint and predictors

The endpoint and predictors were identical to those used in the development cohort. The endpoint was the development of one or more WFs or HRS according the 2012 international Fukuoka guidelines,[27] at any moment during follow-up (Table 1). The five predictors of DART-1 were assessed at baseline. They included: ever having smoked (as reported by the patient, hazard ratio (HR) in the development cohort 1.40, 95%CI 0.95–2.04); a history of acute pancreatitis (HR 2.07, 95%CI 1.21–3.55); a history of extrapancreatic malignancy (both as reported in electronic medical records, HR 1.34, 95%CI 0.91–1.97); BD-IPMN multifocality (unifocal for a single unilocular or multilocular BD-IPMN or a cluster of indiscernible cysts, multifocal for clearly separate cystic lesions, HR 1.49, 95%CI 1.01–2.18); and the size of the largest BD-IPMN (HR 1.12/mm, 95%CI 1.09–1.15).

Statistical analysis

Differences in patient and cyst characteristics between the cohorts were assessed using the independent samples T-test and Mann-Whitney U test for continuous variables, and the Chi-Square test for categorical variables. A p value of < 0.05 (two-sided) was considered statistically significant.

Missing predictor data were imputed using multiple imputation (5 iterations).[28,29] The discriminative performance of the model was assessed with the Concordance statistic (C-statistic), which varies between 0.5 (a non-informative model) and 1.0 (a perfect model), and is the probability that a patient who developed WFs or HRS receives a higher risk score than a patient who did not. The model was calibrated by assessing the agreement of the predicted risk with the observed risk. First, the predicted 3-year and 5-year risk of developing WFs or HRS was calculated for all patients. Subsequently, the patients were divided in quintiles based on these predicted risks. In these five patient groups, the mean observed 3-year and 5-year outcomes were calculated. Calibration plots were created to visualize the agreement between the predicted and observed risks for each quintile. Statistical analysis was performed using Statistical Product and Service Solutions Statistics 23 (IBM Corporation) and R Software version 3.3.5 (R foundation for statistical computing).

RESULTS

Participants and predictors

The validation cohort consisted of 832 patients (363 from Yale and 469 from Italy). Compared to the development cohort (N = 875), they were of the same age (mean 66 years, SD 11.5, p = 0.19, Table 2) and sex (37% male, p = 0.886), but had a lower prevalence of diabetes mellitus (17% vs. 20%, p = 0.006) and body mass index (mean 26, SD 5.2, vs. mean 27, SD 4.9, p = 0.001), as both were lower in Italian participants. There were no differences in smoking and history of extrapancreatic malignancy. A history of acute pancreatitis was less prevalent in the validation cohort (5% vs. 8%, p = 0.004), again because of a low prevalence in the Italian subcohort of 1%. In the

TABLE 1 The composite endpoint of the Dutch-American Risk stratification Tool (DART-1), consisting of the development of one or more worrisome features or high-risk stigmata as listed by the 2012 international Fukuoka criteria.[27]

High-risk stigmata	Worrisome features
Jaundice	Cyst size ≥3 cm
Enhancing solid component	Thickened enhanced cyst wall
Main pancreatic duct ≥10 mm	Main pancreatic duct 5–9 mm
Cytology suspicious or positive for malignancy	Non-enhancing mural nodule
	Abrupt change in pancreatic duct with distal atrophy
	Lymphadenopathy

TABLE 2 Patient and cyst characteristics in comparison to the development cohort

	Validation cohort			Development cohort	Validation versus development
	USA cohort (n = 363)	Italian cohort (n = 469)	Total (N = 832)	(N = 875)	p value
Patient characteristics					
Age, mean (SD), y	68 (12.3)	65 (10.6)	66 (11.5)	66 (11.2)	0.193
Male gender, n (%)	137 (38)	171 (37)	308 (37)	321 (37)	0.886
Diabetes mellitus, n (%)	76 (21)	64 (14)	140 (17)	175 (20)	0.006
Body mass index, mean (SD)	28 (6.1)	25 (4.1)	26 (5.2)	27 (4.9)	0.001
Smoking ever, n (%)	167 (46)	161 (34)	328 (39)	342 (39)	0.563
Alcohol consumption ever, n (%)	161 (44)	129 (28)	290 (35)	372 (43)	<0.001
History of acute pancreatitis, n (%)	35 (10)	4 (1)	39 (5)	70 (8)	0.004
History of extrapancreatic malignancy, n (%)	145 (40)	119 (25)	264 (32)	291 (33)	0.411
Family history of PDAC, n (%)	31 (9)	27 (6)	58 (7)	90 (10)	0.007
Cyst characteristics					
Location dominant cyst, n (%)					
Head	125 (34)	184 (39)	309 (37)	381 (44)	0.058
Body	116 (32.0)	129 (28)	245 (29)	313 (36)	
Tail	59 (16.3)	44 (9)	103 (12)	178 (20)	
Missing	63 (17.4)	112 (24)	175 (21)	3 (0)	
Multifocality, n (%)	104 (29)	270 (58)	374 (45)	335 (38)	0.005
Largest initial size, mean (SD), mm	13 (10.1)	15 (6.1)	14 (8.1)	12 (6.4)	<0.001
Study endpoint					
Follow-up, months, median (IQR)	39 (42)	41 (48)	40 (44)	47 (40)	0.023
Development of WF or HRS, n (%)	90 (25)	73 (16)	163 (20)	116 (13)	<0.001

Abbreviations: HRS, high-risk stigmata; PDAC, pancreatic ductal adenocarcinoma; SD, standard deviation; WF, worrisome feature.

validation cohort, BD-IPMNs were more often multifocal (45% vs. 38%, $p = 0.005$) and larger (mean 14 mm, SD 8.1, vs. 12 mm, SD 6.4, $p < 0.001$). There were no missing data for the predictors cyst size, history of acute pancreatitis, and history of extrapancreatic malignancy. There was less than 1% missing data for the multifocality of the BD-IPMNs and smoking.

Endpoint and clinical outcomes

Patients were followed for 3516 person-years and a median of 40 months (IQR 44, range 12–205), which was shorter than the development cohort (3649 person-years and median 47 months, IQR 40, $p = 0.02$, Table 2). During this time, 163 individuals (20%) met the endpoint of developing one or more WFs or HRS, which was 13% in the development cohort ($p < 0.001$). There were no missing data of the endpoint, and no difference in follow-up time between those reaching the endpoint and those who did not ($p = 0.06$, Table 3).

The clinical outcomes are listed in Table 3. Overall, 13 patients (2%) developed PDAC, which was associated with the development

of WFs or HRS ($p < 0.001$). PDAC was diagnosed in 9 of the 163 (6%) patients who developed WFs or HRS versus in 4 of the 669 (1%) who did not. Seven of the 13 (54%) PDAC cases underwent surgery. The resectability rate was higher in those with WFs or HRS (78%, 7 of 9) than in those without (0%, 0 of 4, $p = 0.02$). Besides the seven operated PDAC patients, surgery was performed in another seven patients: four because of WFs or HRS (all >3 cm in size, one with a solid component), two because of cyst growth (not yet classified as WF in the 2012 Fukuoka criteria), and one for unknown reasons (this concerned a 73-year old male with a unifocal BD-IPMN of 26 mm in the tail). In these seven patients, pathology revealed only low- or moderate-grade dysplasia (Table 3).

Model performance

The C-statistic in the validation cohort was 0.68, which is equal to that in the development subcohorts (0.64 in the Erasmus subcohort, 0.71 in the Columbia subcohort, and 0.72 in the Mayo subcohort).[22]

TABLE 3 Clinical outcomes stratified on the development of worrisome features (WFs) or high-risk stigmata (HRS)

	No development of WF or HRS (n = 669)	Development of WF or HRS (n = 163)	p value
Follow-up, months, median (IQR, range)	39 (41, 12–205)	48 (59, 12–167)	0.055
Largest final cyst diameter, mean (SD), mm	15 (6.6)	27 (13.2)	<0.001
PDAC, n (%)	4 (0)	9 (6)	<0.001
Worrisome features, n (%)			
Cyst size ≥3 cm	-	91 (56)	-
Thickened enhanced cyst walls	-	18 (11)	-
Main pancreatic duct ≥5 mm	-	49 (30)	-
Non-enhancing mural nodule	-	16 (10)	-
Abrupt PD caliber change with distal atrophy	-	8 (5)	-
Abdominal lymphadenopathy	-	7 (4)	-
High-risk stigmata, n (%)			
Jaundice	-	2 (1)	-
Enhancing solid component	-	9 (6)	-
Suspicious or positive cytology	-	5 (3)	-
Underwent surgery, n (%)	3 (0)	11 (7)	<0.001
PDAC	0 (0)	7 (3)	-
Moderate-grade dysplasia	2 (0)	4 (2)	-
Low-grade dysplasia	1 (0)	0 (0)	-
Deceased, n (%)	24 (4)	20 (2)	<0.001
PDAC	2 (0)	3 (2)	-
Other cause	21 (3)	17 (10)	-
Unknown	1 (0)	0 (0)	-

Abbreviations: HRS, high-risk stigmata; IQR, interquartile range; PD, pancreatic duct; PDAC, pancreatic ductal adenocarcinoma; SD, standard deviation; WF, worrisome feature.

The calibration of DART-1 in the validation cohort showed moderate agreement between the predicted and observed 3- and 5-year risk of developing WFs or HRS. The calibration plots are shown in Figure 1 and the ranges of predicted and observed risks in Table 4. For the 3-year risk, DART-1 overestimated the risk in the highest quintile, underestimated it in the second lowest quintile, and correctly estimated the risk in the other three quintiles. For the 5-year risk, the risk was slightly overestimated in the highest two quintiles, underestimated in the lowest two, and correctly estimated in the middle quintile.

DISCUSSION

This study externally validated DART-1, a multivariable Cox-proportional logistic regression prediction model that estimates the risk of developing WFs or HRS within three and five years from diagnosis (notably not the risk of malignancy). The model performed equally well in a large independent cohort of BD-IPMNs that was temporally and geographically different from the development cohort. Instead of focusing on pancreatic cysts at high risk of malignancy, DART-1 was designed to identify the much larger group of BD-IPMNs that is unlikely to progress. Despite the low risk of malignant progression of these cysts, the current European,[13] American,[30,31] and international Fukuoka[14] cyst surveillance guidelines recommend frequent surveillance, even when cysts remain unchanged. The results of the current study suggest that DART-1 can be used to estimate the risk of patients with BD-IPMNs to develop WFs and HRS, with the most accurate prediction in the middle-risk BD-IPMNs, a slight underestimation in the lowest-risk group, and a slight overestimation in the highest-risk BD-IPMNs.

While it is tempting to propose a less intensive follow-up scheme when the DART-1 score suggests a low risk, this has to be considered in light of some potential caveats. First off, our model, although now validated in a second cohort, can be further improved in predictive performance through future updates. Furthermore, it has to be kept in mind that DART-1 does not estimate the risk of malignancy but the development of WFs and HRS. In patients with BD-IPMNs without
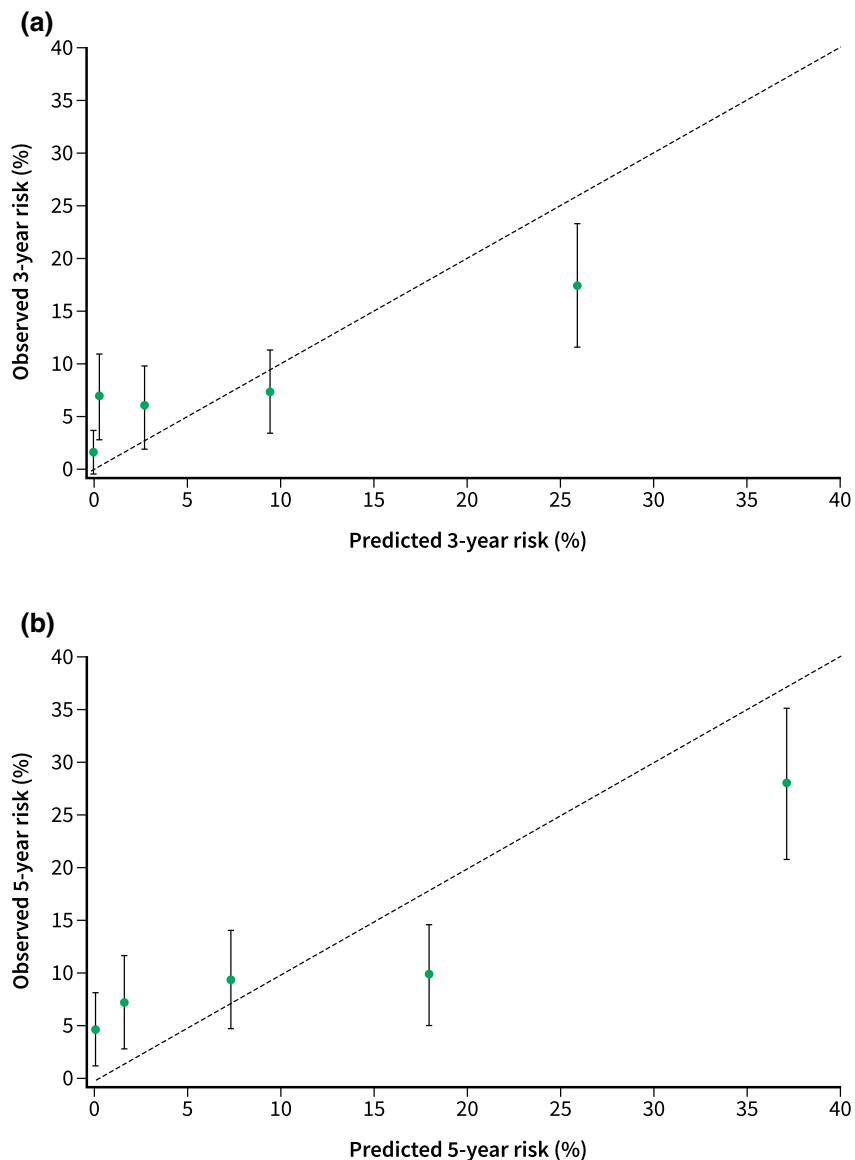
(a)

(b)

FIGURE 1 Calibration plots for the 3-year risk (a) and 5-year risk (b) of developing worrisome features or high-risk stigmata

TABLE 4 Predicted and observed 3-year and 5-year risk of developing worrisome features or high-risk stigmata for the five quintiles

	3-year % (95% CI)		5-year % (95% CI)	
	Predicted	**Observed**	**Predicted**	**Observed**
Q1	0.0 (0.0–0.0)	1.7 (−0.2–3.7)	0.0 (0.0–0.1)	5.1 (1.8–8.5)
Q2	0.3 (0.3–0.4)	7.0 (3.0–11.0)	1.6 (1.4–1.7)	7.6 (3.4–11.9)
Q3	2.8 (2.6–3.0)	6.1 (2.4–9.8)	7.3 (7.0–7.6)	9.8 (5.2–14.4)
Q4	9.5 (9.0–9.9)	7.4 (3.3–11.4)	17.9 (17.2–18.5)	10.2 (5.6–14.9)
Q5	25.8 (24.6–27.1)	17.4 (11.6–23.2)	37.0 (35.7–38.3)	28.1 (21.1–35.0)

WFs and HRS at diagnosis, the risk of malignancy is so low that even in a cohort of more than 800 patients there are too few patients who develop malignancy to perform robust statistical modeling. Instead, DART-1 uses the development of WFs and HRS as surrogate endpoint. Evidence is accumulating that BD-IPMNs that don't develop these features indeed have a very low malignancy

risk.[9,11,15,21,22] The data of our current cohort support this, with only four individuals developing pancreatic cancer out of 669 patients (0.6%) without WFs or HRS. However, most of the previously published studies are retrospective and have relatively short follow-up periods. In our development and validation studies, the follow-up period was a median 47 and 40 months. It has been shown that the risk of malignancy in BD-IPMNs increases over time, and that they can still progress after having been stable on imaging for more than 5 years.[7,11,12,21,32,33] Therefore, prospective cohorts with long-term follow-up, that provide a deeper understanding of the natural history of presumed low-risk BD-IPMNs, are required to draw definite conclusions on the long-term risk of developing a malignancy. These data are currently being accumulated in the ongoing PACYFIC (PAncreatic CYst Follow-up: an International Collaboration) study, an international prospective observational cohort study including 5000 patients under pancreatic cyst surveillance (www.pacyfic.net). Within the next several years, we will be able to update DART-1 based on the data from the PACYFIC study, to provide risk estimations for periods longer than 5 years, and to fine-tune its ability to more reliably predict the risk of developing a malignancy in BD-IPMNs. At one point it may even be possible to identify those individuals in whom surveillance could safely be discontinued.

The current study does not allow us to compare clinical outcomes of the currently recommended surveillance protocol to a hypothetical reduced surveillance frequency. It also did not estimate the optimal cut-off for reducing the intensity of surveillance. However, the advantages and implications for clinical practice can be readily illustrated. If one would use a hypothetical cut-off value of 10% risk of developing WFs or HRS within 3 years, this would

indicate that surveillance could be reduced for 72% of the patients without WFs or HRS at diagnosis. For example, for a patient with a unifocal 1-cm cyst, without a history of acute pancreatitis, smoking, or extrapancreatic malignancy, the number of surveillance investigations prescribed by current clinical guidelines are two (American Gastroenterological Association [AGA]), three (American College of Gastroenterology [ACG]), or four (European and Fukuoka) within the first 3 years. In the first 5 years this would accumulate to three (AGA), five (ACG and Fukuoka) or six (European) investigations. But as estimated by DART-1 (shown in Figure 2), this patient's risk of developing WFs or HRS is only 2.4% in the first 3 years and 5.5% in the first 5 years, and this patient would fall in the third quintile, with an observed risk of 2.5%–10% in 3 years and 5%–14% in 5 years. In addition, based on the data from the development and validation cohorts combined, if the patient does not develop WFs or HRS, the risk of pancreatic cancer within the first 40 months after diagnosis is estimated to be only 0.28% (4/1428 patients). Awaiting studies that identify the optimal risk cut-off for a reduction of surveillance frequency, and the comparison of clinical outcomes to those of current surveillance protocols, based on the current study it seems safe to lengthen intervals within the first three or five years for patients falling in the lowest three risk quintiles as estimated by DART-1.

This validation cohort was similar to the development cohort in age, sex, and the predictors smoking and history of extrapancreatic malignancy. There were also some notable differences (lower prevalence of diabetes mellitus and acute pancreatitis and lower body mass index), which stemmed mostly from differences between the Italian and American patients, as the subcohort from Yale was actually similar to the development cohort, of which 91% was also
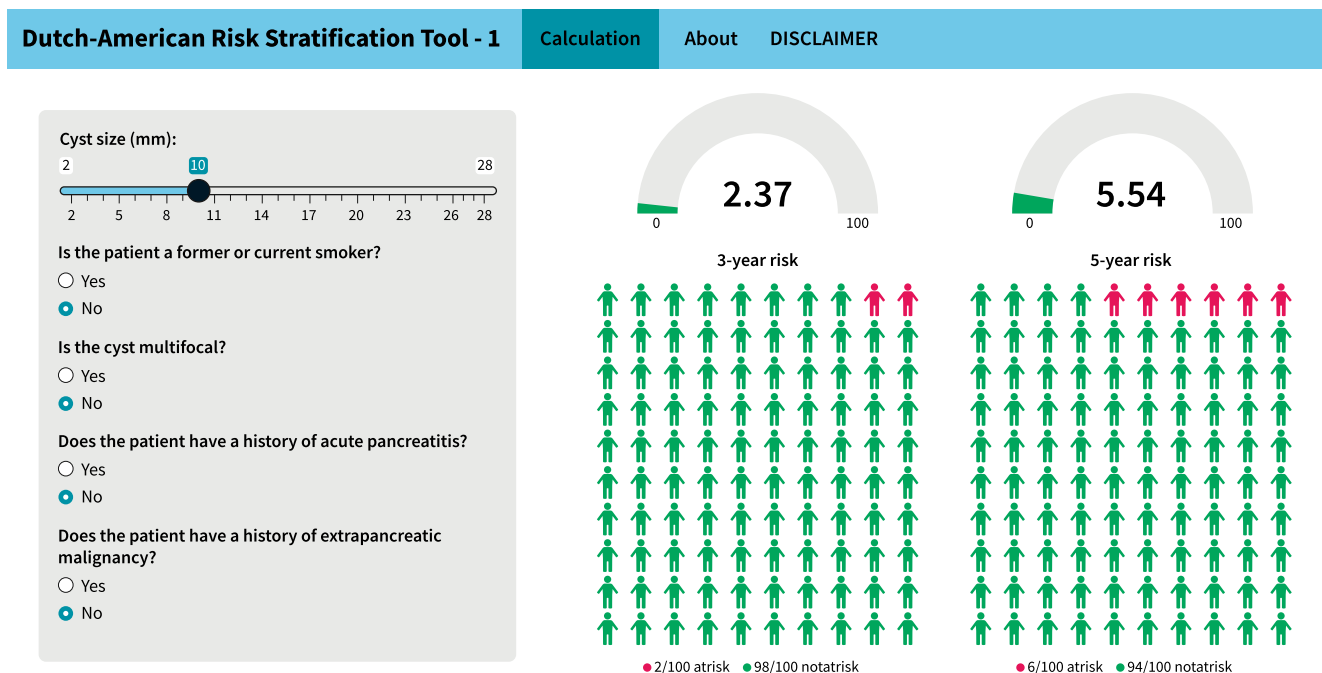

FIGURE 2 Web-based application of the Dutch-American Risk Stratification Tool (DART-1) with an example patient. The application can be found at https://rtools.mayo.edu/DART/

American. Acute pancreatitis is considered a WF when caused by the IPMN, but a past episode of acute pancreatitis is not directly seen as such.[13,14] This is a somewhat gray area, as it is often not certain whether the BD-IPMN was already present at the time of acute pancreatitis. The Italian centers were more stringent in excluding patients with a past episode of acute pancreatitis, leading to a lower prevalence in their cohort. However, when we assessed the model's performance separately for the American and Italian subcohort, this resulted in identical performances. In other words, the differences in patient and cyst characteristics did not lead to a different risk estimation and therefore, did not impact our results.

Another difference between the two cohorts was that cysts in the validation cohort were slightly more often multifocal (45% vs. 38%) and somewhat larger at diagnosis (14 vs. 12 mm). Initial cyst size and multifocality have been identified as predictors for growth and the development of WFs in both our cohorts and others,[32,34–37] and therefore, this difference could (partially) explain why the validation cohort had a higher percentage of patients developing WFs than the development cohort (20% vs. 13%). Nevertheless, both differences are small and therefore may not represent a clinically relevant difference. Overall, the differences between the development and validation cohort are unlikely to represent selection bias for two reasons. First, we included all consecutive patients under pancreatic cyst surveillance. Second, our cohort was highly comparable to a multitude of other large cohorts of BD-IPMNs without WFs or HRS at diagnosis from various geographic regions, in terms of sex (31%–54% male), median age (62–68 years), median initial cyst size (10–20 mm), cyst multifocality (30%–78%), and the development of WFs or HRS (4%–18%).[11,12,15,19,23,24,32] This should be regarded a strength of the current study.

The purpose of an external validation is to demonstrate that a model retains its usefulness in independent cohorts in different settings. In this validation, DART-1 showed an equal performance in a cohort that was completely independent from the development cohort, was under surveillance in a slightly different time period, came from a different geographical region, and had somewhat different patient and cyst characteristics. This illustrates the robustness of the model and proves that these results can be generalized to other cohorts of BD-IPMNs without WFs or HRS at diagnosis. Other strengths include that our validation cohort was equally large as the development cohort. There were almost no missing data for the predictors and the endpoint of the study, and there was a large number of index cases. As a result, the data of this cohort were of high quality and allowed for a statistically robust validation.

As mentioned, this study is limited by the incomplete knowledge of the natural history of BD-IPMNs and how long the cyst was already present at the time of diagnosis, which hampers the estimation of the long-term risk of developing WFs and HRS or malignancy. Fortunately, DART-1 does not primarily aim to identify those cysts harboring malignancy, but instead to identify those at lowest risk. Worrisome features and HRS have been proven to correlate well with malignancy,[4,19] and therefore, the continued

absence of these features during follow-up is a good indication of a stable cyst that carries no evidence of malignant progression. A limitation of our study design is that we could not include the WFs high cyst growth rate, a new diagnosis of diabetes mellitus, and elevated serum carbohydrate antigen 19-9 level, as these features were only recently added as predictive parameters in the updates of cyst surveillance guidelines.[13,14] Because of the recent introduction, there are not yet large enough cohorts of individuals with BD-IPMNs in whom all these parameters have been recorded during a sufficient follow-up period to enable statistical modeling. In future updates and iterations of DART-1, these WFs will be incorporated. A last limitation is that the long follow-up period of the surveillance cohorts means there have been slight differences in the management of BD-IPMNs due to the updates of cyst surveillance guidelines. In addition, there may have been some differences in management between the American and Italian centers. However, we do not expect this to have introduced a relevant bias, as all guidelines propose similar surveillance modalities and intervals, targeting the same WFs and HRS.

In conclusion, we externally validated the performance of DART-1 in identifying BD-IPMNs at lowest risk of developing WFs or HRS within three or five years. DART-1 had an equal discriminative ability when compared to the development cohort. Risk estimations were most accurate for the patients with BD-IPMNs in the middle risk quintiles, with slight underestimations in the lowest quintiles and slight overestimations in the highest quintiles. Based on this independent confirmation, it now seems prudent to start assessing the outcome of strategies with personalized surveillance intensities based on the risk of progression as estimated by DART-1.

CONFLICT OF INTEREST
Michael B. Wallace received research funding from Olympus, Medtronic, Boston Scientific, Fujifilm, and ChiroChem. He is a consultant to Lumendi, Virgo, Cosmo, and GI Supply. Gabriele Capurso is a consultant to Mylan. Djuna L. Cahen is a consultant to Tramedico. Marco J. Bruno received research funding from Boston Scientific, Cook Medical, and Pentax Medical. He is a consultant to Boston Scientific, Cook Medical, Pentax Medical, and Mylan. The other authors have nothing to disclose.

AUTHOR CONTRIBUTIONS
The study was initiated and designed by Djuna L. Cahen and Marco J. Bruno. The study was supervised by the principal investigators Paolo Giorgio Arcidiacono, Tamas A. Gonda, Michael B. Wallace, Gabriele Capurso, James J. Farrell, Djuna L. Cahen, and Marco J. Bruno. The study was coordinated by Kasper A. Overbeek. Data was collected by Matteo Tacelli, Muhammad S. Anwar, Muhammad N. Yousaf, and Ankit Chhoda, and critically reviewed by Kasper A. Overbeek. Data analysis was performed by Kasper A. Overbeek and Nikki van

Leeuwen. Results were interpreted by Kasper A. Overbeek, Nikki van Leeuwen, Djuna L. Cahen, and Marco J. Bruno. The manuscript was drafted by Kasper A. Overbeek and Nikki van Leeuwen and critically reviewed by Djuna L. Cahen and Marco J. Bruno. All authors reviewed and approved the final submitted manuscript.

DATA AVAILABILITY STATEMENT

Data are available upon reasonable request.

ORCID

Kasper A. Overbeek https://orcid.org/0000-0003-1829-9963
Paolo Giorgio Arcidiacono https://orcid.org/0000-0001-6692-7720
Michael B. Wallace https://orcid.org/0000-0002-6446-5785
Gabriele Capurso https://orcid.org/0000-0002-0019-8753
Djuna L. Cahen https://orcid.org/0000-0001-6753-4857
Marco J. Bruno https://orcid.org/0000-0001-9181-5499

REFERENCES

1. Zerboni G, Signoretti M, Crippa S, Falconi M, Arcidiacono PG, Capurso G. Systematic review and meta-analysis: prevalence of incidentally detected pancreatic cystic lesions in asymptomatic individuals. Pancreatology. 2019;19(1):2–9.
2. Hruban RH, Maitra A, Kern SE, Goggins M. Precursors to pancreatic cancer. Gastroenterol Clin N Am. 2007;36(4):831–49, vi.
3. Hackert T, Fritz S, Klauss M, Bergmann F, Hinz U, Strobel O, et al. Main-duct intraductal papillary mucinous neoplasm: high cancer risk in duct diameter of 5 to 9 mm. Ann Surg. 2015;262(5):875–80. Discussion 80-1.
4. Seo N, Byun JH, Kim JH, Kim HJ, Lee SS, Song KB, et al. Validation of the 2012 international consensus guidelines using computed tomography and magnetic resonance imaging: branch duct and main duct intraductal papillary mucinous neoplasms of the pancreas. Ann Surg. 2016;263(3):557–64.
5. Marchegiani G, Mino-Kenudson M, Sahora K, Morales-Oyarvide V, Thayer S, Ferrone C, et al. IPMN involving the main pancreatic duct: biology, epidemiology, and long-term outcomes following resection. Ann Surg. 2015;261(5):976–83.
6. Crippa S, Capurso G, Camma C, Delle Fave G, Castillo CF, Falconi M. Risk of pancreatic malignancy and mortality in branch-duct IPMNs undergoing surveillance: a systematic review and meta-analysis. Dig Liver Dis. 2016;48(5):473–9.
7. Pergolini I, Sahora K, Ferrone CR, Morales-Oyarvide V, Wolpin BM, Mucci LA, et al. Long-term risk of pancreatic malignancy in patients with branch duct intraductal papillary mucinous neoplasm in a referral center. Gastroenterology. 2017;153(5):1284–94.e1.
8. Malleo G, Marchegiani G, Borin A, Capelli P, Accordini F, Butturini G, et al. Observational study of the incidence of pancreatic and extrapancreatic malignancies during surveillance of patients with branch-duct intraductal papillary mucinous neoplasm. Ann Surg. 2015;261(5):984–90.
9. Kwong WT, Hunt GC, Fehmi SM, Honerkamp-Smith G, Xu R, Lawson RD, et al. Low rates of malignancy and mortality in asymptomatic patients with suspected neoplastic pancreatic cysts beyond 5 years of surveillance. Clin Gastroenterol Hepatol. 2016;14(6):865–71.
10. Munigala S, Gelrud A, Agarwal B. Risk of pancreatic cancer in patients with pancreatic cyst. Gastrointest Endosc. 2016;84(1):81–6.
11. Del Chiaro M, Ateeb Z, Hansson MR, Rangelova E, Segersvard R, Kartalis N, et al. Survival analysis and risk for progression of intraductal papillary mucinous neoplasia of the pancreas (IPMN) under surveillance: a single-institution experience. Ann Surg Oncol. 2017;24(4):1120–6.
12. Oyama H, Tada M, Takagi K, Tateishi K, Hamada T, Nakai Y, et al. Long-term risk of malignancy in branch-duct intraductal papillary mucinous neoplasms. Gastroenterology. 2020;158(1):226–37.e5.
13. European evidence-based guidelines on pancreatic cystic neoplasms. Gut. 2018;67(5):789–804.
14. Tanaka M, Fernández-Del Castillo C, Kamisawa T, Jang JY, Levy P, Ohtsuka T, et al. Revisions of international consensus Fukuoka guidelines for the management of IPMN of the pancreas. Pancreatology. 2017;17(5):738–53.
15. Marchegiani G, Andrianello S, Pollini T, Caravati A, Biancotto M, Secchettin E, et al. "Trivial" cysts redefine the risk of cancer in presumed branch-duct intraductal papillary mucinous neoplasms of the pancreas: a potential target for follow-up discontinuation? Am J Gastroenterol. 2019;114(10):1678–84.
16. Suzuki Y, Nakazato T, Yokoyama M, Kogure M, Matsuki R, Abe N, et al. Development and potential utility of a new scoring formula for prediction of malignant intraductal papillary mucinous neoplasm of the pancreas. Pancreas. 2016;45(9):1227–32.
17. Rodriguez JR, Salvia R, Crippa S, Warshaw AL, Bassi C, Falconi M, et al. Branch-duct intraductal papillary mucinous neoplasms: observations in 145 patients who underwent resection. Gastroenterology. 2007;133(1):72–9. Quiz 309-10.
18. Ridtitid W, DeWitt JM, Schmidt CM, Roch A, Stuart JS, Sherman S, et al. Management of branch-duct intraductal papillary mucinous neoplasms: a large single-center study to assess predictors of malignancy and long-term outcomes. Gastrointest Endosc. 2016;84(3):436–45.
19. Mukewar S, de Pretis N, Aryal-Khanal A, Ahmed N, Sah R, Enders F, et al. Fukuoka criteria accurately predict risk for adverse outcomes during follow-up of pancreatic cysts presumed to be intraductal papillary mucinous neoplasms. Gut. 2017;66(10):1811–7.
20. Attiyeh MA, Fernandez-Del Castillo C, Al Efishat M, Eaton AA, Gonen M, Batts R, et al. Development and validation of a multi-institutional preoperative nomogram for predicting grade of dysplasia in intraductal papillary mucinous neoplasms (IPMNs) of the pancreas: a report from the pancreatic surgery consortium. Ann Surg. 2018;267:157–63.
21. Yoshioka T, Shigekawa M, Ikezawa K, Tamura T, Sato K, Urabe M, et al. Risk factors for pancreatic cancer and the necessity of long-term surveillance in patients with pancreatic cystic lesions. Pancreas. 2020;49(4):552–60.
22. Overbeek KA, Alblas M, Gausman V, Kandel P, Schweber AB, Brooks C, et al. Development of a stratification tool to identify pancreatic intraductal papillary mucinous neoplasms at lowest risk of progression. Aliment Pharmacol Therapeut. 2019;50(7):789–99.
23. Ohno E, Hirooka Y, Kawashima H, Ishikawa T, Kanamori A, Ishikawa H, et al. Natural history of pancreatic cystic lesions: a multicenter prospective observational study for evaluating the risk of pancreatic cancer. J Gastroenterol Hepatol. 2018;33(1):320–8.
24. Boraschi P, Tarantini G, Donati F, Scalise P, Cervelli R, Caramella D. Side-branch intraductal papillary mucinous neoplasms of the pancreas: outcome of MR imaging surveillance over a 10 years follow-up. Eur J Radiol Open. 2020;7:100250.
25. Capurso G, Crippa S, Vanella G, Traini M, Zerboni G, Zaccari P, et al. Factors associated with the risk of progression of low-risk branch-duct intraductal papillary mucinous neoplasms. JAMA Netw Open. 2020;3(11):e2022933.
26. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Br J Surg. 2015;102(3):148–58.
27. Tanaka M, Fernandez-del Castillo C, Adsay V, Chari S, Falconi M, Jang JY, et al. International consensus guidelines 2012 for the management of IPMN and MCN of the pancreas. Pancreatology. 2012;12(3):183–97.

28. Steyerberg EW, van Veen M. Imputation is beneficial for handling missing data in predictive models. J Clin Epidemiol. 2007; 60(9):979.

29. van der Heijden GJ, Donders AR, Stijnen T, Moons KG. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. J Clin Epidemiol. 2006;59(10):1102–9.

30. Elta GH, Enestvedt BK, Sauer BG, Marie Lennon A. ACG clinical guideline: diagnosis and management of pancreatic cysts. Am J Gastroenterol. 2018;113(4):464–79.

31. Vege SS, Ziring B, Jain R, Moayyedi P, Clinical Guidelines C, American Gastroenterology A. American gastroenterological association institute guideline on the diagnosis and management of asymptomatic neoplastic pancreatic cysts. Gastroenterology. 2015;148(4): 819–22.

32. Crippa S, Pezzilli R, Bissolati M, Capurso G, Romano L, Brunori MP, et al. Active surveillance beyond 5 years is required for presumed branch-duct intraductal papillary mucinous neoplasms undergoing non-operative management. Am J Gastroenterol. 2017;112(7): 1153–61.

33. Lawrence SA, Attiyeh MA, Seier K, Gonen M, Schattner M, Haviland DL, et al. Should patients with cystic lesions of the pancreas undergo long-term radiographic surveillance?: results of 3024 patients evaluated at a single institution. Ann Surg. 2017;266(3): 536–44.

34. Park HW, Lee JS, Park SY, Kim TH, Lee JY, Koo JE, et al. Progression of pancreatic cystic lesions without any risk features is associated with initial cyst size and obesity. J Gastroenterol Hepatol. 2020; 35(5):877–84.

35. Kim GE, Shin SS, Kim JW, Heo SH, Lim HS, Jun CH, et al. Incidental, small (<3 cm), unilocular, pancreatic cysts: factors that predict lesion progression during imaging surveillance. Korean J Radiol. 2017; 18(6):915–25.

36. Han Y, Lee H, Kang JS, Kim JR, Kim HS, Lee JM, et al. Progression of pancreatic branch duct intraductal papillary mucinous neoplasm associates with cyst size. Gastroenterology. 2018;154(3):576–84.

37. Kayal M, Luk L, Hecht EM, Do C, Schrope BA, Chabot JA, et al. Long-term surveillance and timeline of progression of presumed low-risk intraductal papillary mucinous neoplasms. AJR. 2017;209(2):320–6.