Accepted Manuscript

Model to Determine Risk of Pancreatic Cancer in Patients with New-onset Diabetes

Ayush Sharma, Harika Kandlakunta, Sajan Jiv Singh Nagpal, Feng Ziding, William Hoos, Gloria M. Petersen, Suresh T. Chari

PII: S0016-5085(18)34543-8
DOI: 10.1053/j.gastro.2018.05.023
Reference: YGAST 61889

To appear in: *Gastroenterology*
Accepted Date: 9 May 2018

Please cite this article as: Sharma A, Kandlakunta H, Singh Nagpal SJ, Ziding F, Hoos W, Petersen GM, Chari ST, Model to Determine Risk of Pancreatic Cancer in Patients with New-onset Diabetes, *Gastroenterology* (2018), doi: 10.1053/j.gastro.2018.05.023.

Title: Model to Determine Risk of Pancreatic Cancer in Patients with New-onset Diabetes

Running title: END-PAC Risk Stratification Score

Authors: Ayush Sharma[1], Harika Kandlakunta[1], Sajan Jiv Singh Nagpal[1], Feng Ziding[2], William Hoos[3], Gloria M. Petersen[4], Suresh T. Chari[1]

Author Affiliations: [1]Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN; [2]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX; [3]Pancreatic Cancer Action Network, Manhattan Beach, CA; [4]Department of Health Science Research, Mayo Clinic, Rochester, MN

Reprints: Suresh T. Chari, MD, Division of Gastroenterology and Hepatology, Mayo Clinic, 200 First St SW, Rochester, MN 55905 (chari.suresh@mayo.edu)

Phone: 507-255-5713 Fax: 507-255-6318

Author Contributions:

Study concept and design: STC and AS

Acquisition, analysis and interpretation of data: AS, HK, SJSN and STC

Drafting of the manuscript: STC and AS

Critical revision of the manuscript for important intellectual concept: all authors

Conflict of interest: No conflict of interest declared

Text word count: 4,066

Abstract word count: 392

Abbreviations

BG, blood glucose
EAG, estimated average glucose
END-PAC, Enriching New-onset Diabetes for pancreatic ductal adenocarcinoma
FBG, fasting blood glucose
PC-NOD, pancreatic cancer new-onset diabetes
T2-NOD, type 2 new-onset diabetes

Grant support

This work was supported by the NIH U01 Consortium for Study of Chronic Pancreatitis, Diabetes and Pancreatic cancer (CPDPC) (Chari and Ziding), Pancreatic Cancer Action Network (Chari and Hoos), Kenner Family Research Fund (Chari), Prokopanko Gift to Mayo Foundation (Chari and Petersen), U01DK108328 (Ziding) and U24086368 (Ziding). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or any other funding source.

This study was made possible by the Rochester Epidemiology Project (grant no. RO1-AG034676; Principal Investigators: Walter A. Rocca, MD, MPH, and Jennifer L. St Sauver, PhD.

Abstract

Background & Aims: Of subjects with new-onset diabetes (based on glycemia) over the age of 50 years, approximately 1% are diagnosed with pancreatic cancer within 3 years. We aimed to develop and validate a model to determine risk of pancreatic cancer in individuals with new-onset diabetes.

Methods: We retrospectively collected data from 4 independent, non-overlapping cohorts of patients (n=1561) with new-onset diabetes (based on glycemia; data collected at date of diagnosis and 12 months before) in the Rochester Epidemiology Project, from January 1, 2000 through December 31, 2015 to create our model. The model weighed scores for the 3 factors identified in the discovery cohort to be most strongly associated with pancreatic cancer (64 patients with pancreatic cancer and 192 with type-2 diabetes): change in weight, change in blood glucose, and age at onset of diabetes. We called our model enriching new-onset diabetes for pancreatic cancer (END-PAC). We validated the locked-down model and cutoff score in an independent population-based cohort of 1096 patients with diabetes; of these 9 patients (.82%) had pancreatic within 3 years of meeting the criteria for new-onset diabetes.

Results: In the discovery cohort the END-PAC model identified patients who developed pancreatic cancer within 3 years of onset of diabetes with an area under the receiver operating characteristic curve value of 0.87; a score of $\geq$3 identified patients who developed pancreatic cancer with 80% sensitivity and specificity. In the validation cohort, a score of $\geq$3 identified 7/9 patients with pancreatic cancer (78%), with 85% specificity; the prevalence of pancreatic cancer in subjects with score of $\geq$3 (3.6%) was 4.4-fold more than in patients with new-onset diabetes. A high END-PAC score in subjects who did not have pancreatic cancer (false positives) was often due to such factors as recent steroid use or different malignancy. An END-PAC score <0 (in 49% of subjects) meant that patients had an extremely low-risk for pancreatic cancer. An END-PAC score $\geq$3 identified 75% of subjects in the discovery cohort >6 months before a diagnosis of pancreatic cancer.

Conclusions: Based on change in weight, change in blood glucose, and age at onset of diabetes, we developed and validated a model to determine risk of pancreatic cancer in patients with new-onset diabetes, based on glycemia (the END-PAC model). An independent, prospective study is needed to further validate this model, which could contribute to early detection of pancreatic cancer.

Keywords: END-PAC, biomarker, pancreas, screening

Introduction

Pancreatic ductal adenocarcinoma has a dismal (9%) 5-year survival[1], largely because the majority (85%) of pancreatic cancer is diagnosed at an advanced stage. Developing strategies for early detection of resectable sporadic pancreatic cancer are critical for improving survival[2]. Since pancreatic cancer is uncommon (annual incidence 37/100,000 in subjects >50 years of age[3]), a 3-step (DEF) approach to its early detection has been suggested[4]: (1) Define a high-risk group for pancreatic cancer, (2) Enrich the high-risk group further for pancreatic cancer and (3) Find the lesion in the highly enriched cohort.

The only known high-risk group for sporadic pancreatic cancer is that of subjects ≥50 years of age with glycemically defined new-onset diabetes[2]. Compared to the general population, such subjects have a 6-8-fold higher risk of being diagnosed with pancreatic cancer within 3 years of first meeting glycemic criteria for new-onset diabetes, with a 3-year incidence of pancreatic cancer being ~1% [2]. Currently new-onset diabetes in type 2 diabetes (T2-NOD) is indistinguishable from new-onset diabetes in pancreatic cancer (PC-NOD). Facilitating the utility of a clinical work-up for pancreatic cancer in new-onset diabetes requires identifying a very high-risk group for pancreatic cancer.

Three previous prospective studies [5-7] have included some form of enrichment strategy to identify pancreatic cancer among those with incident, physician-diagnosed new-onset diabetes. While the cohorts were clearly enriched for pancreatic cancer (prevalence 2.5-12%), all identified pancreatic cancers were at advanced stage[5, 6], likely due to use of markers of late cancer for risk-stratification. Two recent retrospective studies using large databases, the Veterans Administration database[8] and The Health Improvement Network (THIN) database[9] in the United

Kingdom, estimated pancreatic cancer incidence in physician-diagnosed new-onset diabetes, and proposed models for enriching the cohort for pancreatic cancer. They found the 3-year incidence of pancreatic cancer to be 0.25% and 0.4%, respectively consistent with incidence reported in studies using physician diagnosed diabetes[10], but these are significantly lower than in studies using glycemically-defined new-onset diabetes, both previous[2] and this current study. Munigala et al[8] concluded that despite a 4-fold enrichment, the incidence of pancreatic cancer in physician diagnosed new-onset diabetes is too low to warrant further study.

Nearly 60% of pancreatic cancer in new-onset diabetes occurs within 12 months of its glycemic onset[2, 11]. Since physician diagnosis of diabetes occurs months to years after diabetes onset[11-14], the strategy of using new-onset diabetes as a clue for early diagnosis of pancreatic cancer would be most effective if new-onset diabetes is identified at its glycemic onset rather than at its clinical diagnosis. Our goal was to develop a model that can be used concurrently with glycemic onset of new-onset diabetes.

We developed our model based on three previously noted features that distinguish T2-NOD from PC-NOD. While T2-NOD is often accompanied by weight gain[15], PC-NOD paradoxically occurs in the face of weight loss[16, 17]. Progression from normal fasting glucose to T2-NOD is a slow process occurring over ~8 years[18, 19], while PC-NOD progresses rapidly, over 2-3 years[17, 20]. Patients with T2-NOD are younger at diabetes diagnosis[21] than patients with pancreatic cancer[22]. In our Discovery Set of T2- and PC-NOD subjects we confirmed these features.

We created and tested various models based on these features. The best predictiveness for pancreatic cancer was provided by a model that included age, Δweight and Δblood glucose over

previous year as categorical variables. The weighted score, we call the Enriching New-onset Diabetes for Pancreatic Cancer (END-PAC) Score, classifies new-onset diabetes subjects into high-, intermediate- and low-risk groups for pancreatic cancer. We validated the score in a population-based glycemically-defined new-onset diabetes cohort. An END-PAC Score of $\geq$3 significantly enriched the new-onset diabetes cohort for pancreatic cancer, even those with >6 months lead time to pancreatic cancer diagnosis. If the extremely high risk of pancreatic cancer in the END-PAC cohort is validated, we believe that it would warrant clinical work-up for pancreatic cancer.

Patients and Methods

This study was approved by the Mayo Clinic Foundation Institutional Review Board and Olmsted Medical Center Institutional Review Board. The Rochester Epidemiology Project (REP), a unique medical records linkage system funded by NIH since 1966, collects, collates, and indexes patient-level data from all health care providers in Olmsted County, Minnesota and the surrounding 27 county area[23, 24] and allows for accurate population-based epidemiologic research.

Cohorts assembled:

We assembled the following 4 independent, non-overlapping cohorts from the REP resources: Three retrospectively identified and annotated cohorts (Discovery Set of PC-NOD [n=64], Discovery Set of T2-NOD [n=192], and a population-based new-onset diabetes Validation Set [n=1096]) and a prospectively identified cohort of new-onset diabetes subjects recruited into a pilot screening study for pancreatic cancer in new-onset diabetes (Examination of the Pancreas in New-onset Diabetes [EXPAND] trial) (n=100) (NCT0200133). All new-onset

diabetes subjects in Olmsted County between January 1st, 2000 to December 31st, 2015 (n=1561) were identified using a glycemic definition of diabetes (Supplementary material, Table 1). Among these subjects, 1288 (83%) had available the prior data on weight and blood glucose between 3 and 18 months prior as well as when first meeting new-onset diabetes criteria (paired data). Those T2-NOD with data between January 1st, 2007 to December 31st, 2008 formed the T2-NOD Discovery Set (n=192), and all new-onset diabetes subjects in the remaining years formed the Validation Set (n=1096). An independent cohort of pancreatic cancer with new-onset diabetes identified from the 28 counties covered by REP formed the Discovery Set for PC-NOD.

Clinical and laboratory data were abstracted as model development required. Paired data on weight plus values for fasting blood glucose (FBG) and/or estimated average glucose (EAG) (both generically referred to as blood glucose [BG]) at new-onset diabetes date and between 3 and 18 months prior to new-onset diabetes date. EAG was calculated as 28.7*HbA1c-46.7. Supplementary Figure 1 provides further details of how we calculated EAG and the algorithm for selection of BG and weights for calculating the final score. Overall, 1288 T2-NOD and 73 PC-NOD could be scored.

The subjects in one or more of the above-described independent cohorts were used to determine the following:

1. *Incidence of pancreatic cancer in new-onset diabetes*: This was defined from the population-based new-onset diabetes cohort from January 1st, 2000 to December 31st, 2015 (n=1561).

2. *Development of models*: Using Discovery Sets for T2- and PC-NOD a univariate (parametric and non-parametric) analysis was performed and characteristics showing significant

differences ($p<.05$) between PC- and T2-NOD subjects were included for further analysis. The 3 highly discriminatory characteristics (age, BG progression [ΔBG] and change in weight [Δweight]) of PC and T2-NOD were analyzed as continuous and as categorical variables. Three new-onset diabetes models were created and compared: Model A) weight loss of ≥2.5 kg, Model B) Δweight (kg in categories) + ΔBG (mg/dl) + age (categories) and Model C) Δweight (kg in categories) + ΔBG category + age (categories) (Tables 1 and 2). For model B and C, Δweight was categorized based on the distribution of weight change prior to new-onset diabetes date in the PC- and T2-NOD Discovery sets (supplementary Figure 2). For model C, the ΔBG categories were based on the American Diabetes Association (ADA) classification but adding 2 categories: 100 to 109 mg/dl and >160 mg/dl (Table 1). Model A had a binary score (1 or 0). For Models B and C criterion (cutoff) scores were defined from receiver operating characteristic (ROC) curves based on Youden index. Models and criterion scores defined in the Discovery Cohort were validated in the Validation cohort without further adjustments to the model.

3. *Validation of Model score*: Using the population-based glycemically defined new-onset diabetes cohort (n=1096) we scored all subjects as per the three model scoring systems (Table 1 and 2). We compared performance of the 3 models in the Validation Cohort based on 4 parameters, viz. sensitivity (proportion of pancreatic cancer identified), specificity (% of T2-NOD with scores below cutoff), enrichment (cancer prevalence in Model-defined cohort above cutoff) i.e. positive predictive value (PPV), and proportion of T2-NOD in low-risk category, i.e. negative predictive value (NPV).

4. *Risk-stratification by model in all new-onset diabetes subjects*: Distribution of scores was analyzed in all identified T2-NOD (n=1288) and PC-NOD (n=73) subjects to validate our criterion score that defined high-, intermediate- and low- risk categories.

5. *Causes of false positive score for high-risk category*: A random selection of T2-new-onset diabetes with false positive END-PAC scores (n=100) were manually reviewed to determine etiology of false positives and the proportion that might be excluded in a prospectively done study (e.g., concurrent non-pancreatic cancer). This was tested in the prospectively assembled EXPAND trial cohort.

6. *Sensitivity of score with increasing lead time:* All identified PC-NOD (n=73) were classified by lead time defined as interval between date of first meeting glycemic criteria for new-onset diabetes and cancer diagnosis (<6, 6-12, 12-18 and >18 months). The sensitivity of the chosen criterion score was determined.

Sensitivity analysis

Selection of BG in the preceding ~1 year: We compared the final END-PAC score in the population-based validation Cohort using the lowest, highest and mean BG value in the preceding ~1 year.

Use of similar paired BG values: We compare the final END-PAC score in the population-based validation Cohort using the subjects with paired FBG data only *vs.* paired EAG only *vs.* mixing FBG and EAG.

Statistical analysis

Statistical analyses were carried out using commercial software (JMP, version 10.0, SAS Institute Inc.). All the results are expressed as mean (standard deviation [SD]) or median (interquartile range [IQR]) as appropriate. The Pearson's χ^2 test was used to compare categorical variables. The two-tailed *t* test was used to compare continuous variables. Given the binary

outcome, the final predictors for the model were estimated using a logistic regression analyses. A p value of <.05 indicated statistical significance.

Results

Incidence of pancreatic cancer in new-onset diabetes: Between January 1st, 2000 and December 31st, 2015, there were 1561 Olmsted County residents ≥50 years of age who first met the glycemically-defined new-onset diabetes criteria, of whom 16 (1.0%) developed pancreatic cancer with 3 years of meeting criteria for new-onset diabetes.

Development of models: Significant differences between PC- and T2-NOD were identified in the Discovery Set (Table 3). Multivariate logistic regression showed weight loss, ΔBG category, age and ΔBG to have significant Likelihood Ratio in the descending order (Table 3). Proportionate to their likelihood ratios, Δweight scores increased by 2 points per category compared to ΔBG (1 point). From these variables we created and compared 3 models created (A, B and C described in Methods) in the Discovery and Validation cohorts. Model C performed the best and was labeled Enriching New-onset Diabetes in Pancreatic Cancer (END-PAC) model (Table 2). In the Discovery cohort, subjects with T2-NOD had a lower mean END-PAC score compared with PC-NOD (5 vs 0; p<.001). The area under the Receiver Operating Characteristic curve (ROC) for the END-PAC Model was 0.87; a cutoff score of ≥3 had a sensitivity of 80%, and specificity of 80%. This cutoff score was validated in an independent population-based glycemically-defined new-onset diabetes cohort that did not differ in profile from the Discovery cohort (supplementary table 1).

Validation of END-PAC Cohort: Of 1096 glycemically-defined new-onset diabetes subjects in the validation set, 9 pancreatic cancers were identified (0.82%). An END-PAC score of ≥3

identified 7 pancreatic cancers with a sensitivity of 78%, specificity of 82% and enriched the pancreatic cancer prevalence of 0.82% in the population-based cohort to 3.6% (4.4 fold) in END-PAC model-defined cohort, (predictiveness curve illustrated in Figure 1). A total of 370 subjects (33%) had an END-PAC score of 2 or 1, with 2 subjects developing pancreatic cancer within 3 years (0.54%). A total of 530 subjects (48%) had an END-PAC score of $\leq$0 with no pancreatic cancer was identified in this low risk group.

Risk stratification by END-PAC score in all new-onset diabetes subjects: When the distribution of scores were analyzed in all T2-NOD (n=1288) and PC-NOD (n=73), 56 PC-NOD (77%) had an END-PAC score of $\geq$3 compared to 248 T2-NOD (19%) (Figure 2A). Fifteen PC-NODs subjects (21%) had an END-PAC score of 1 or 2 compared to 408 T2-NOD subjects (32%). Two PC-NODs subjects (2%) had an END-PAC score $\leq$0 to 632 T2-NOD subjects (49%).

Causes of false positive scores in the high-risk category: Of 100 T2-NODs from the Discovery and Validation set who had high END-PAC scores ($\geq$3), 22 had an active malignancy, 10 had end-stage disease (e.g., heart, kidney, liver, advanced dementia), 8 had steroid-induced diabetes, and 2 were severely ill needing parenteral nutrition. Twelve patients had rapid gain in weight (>4 kg) in months preceding date of glycemically-defined new-onset diabetes followed by weight loss. In 45 T2-NOD patients (45%), no identifiable cause was found to explain high the END-PAC score.

The proportion of false positive fraction was validated in a prospectively identified glycemically-defined new-onset diabetes cohort of 74 subjects from the EXPAND trial which included 1 pancreatic cancer (1.3%). The false positive rate (T2-NOD with an END-PAC $\geq$3) was 13 (17%) while 42 subjects (57%) had an END-PAC score of $\leq$0.

Sensitivity of END-PAC score with increasing lead time: The END-PAC score had a higher sensitivity in subjects with lead time <6 months (83%) compared with lead time (months) of 6 to 12 (73%), 12 to 18 (70%) and >18 (71%) (Figure 2B).

Sensitivity analysis

Selection of BG in the preceding ~1 year: There was no difference in the mean END-PAC score of subjects after lowest, highest and mean BG value in proceeding ~1 year (0.6 *vs.* 0.5 *vs.* 0.5; p=.52)

Use of similar paired BG values: There was no difference in the mean END-PAC score of subjects with paired FBG only *vs.* paired EAG only *vs.* mixed FBG or EAG (0.8 *vs.* 0.4 *vs.* 0.5; p=.24).

Discussion

We have developed and validated a score which stratifies subjects over age 50 years into high-, intermediate- or low-risk groups for pancreatic cancer at the time they first meet glycemic criteria for diabetes. The 3-year incidence of pancreatic cancer in glycemically-defined new-onset diabetes was ~1% and increased to 3.6% in those with an END-PAC score of $\geq$3 (END-PAC Cohort), which would warrant clinical work-up (Figure 3). A negative END-PAC score ($\leq$0) has a very high negative predictive value for pancreatic cancer; subjects with these scores should be managed as T2-NOD. The remaining 25% of PC-NOD patients have an END-PAC Score of 1 or 2; the 3-year incidence of pancreatic cancer in this group is 0.5%; biomarkers may help enrich this cohort for pancreatic cancer.

Our model has a strong clinical rationale. It is based on the features of PC- and T2-NOD that others and we have consistently observed in various cohorts studied over past two decades[17, 25-28], viz., the paradoxical development of diabetes in the face of weight loss in PC-NOD [16, 17] and, weight gain in T2-NOD [15]. This paradox provides clues to the as-yet unknown pathogenesis of diabetes in pancreatic cancer. It is also the reason the model maintains its sensitivity even in those who develop PC-NOD >12 months before clinical diagnosis. The model also accounts for the fact that mean age at T2-NOD (~52 years) is lower than the mean age at PC-NOD diagnosis (~71 years).

However, if one simply summarized the model by the concept of "new-onset diabetes + weight loss = pancreatic cancer", one would seriously underutilize the power of the model to both enhance sensitivity and enrich the cohort for pancreatic cancer (Figure 4). This is because the concept fails to capture two characteristic differences between T2-NOD and PC-NOD with regard to their glycemic progression, viz. i) slow progression from normal fasting glucose to new-onset diabetes over ~8 years in T2-NOD[18, 19] *vs.* rapid worsening of prediabetes in PC-NOD over 2-3 years leading to ii) significantly higher blood sugars at PC- vs T2-NOD[17, 20]. Here too there is a nuance that enhances the performance of the model. A simple ΔBG is not as informative as ΔBG category used in the model (Table 3). For example, a ΔBG of 30 mg/dl over 1 year does not discriminate between an increase in FBG from 118 mg/dl to 148 mg/dl (common in T2-NOD) from an increase in FBG from 105 mg/dl to 135 mg/dl (frequently seen in PC-NOD, but not in T2-NOD). The model uses FBG and EAG, two quite different glycemic parameters, interchangeably. While EAG is a glycemic indicator of average glucose control derived from HbA1c, FBG provides an indication of glucose control at the time of the blood test. Our data show that their interchangeable use in the model did not affect the score (see Results). This likely

reflects the use of categories for BG rather than absolute values and stronger influence of weight loss on the score. The use of FBG, EAG or higher value of the two when both are available reduces the proportion of subjects with missing paired BG values to <15% enhancing model performance compared to use of paired FBG or EAG.

When more than one BG value was available 1 year prior to new-onset diabetes, we chose the highest value. One of the discriminating features between T2- and PC-NOD is the high BG value 1 year prior to new-onset diabetes diagnosis in T2-NOD, reflecting gradual progression to new-onset diabetes in T2 *vs.* rapid progression in pancreatic cancer (Table 2). Our choice of the highest BG takes this into account to exclude T2-NOD. However, when we tested the END-PAC model in our population-based validation cohort using the lowest, highest or mean BG we found no difference in the mean END-PAC score between the 3 strategies (see Results section). If future validation confirms this, one could potentially use any BG at -12 months.

The false positive fraction of T2-NOD increases the cost of finding pancreatic cancer in the END-PAC Cohort. To study this further we reviewed medical records of 100 T2-NOD subjects with a score of $\geq$3: 60% of false positives were due to profound weight loss and 40% due to rapid development of DM. We determined that, if studied prospectively, ~40% of T2-NOD subjects in the END-PAC Cohort would have been excluded from work-up of pancreatic cancer because they were already being investigated for another illness that explained their weight loss. Rapid development of DM in T2-NOD could be attributed to recent steroid use in 10% and rapid gain in weight in the months preceding new-onset diabetes development followed by weight loss due to uncontrolled DM in 15%. In 50% T2-NOD and PC-NOD were indistinguishable. Thus, it appears that false positive END-PAC scores can be reduced by 50% resulting in an increase in the positive predictive value of an END-PAC score of $\geq$3 to 10%.

At the other end of the spectrum are subjects with an END-PAC Score of ≤0. The clinical profile of these patients is typically that of a younger patient who has gained weight and gradually progressed to T2-NOD. Such subjects are typical of T2-NOD. They have negligible to no risk of pancreatic cancer and should be managed as T2-NOD; clinical work-up for pancreatic cancer is not warranted. About 50% of T2-NOD subjects and 25% of PC-NOD have scores in the intermediate range. These subjects have lost a modest amount of weight (<4 kg) while having gradual increase in their BG category. Their risk of pancreatic cancer is too low (1 in 250) to justify clinical work up. However, biomarkers such as CancerSeeK[29] or others in development may help enrich this cohort for pancreatic cancer (Figure 3).

Recently Pepe et al reported "predictiveness" as a better yardstick to compare risk prediction models than the traditional area under the receiver operating characteristic curve (AUC)[30]. We fully concur with this concept which is well illustrated in our model comparison (Figure 1 and 5). However, additional characteristics are important to consider when comparing models. Is the model's sensitivity adequate with increasing lead time? Most models will have high sensitivity close to diagnosis of cancer. But for early detection, the model with a high sensitivity despite a long (>6 months) lead time will be most beneficial for early detection. The END-PAC score has a ~75% sensitivity for subjects with 6, 12 and 18 months lead time.

In another paper Pepe et al[31] comment that a model or biomarker has clinical utility if

$$\frac{True\ positive\ fraction\ (sensitivity)}{False\ positive\ fraction\ (1-specificity)} > \left(\frac{1-\rho\ (prevalence\ of\ disease)}{p\ (prevalence\ of\ disease}\right) * \text{r (cost/benefit ratio)}$$

where r is reciprocal of the number false positives that are acceptable to be worked up to find one true positive. The END-PAC model (80% sensitivity, 85% specificity, disease prevalence 1.0% in new-onset diabetes cohort and r of 1/24) meets the bar for clinical utility noted above. In

prospective studies we expect the END-PAC model to perform better due to markedly reduced fraction of False Positive subjects. We observed this in the prospectively conducted EXPAND trial where the false positives proportion was 10%. This will lead to nearly doubling the pancreatic cancer prevalence in a model-defined high-risk cohort. If a validated biomarker of pre-symptomatic pancreatic cancer is identified, this would allow the intermediate risk group to be enriched and worked up.

How would one find pancreatic cancer in subjects in those with very high risk? In a recent study (under review), we showed that the volume of tumors whose mean fasting glucose was above 126 mg/l was 4.1 to 8.0 cc (median diameter of largest dimension 26 mm) compared to 12.0 cc (median diameter of largest dimension 34 mm) at clinical diagnosis. The current modalities available for screening include computerized tomography (CT), magnetic resonance imaging (MRI) and endoscopic ultrasound EUS. The average volume and largest diameter of tumors at new-onset diabetes appear to be large enough to be detectable by these modalities, especially endoscopic ultrasound.

What proportion of pancreatic cancers might be detected and how earlier by this strategy? We have observed that ~20% of pancreatic cancer subjects meet criteria for new-onset diabetes (unpublished data), ~70% of whom could potentially be diagnosed >3 months prior to clinical diagnosis if END-PAC score were to be applied at onset of diabetes. Our previous studies of CT scans review done prior to diagnosis show that there was no evidence of un-resectability >6 months prior to diagnosis[32]. This would suggest that earlier detection will lead to improved resectability and consequent better survival. However, the true benefit of this strategy can only be determined by a prospective study, which is currently underway (http://cpdpc.mdanderson.org/index.html).

In a cancer where 85% patients are diagnosed at an advanced stage, the need for earlier diagnosis is unquestionable. While the model is easy to use, we believe physician and patient education will be required to appropriately apply the model only in true new-onset diabetes subjects as it is unclear how it will perform in long-standing diabetes or patients with unknown duration of diabetes. The economic burden of earlier detection appears very reasonable considering that a recent study determined the cost of screening for pancreatic cancer in an unenriched population of new-onset diabetes (8-fold increased risk) to be $356 to $1569 per year life added[33]. However, the social impact of implementing an early detection strategy using a risk prediction model in clinical practice needs a prospective study.

The strength of our study is that it is based on well-known clinical features of T2-NOD and PC-NOD which were confirmed in carefully constructed cohorts from a population-based setting. All pancreatic cancer diagnoses were manually verified to exclude common mimics (such as IPMN, ampullary cancer, islet cell cancer) that otherwise constitute ~20% of unverified pancreatic cancer cohorts. The validation set was also population-based with well documented diabetes status and cancer outcomes for all subjects. Despite the inherent limitations related to retrospective nature of our study, the REP resources, which capture, all blood sugar measurements performed at different institutions, minimized the proportion of subjects with missing values to <15%. While this attests to its applicability in a prospective study in this population, it might be a possible source of missing values in single center studies. A limitation of the study is that the population of Olmsted County, Minnesota is predominantly Caucasian; its validation in a more diverse population is being planned.

In summary, an END-PAC score based on three easily accessible and highly discriminatory factors readily risk-stratifies subjects with glycemically-defined new-onset

diabetes, clearly defining those who warrant clinical work-up for pancreatic cancer and those who could managed as T2-NOD. The intermediate-risk group will have to await further research to enrich the population for pancreatic cancer. However, 75% of pancreatic cancer have an END-PAC score of $\geq$3. If validated, the END-PAC can be utilized in clinical care to identify a very high risk group for pancreatic cancer that warrants clinical work up.

References

1. Rahib L, Smith BD, Aizenberg R, et al. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer research 2014;74:2913-2921.
2. Chari ST, Leibson CL, Rabe KG, et al. Probability of pancreatic cancer following diabetes: a population-based study. Gastroenterology 2005;129:504-511.
3. SEER. Fast Facts: Pancreatic Cancer. Available at: http://seer.cancer.gov/statfacts/html/pancreas.html. Accessed on January 17, 2018.
4. Chari ST. Detecting early pancreatic cancer: Problems and prospects. Seminars in Oncology 2007;34:284-294.
5. Ogawa Y, Tanaka M, Inoue K, et al. A prospective pancreatographic study of the prevalence of pancreatic carcinoma in patients with diabetes mellitus. Cancer 2002;94:2344-9.
6. Damiano J, Bordier L, Le Berre JP, et al. Should pancreas imaging be recommended in patients over 50 years when diabetes is discovered because of acute symptoms? Diabetes Metab 2004;30:203-7.
7. Illes D, Terzin V, Holzinger G, et al. New-onset type 2 diabetes mellitus--A high-risk group suitable for the screening of pancreatic cancer? Pancreatology 2016;16:266-71.
8. Munigala S, Singh A, Gelrud A, et al. Predictors for Pancreatic Cancer Diagnosis Following New-Onset Diabetes Mellitus. Clin Transl Gastroenterol 2015;6:e118.
9. Boursi B, Finkelman B, Giantonio BJ, et al. A clinical prediction model to assess risk for pancreatic cancer among patients with new-onset diabetes. Gastroenterology 2017;152:840-850. e3.
10. Gupta S, Vittinghoff E, Bertenthal D, et al. New-onset diabetes and pancreatic cancer. Clinical Gastroenterology and Hepatology 2006;4:1366-1372.
11. Aggarwal G, Rabe KG, Petersen GM, et al. New-onset diabetes in pancreatic cancer: a study in the primary care setting. Pancreatology 2012;12:156-61.
12. Fraser LA, Twombly J, Zhu M, et al. Delay in diagnosis of diabetes is not the patient's fault. Diabetes Care 2010;33:e10.
13. Harris MI, Klein R, Welborn TA, et al. Onset of NIDDM occurs at least 4-7 yr before clinical diagnosis. Diabetes Care 1992;15:815-9.
14. Porta M, Curletto G, Cipullo D, et al. Estimating the delay between onset and diagnosis of type 2 diabetes from the time course of retinopathy prevalence. Diabetes Care 2014;37:1668-74.
15. Ford ES, Williamson DF, Liu S. Weight change and diabetes incidence: findings from a national cohort of US adults. Am J Epidemiol 1997;146:214-22.
16. Sah RP, Nagpal SJS, Mukhopadhyay D, et al. New insights into pancreatic cancer-induced paraneoplastic diabetes. Nature Reviews Gastroenterology and Hepatology 2013;10:423-433.
17. Hart PA, Kamada P, Rabe KG, et al. Weight loss precedes cancer-specific symptoms in pancreatic cancer-associated diabetes mellitus. Pancreas 2011;40:768-72.
18. DeJesus RS, Breitkopf CR, Rutten LJ, et al. Incidence Rate of Prediabetes Progression to Diabetes: Modeling an Optimum Target Group for Intervention. Popul Health Manag 2017;20:216-223.
19. Nichols GA, Hillier TA, Brown JB. Normal fasting plasma glucose and risk of type 2 diabetes diagnosis. American Journal of Medicine 2008;121:519-524.
20. Pannala R, Leirness JB, Bamlet WR, et al. Prevalence and clinical profile of pancreatic cancer–associated diabetes mellitus. Gastroenterology 2008;134:981-987.
21. CenterforDiseaseControl. https://gis.cdc.gov/grasp/diabetes/diabetesatlas.html accessed on January 17, 2018.
22. https://www.cancer.org/cancer/pancreatic-cancer/causes-risks-prevention/risk-factors.html. Accessed on January 17, 2018.

23. Melton LJ. History of the Rochester epidemiology project, In Mayo Clinic Proceedings, Elsevier, 1996.
24. Rocca WA, Yawn BP, Sauver JLS, et al. History of the Rochester Epidemiology Project: half a century of medical records linkage in a US population, In Mayo Clinic proceedings, Elsevier, 2012.
25. Chari ST, Klee GG, Miller LJ, et al. Islet amyloid polypeptide is not a satisfactory marker for detecting pancreatic cancer. Gastroenterology 2001;121:640-5.
26. Aggarwal G, Ramachandran V, Javeed N, et al. Adrenomedullin is up-regulated in patients with pancreatic cancer and causes insulin resistance in β cells and mice. Gastroenterology 2012;143:1510-1517. e1.
27. Pannala R, Leibson CL, Rabe KG, et al. Temporal association of changes in fasting blood glucose and body mass index with diagnosis of pancreatic cancer. The American journal of gastroenterology 2009;104:2318.
28. Chari ST, Leibson CL, Rabe KG, et al. Pancreatic cancer–associated diabetes mellitus: prevalence and temporal association with diagnosis of cancer. Gastroenterology 2008;134:95-101.
29. Cohen JD, Li L, Wang Y, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science 2018.
30. Pepe MS, Feng Z, Huang Y, et al. Integrating the predictiveness of a marker with its performance as a classifier. Am J Epidemiol 2008;167:362-8.
31. Pepe MS, Janes H, Li CI, et al. Early-Phase Studies of Biomarkers: What Target Sensitivity and Specificity Values Might Confer Clinical Utility? Clin Chem 2016;62:737-42.
32. Pelaez-Luna M, Takahashi N, Fletcher JG, et al. Resectability of presymptomatic pancreatic cancer and its relationship to onset of diabetes: a retrospective review of CT scans and fasting glucose values prior to diagnosis. Am J Gastroenterol 2007;102:2157-63.
33. Bruenderman E, Martin RC, 2nd. A cost analysis of a pancreatic cancer screening protocol in high-risk populations. Am J Surg 2015;210:409-16.

Figures

Figure 1. Predictiveness curve and sensitivity of END-PAC model

In Figure 1 at 80% on the x-axis [risk percentile] the pancreatic cancer risk value is 4.0% indicating that, after the END-PAC score, 80% of subjects in the cohort have a calculated risk below 4.0% and only 20% have risk at or above 4.0. The line intersecting at the 80^{th} percentile on the lower half of the graph gives the sensitivity (80%) and specificity (80%) of the END-PAC score at the corresponding risk percentage.

Abbreviations: NOD, new-onset diabetes; PC, pancreatic cancer;

Figure 2. A. Distribution of score in all patients; B. Sensitivity of END-PAC score in PC-NOD based on lead time

Abbreviations: PC-NOD, pancreatic cancer new-onset diabetes; T2-NOD, type 2 new-onset diabetes

Figure 3. Guidelines for clinical workup on new-onset diabetes for pancreatic cancer

Abbreviations: CT, computerized tomography; EUS, endoscopic ultrasound; FBG, fasting blood glucose, RBG, random blood glucose;

Figure 4. Comparison of predictiveness curves of new-onset diabetes + weight loss *vs.* END-PAC model

Abbreviations: NOD, new-onset diabetes; PC, pancreatic cancer

Tables

Table 1. Enriching New-onset Diabetes for Pancreatic Cancer (END-PDAC) score parameters

Blood Glucose (BG) Categories		**Δ BG Category Score (NOD-1y) (A)**
BG range (mg/dl)	Score	Score Range
BG category at -1 years		1-4
<100	1	
100-109	2	
110-125	3	
BG category at glycemically-defined new-onset diabetes		
126-160	4	
>160	5	
Δ Weight Categories		**Δ Weight score (B)**
Δ Weight (kg)	Score	Score Range
≤ -6.0	+6	-6 to +6
-5.9 to -4.0	+4	
-3.9 to -2.0	+2	
-1.9 to +1.9	0	
+2.0 to +3.9	-2	
+4.0 to +5.9	-4	
≥ +6.0	-6	
Age (years) at glycemically-defined new-onset diabetes		**Age score (C)**
Age range	Score	Score Range
≤59	-1	-1 to +1
60 to 69	0	
≥70	+1	
Total Score		**A + B + C**

Abbreviations: BG, blood glucose; NOD, new-onset diabetes

Table 2. Comparison of performance characteristics of different classifier models for pancreatic cancer in new-onset diabetes

Model	Sensitivity	Specificity	Pancreatic cancer prevalence	Sensitivity with >12-month lead time	Proportion in low risk group	AUC
Glycemically-defined New-onset Diabetes	…	…	0.9%	…	…	…
A: + weight loss ≥2.5 kg	44%	84%	1.9%	47%	0	.75
B: + ΔBG^ + Δ weight loss*	78%	75%	2.1%	71%	35%	.86
C: + ΔBG categories* + Δ weight loss*	78%	80%	4.5%	71%	48%	.87

*Please refer to Table 1 for model A, Δ weight loss categories and Δ BG categories and respective assigned score

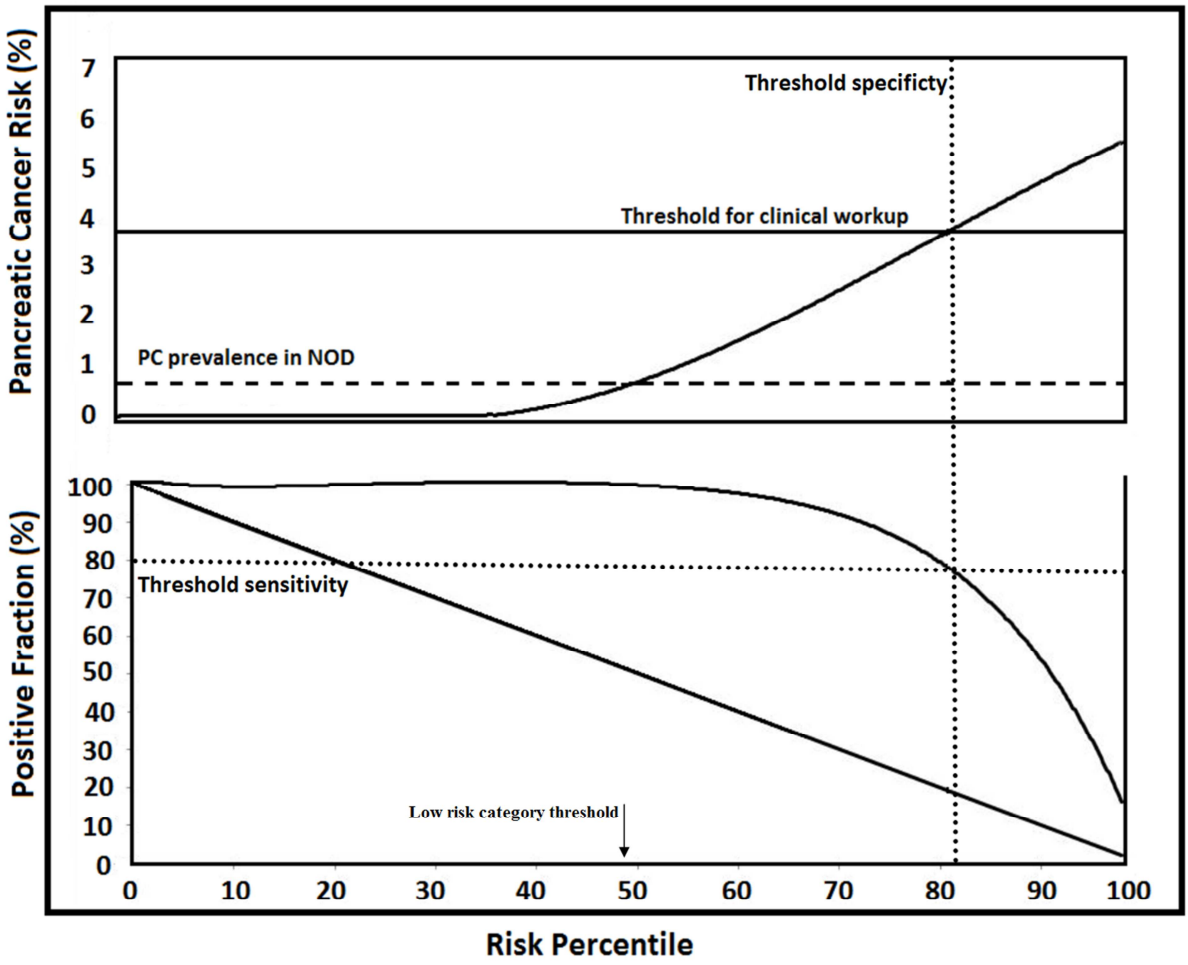
^Δ FBG was calculated subtracting the blood glucose at -1 years from blood glucose at New-onset diabetes. For every 10 mg/dl difference, 1 point was assigned with highest point being 10 for anyone with a Δ FBG of ≥100 mg/dl.

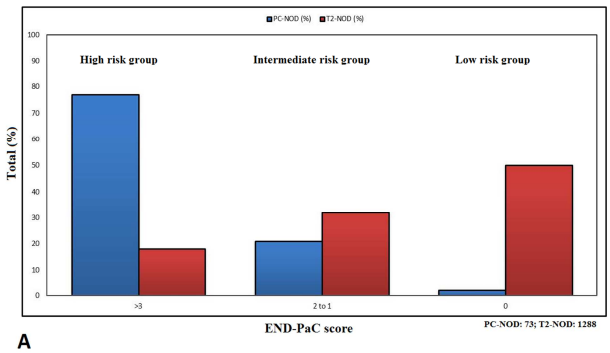
Abbreviations: AUC, area under curve; BG, blood glucose;

Table 3. Univariate model for predictors of pancreatic cancer in new-onset diabetes

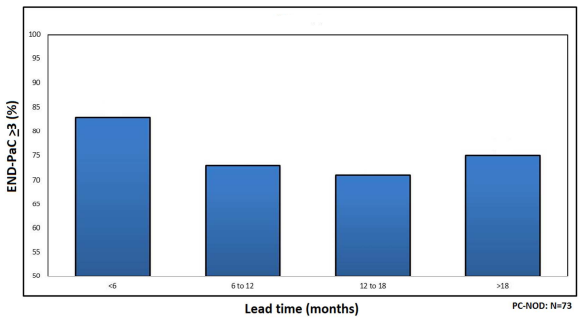
Characteristics	**PC-NOD**	**T2-NOD**	**P-value**
Total, N	64	192	
Age, years, mean (±SD)	72.4 (±9.3)	65.6 (±10.1)	**<.001**
Gender, Female (%)	36 (56%)	99 (52%)	.51
Blood Glucose (BG) (mean ± SD) (mg/dl)			
at New-onset diabetes	162 (±69)	143 (±25)	**.002**
at 1y	110 (±12)	118 (±11)	**<.001**
ΔBG	53 (±69)	26 (±27)	**<.001**
Body weight (mean ± SD) (kg)			
at New-onset diabetes	82.5 (±17.4)	95.3 (±22.9)	**<.001**
at 1y	87.2 (±19.3)	94.2 (±21.9)	**.02**
ΔWeight	-4.7 (±5.9)	1.1 (±4.4)	**<.001**
Blood Glucose category at NOD (%)			
>160 (Category 5)	18 (28%)	17 (9%)	
125-160 (Category 4)	46 (72%)	175 (91%)	
Blood Glucose category at -1y (%)			
≥126 (Category 4)	3 (5%)	45 (23%)	**<.001**
110-125 (Category 3)	17 (45%)	106 (55%)	
101-109 (Category 2)	20 (31%)	32 (17%)	
<100 (Category 1)	12 (19%)	9 (5%)	
Δ Blood Glucose category	2.0 (±0.9)	1.0 (±0.8)	**<.001**
Multivariable regression analysis			
Parameter	Likelihood Ratio		
ΔWeight	45.5		**<.001**
Δ Blood Glucose category	24.0		**<.001**
Age at New-onset diabetes	8.4		**.003**
Δ Blood Glucose	1.1		.28

Abbreviations: BG, blood glucose; CI, confidence interval; NOD, new-onset diabetes; PC-NOD, pancreatic cancer new-onset diabetes; SD, standard deviation; T2-NOD, type 2 new-onset diabetes

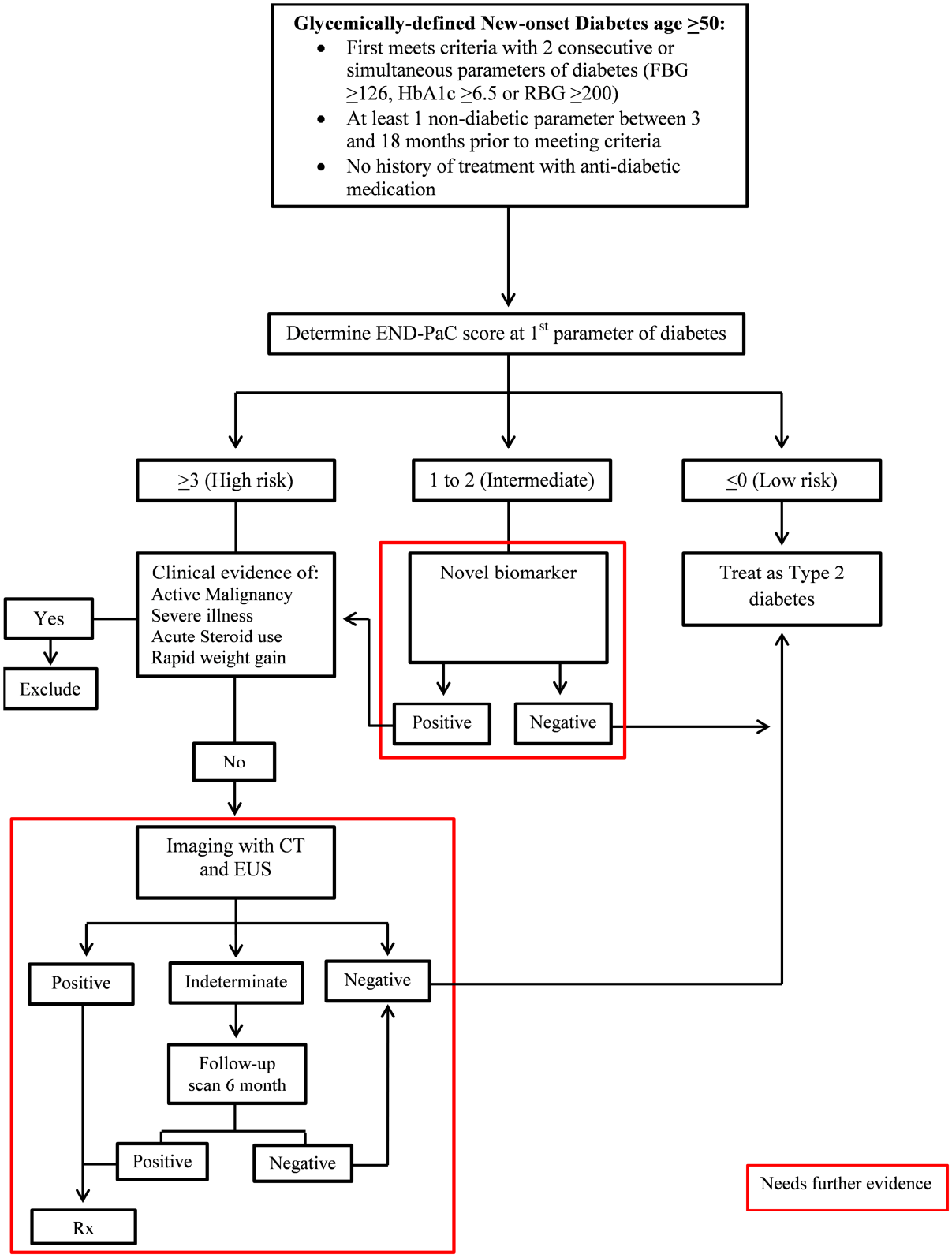

Pancreatic Cancer Risk (%)
7
6
5
4
3
2
1
0
Threshold specificty
Threshold for clinical workup
PC prevalence in NOD
Positive Fraction (%)
100
90
80
70
60
50
40
30
20
10
0
Threshold sensitivity
Low risk category threshold
0
10
20
30
40
50
60
70
80
90
100
Risk Percentile

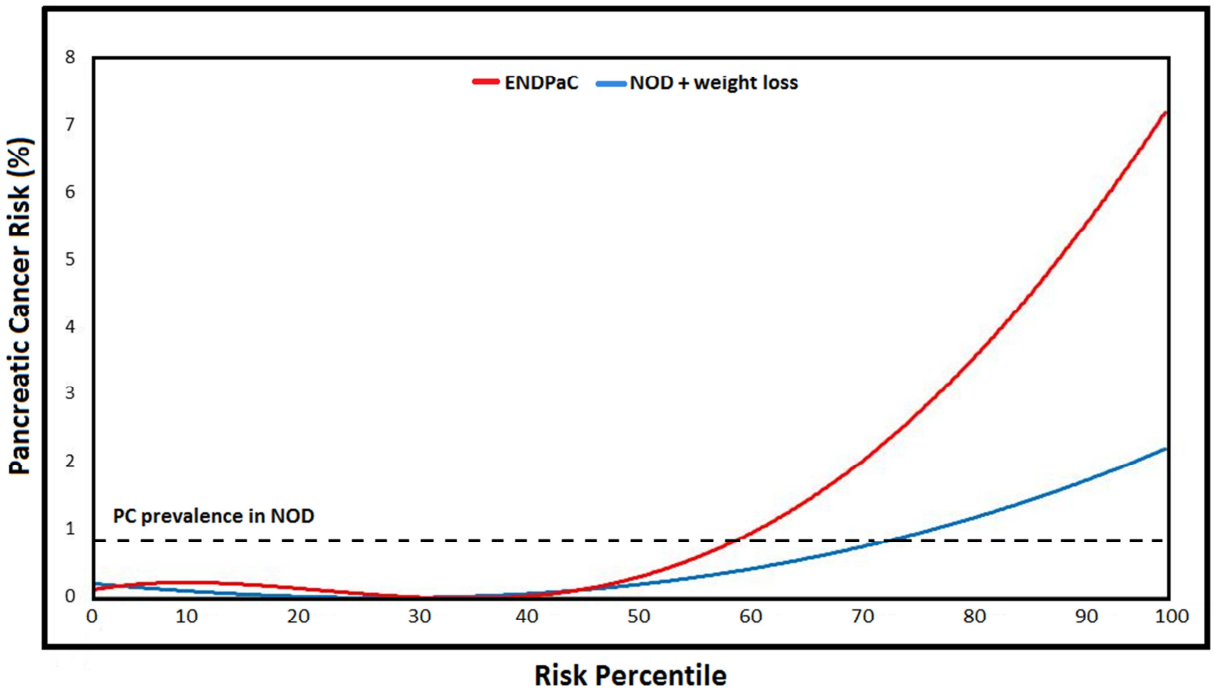


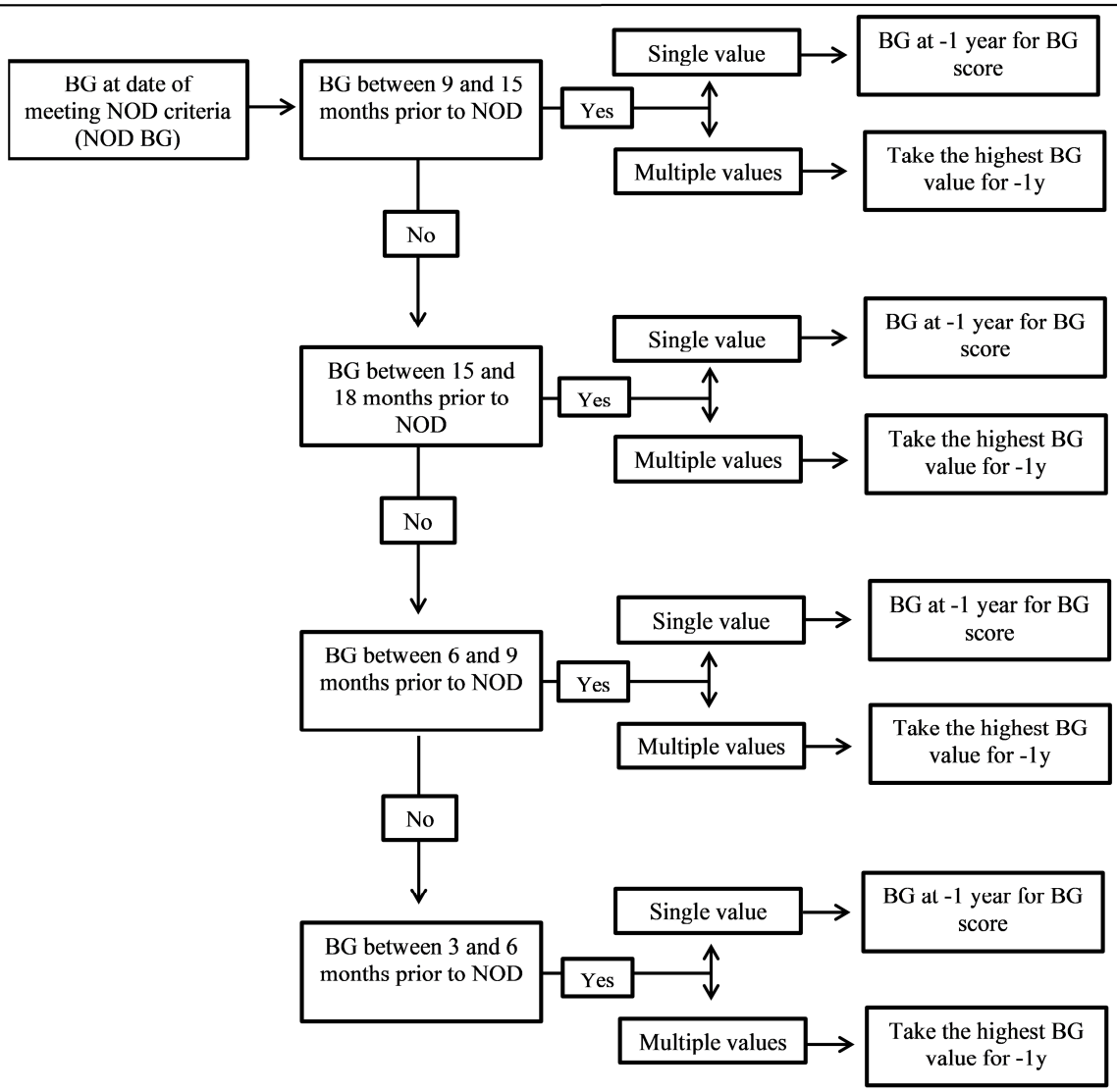
A



B



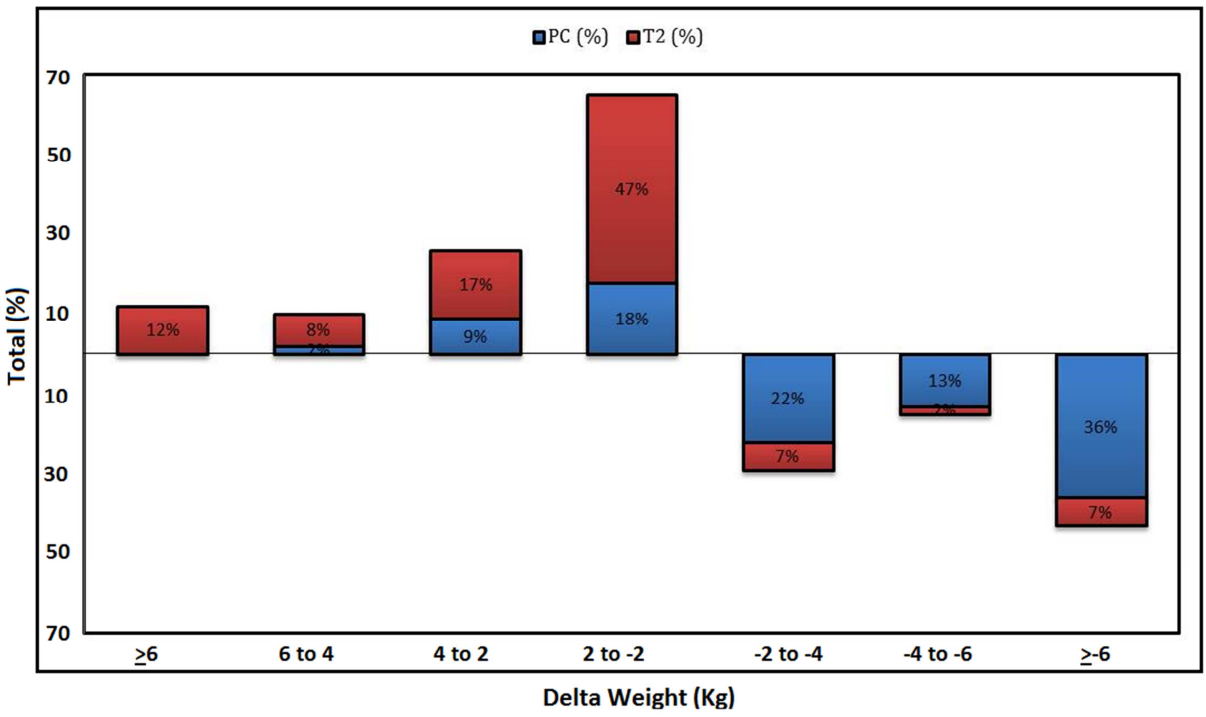




BG: blood glucose includes fasting blood glucose (mg/dl) or estimated average glucose (mg/dl)

To calculate estimated average glucose (mg/dl), use formula: 28.7*HbA1c-46.7

(ref: David M. Nathan, Judith Kuenen, Rikke Borg, Hui Zheng, David Schoenfeld, and Robert J. Heine, for the A1c-Derived Average Glucose (ADAG) Study Group. Diabetes Care 2008)


PC (%)
T2 (%)
70
50
30
10
10
30
50
70
Total (%)
12%
8%
17%
9%
47%
18%
22%
7%
13%
36%
7%
≥6
6 to 4
4 to 2
2 to -2
-2 to -4
-4 to -6
≥-6
Delta Weight (Kg)

Criteria for diabetes: Subjects had diabetes if they met one of these:

i) Defining glycemic criteria: two of the following values (excluding those drawn as inpatient or in ER) measured simultaneously (same day) or consecutively: fasting blood glucose (≥126 mg/dl), random blood glucose (≥200mg/dl), glycosylated hemoglobin (HbA1c) (≥6.5%), 2h Post oral glucose load (PG) ≥200 mg/dL (11.1 mmol/L) during OGTT

or

ii) Therapeutic criteria: were started on anti-diabetes medication after at least one of the above parameters being met within 90 days.

Definition of new-onset diabetes: Patients who met above criteria for diabetes who also met all the following criteria were considered new-onset:

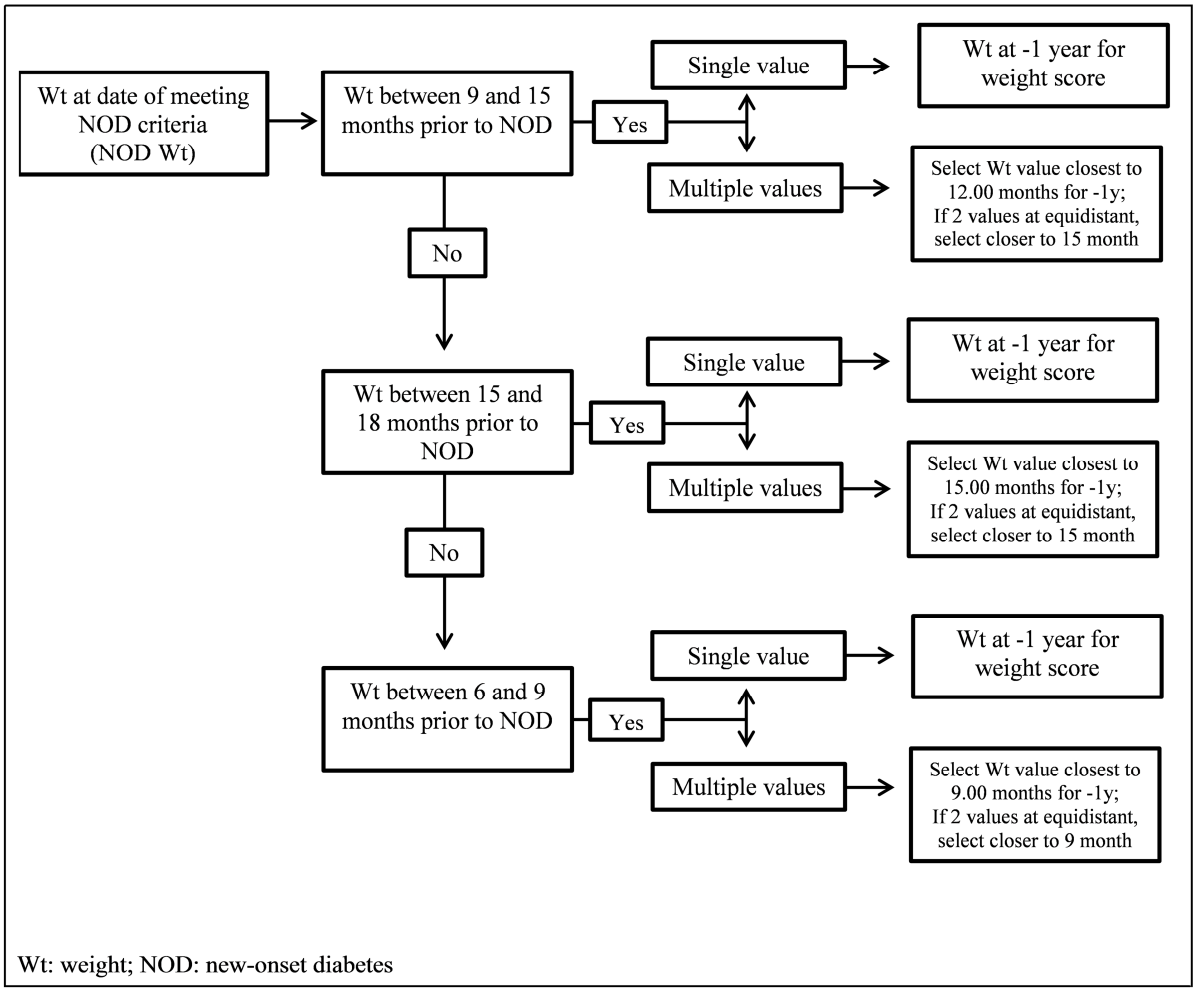
1. Interval between the two defining glycemic criteria is <90 days.
2. One or more of fasting glucose or HbA1c values measured in the past 18 months
3. Did not previously meet glycemic criteria for diabetes (as defined above)
4. Were not previously treated with anti-diabetes medications

Date of onset of diabetes: Date of first abnormal glycemic parameter was considered as date of onset of diabetes.

Supplementary Material

Supplementary Table 1. Comparison of characteristics in subjects with type 2 new-onset diabetes (T2-NOD) in validation and discovery cohort

Characteristics	**Validation**	**Discovery**	**P-value**
Total, N	1087	192	
Age, years, mean (±SD)	65.6 (±9.6)	65.6 (±10.1)	.94
Gender, Female (%)	540 (50%)	99 (52%)	.76
Blood glucose (mean ± SD) (mg/dl)			
At NOD	147 (±32)	143 (±25)	.11
At 1y	116 (±11)	118 (±11)	**.05**
Body weight (mean ± SD) (Kg)			
At NOD	96.7 (±23.5)	95.3 (±22.9)	.43
At 1y	95.4 (±22.5)	94.2 (±21.9)	.51
ΔWeight	1.3 (±4.7)	1.1 (±4.4)	.44
Blood Glucose categories at NOD (%)			
125-160	928 (86%)	175 (91%)	.09
>160	157 (14%)	17 (9%)	
Blood Glucose categories at -1y (%)			
<100	84 (7%)	9 (5%)	.21
101-109	193 (18%)	32 (17%)	
110-125	609 (56%)	106 (55%)	
>125	199 (18%)	45 (23%)	
Mean END-PDAC score (±SD)	0.4 (±3.3)	0.4 (±3.3)	.89



Wt: weight; NOD: new-onset diabetes