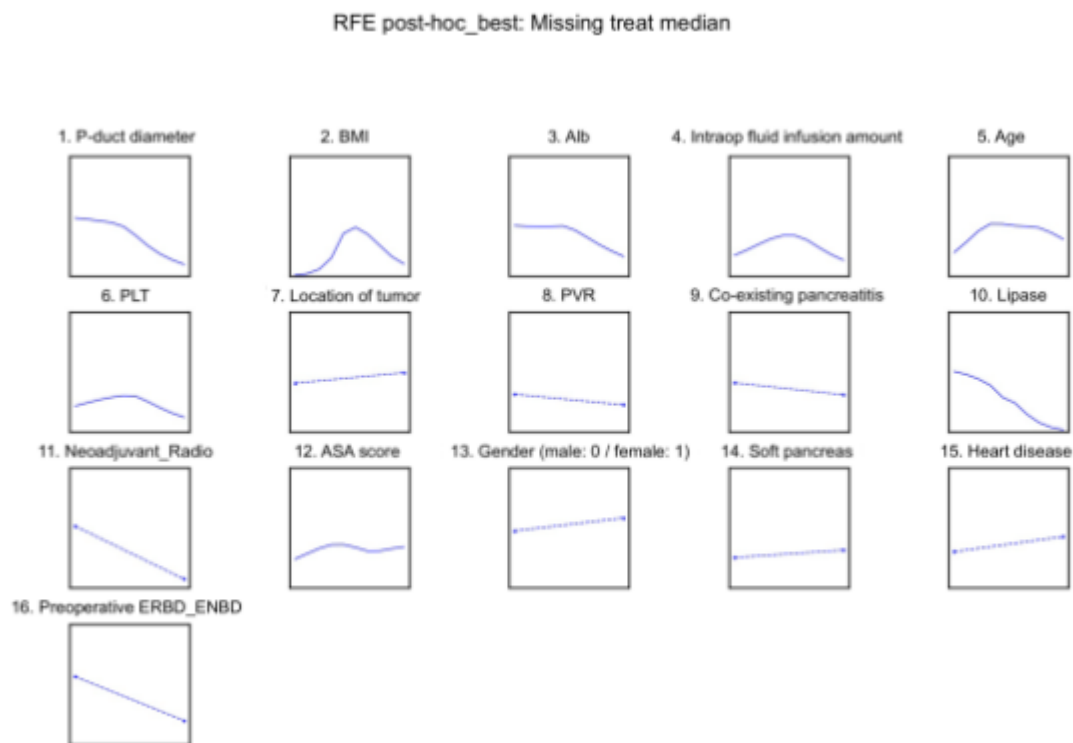Response letter for reviewer's comments

Reviewer #1:

1. The main limitation is the lack of clinical utility. Although a web-based calculator was created, a tool with 16 variables is very cumbersome to use. Especially when its predictive ability is not significantly better (AUC 0.74) than current models. I'm sure many would be willing to reduce the AUC a bit in order to have fewer factors (see figure 1).

Response:

We all authors in this study absolutely agree on reviewer's concern. In this study, a total of 38 variables were inserted into AI-driven algorithms. Of course, the inclusion of many variables could be the reason for lowering the AUC value for this study. Also, it may be inconvenient to calculate the risk of POPF by applying 16 risk factors in actual clinical care. However, the basic hypothesis of our study is to identify how many variables interact with each other and affect POPF, and we believe that these interactions would not be linear like previous models which could oversimplify complex relationships among many risk factors. An example of nonlinearity is shown in supplement Figure 1, and note that each variable has a variable effect. As a result, we believe that our AI-driven risk platform better incorporates multiple risk factors and can account for more nuanced relationships between the risk factors and POPF (main text Figure 2). Additionally, to improve this cumbersome procedure to use 16 variables, we are currently proceeding on interconnecting the hospital's EMR system with the web-based POPF application. The above is reflected in the text (highlighted) as followed: The post hoc analysis revealed a nonlinear relationship by showing the response of NN model to each input variable at every RFE step. Ten discrete points cover the observed range of variation for each corresponding variable. We found several patterns of NN output response, in which the predicted POPF risk seemed to have a positive, negative, or biphasic relationship with each variable. The contribution profiles of the top 16 variables are shown in supplement Figure 1.

Supplement Figure 1. Post-hoc analysis of best NN model with RFE and the contribution profile of the risk factors.

Top 16 variables from the best NN model with RFE and the contribution profile of each simulated variable to the prediction of POPF risk. For example, the profile of the p-duct diameter (top left corner) shows a negative relationship, i.e., if the diameter is longer, the risk is lower.

RFE post-hoc_best: Missing treat median

NN, neural network; RFE, recursive feature elimination; POPF, postoperative pancreatic fistula; BMI, body mass index, PLT, platelet; PVR, portal vein resection; ASA, American Society of Anesthesiologists; ERBD, endoscopic retrograde biliary drainage; ENBD, endoscopic nasobiliary drainage

2. The other limitation is the lack of external validation. Also, the methods stated that the cohort was split into a training and validating set, but the results are presented as just one large cohort analysis.

Response:

We understand reviewer's point very deeply. As mentioned in the Limitation section, this study had the disadvantage of not performing external validation. As a result, the follow-up study will be conducted by performing an external validation on patients in multiple institutions.

In our study, the number of samples of Table 2 appears to be 1769 - the sample number of the entire cohort - is expressed as such because it is the result of the 5-fold cross validation. The above 5-fold cross validation process was described in the Data analysis and statistical methods section as followed (highlighted):

This randomly divides all the data into 5 partitions (folds) keeping each one with similar positive and negative data distribution. Then, we train a model with four of the partitions and test the model with the remaining fold. By changing the folds for training and testing, this process is performed 5 times. Also, the whole cross-validation was repeated 10 times by random split of the dataset, evaluating the performance of the models at the end. These processes ensure the generalized performance of a model by preventing overfitting to the samples.

3. The factors identified are largely known risk factors for POPF. Interestingly, the authors bring up sarcopenia but don't provide data on that factor.

Response:

Thank you for reviewer's comment for sarcopenia. Recently, we just got our own data for preoperative sarcopenia as a risk factor for POPF (PMID: 32354655 DOI: 10.1016/j.hpb.2020.04.004). Unfortunately, in the design phase of our study at that time, it was before we got this data about sarcopenia and did not reflect it. Therefore, we will sooner or later conduct modeling for only preoperative variables including preoperative sarcopenia with multicenter data.

4. The authors also bring up mitigation factors which have largely been proven unsuccessful for POPF prevention.

Response:
We absolutely agree on reviewer's opinion. In this study, we described several mitigation strategies, such as, pancreaticogastrostomy reconstruction, dunking/invaginating anastomosis, absorbable mesh patches, and the use of intraperitoneal drains, anastomotic stents, and prophylactic somatostatin analogues. We acknowledge those methods somewhat questionable. However, the main purpose of this study is not to find efficacy of mitigation strategy but to discover the risk factor and build prediction platform for POPF through artificial intelligence. Further research needs to be done to prove efficacy of various mitigation strategy. As part of those efforts, we have carrying out the ongoing clinical trial of this risk score wherein we are applying a somatostatin analogue during postoperative days 0–3 in high-risk patients.

5. Since duration of drain placement influences POPF risk, what was the authors' drain management practice?

Response:
The duration of drain placement of our institution is determined on clinical condition and CT scan on postoperative 5 days. Without the clinical and radiological evidence of POPF, the drain had been removed immediately on postoperative 5 days. If there were evidence of POPF and the amount of drainage were large, the drain tube would be maintained for a certain period of time.

6. Caution stating first study to…..See recent paper: Surgery. 2020 Feb;167(2):448-454

Response:
We absolutely agree on reviewer's point. Before we began to write this manuscript, we conducted a thorough literature review as much as possible. But nevertheless, we

made the mistake of omitting this recently published study from Switzerland. We have quoted and included Switzerland study in the reference list (ref. no 5). However, one thing to note is that this Switzerland study did not focus on multiple risk factors and relationships between the risk factors and POPF, but focused on machine learning-based pancreatic parenchyma texture analysis as a risk factor for POPF. As a result, we believe that this study is the first to predict POPF <mark>with multiple risk factors</mark> using AI. . The above is reflected in the text (highlighted).