LETTER OF INTENT
DETERMINATION LETTER

DDT BMQ000105
September 16, 2020

PathAI
Attention: Katy Wack
120 Brookline Ave.
Boston, MA 02215

Dear Dr. Wack:

We are issuing this letter to PathAI to notify you of our determination on the project submitted to the Center for Drug Evaluation and Research (CDER) and the Center for Biologics Evaluation and Research (CBER) Biomarker Qualification Program (BQP). We have completed our review of the Letter of Intent (LOI) deemed reviewable on June 8, 2020 and have determined to accept it into the CDER and CBER BQP[1].

Your next submission, a Qualification Plan (QP), contains details of the analytical validation plan for the biomarker measurement method, detailed summaries of existing data that will support the biomarker and its context of use (COU), and includes descriptions of knowledge gaps and how you propose they will be mitigated. If future clinical studies are planned, please include detailed study protocols and the statistical analysis plan for each study as part of your QP submission.

Below, we provide you with specific considerations and recommendations to help improve your preparation for, and submission of the QP. As this biomarker development effort is refined, the submitted data, the specifics of your context of use (including the target patient population), and the design of study(ies) used in the clinical validation of the biomarker will ultimately determine which of these considerations and recommendations are most applicable. For more information about your next submission and a QP content element outline, please see the BQP Resources for Biomarker Requestors web page.[2]

[1] In December 2016, the 21st Century Cures Act added section 507 to the Food, Drug, Cosmetic Act (FD&C Act). FDA is now operating its drug development tools (DDT) programs under section 507 of the FD&C Act.

[2] https://www.fda.gov/drugs/cder-biomarker-qualification-program/resources-biomarker-requestors

CONSIDERATIONS & RECOMMENDATIONS

1. Drug Development Need Considerations

Currently, there are no approved therapies for non-alcoholic steatohepatitis (NASH). Pathology plays a critical role in NASH clinical trials with histology being the current reference method to determine inclusion in trials and change in disease activity and fibrosis stage. Manual histological review is complex, subjective, and prone to inter- and intra-reader variability and error. Existing pathology scoring systems and practices show only moderate to fair reproducibility, limiting their utility for clinical research and practice. Thus, there is a critical need for tools that would improve upon histological interpretation to achieve greater consistency and standardization.

2. Biomarker Description and Interpretation Considerations

The biomarker described by this submission is a histological biomarker consisting of the features comprising the Non-Alcoholic Fatty Liver Disease Activity Score (NAS) (the grade of steatosis, hepatocellular ballooning, inflammation) and fibrosis staging (via NASH/CRN Brunt/Kleiner scale), assessed on liver biopsy as interpreted by Artificial Intelligence (AI). See Tables 1 and 2 below:

Table 1: NAS scoring, summarized from Kleiner, 2005

Non-Alcoholic Fatty Liver Disease Activity Score (NAS)		
Item	**Definition**	**Score**
Steatosis	< 5%	0
	5%-33%	1
	> 33%-66%	2
	> 66%	3
Lobular inflammation	No foci	0
	< 2 foci per 200 x field	1
	2-4 foci per 200 x field	2
	> 4 foci per 200 x field	3
Ballooning	None	0
	Few balloon cells	1
	Many cells / prominent	2

Table 2: NASH/CRN Brunt/Kleiner Stages

Fibrosis Stage	
0	None
1A	Mild, zone 3 perisinusoidal
1B	Moderate, zone 3 perisinusoidal
1C	Periportal sinusoidal fibrosis without accompanying zone 3 fibrosis
2	Zone 3 perisinusoidal and portal/periportal
3	Bridging fibrosis
4	Cirrhosis

2.1 The inputs, algorithm description, algorithm training, and algorithm output and interpretation need to be specified as components contributing to the biomarker interpretation. See Section 4. Analytical Considerations, below, for more details.

2.2 Please plan to provide clear definitions for each of the features that comprise the biomarker to support your comparator pathology assessments. Of the three features that comprise the NAS, the degree of steatosis appears to be the most reproducible. However, agreement on the reading of all the components of the NAS (steatosis, lobular inflammation and ballooning degeneration) are suboptimal. The variability in reproducibility in pathology assessment may, in part, stem from imprecise definitions. For example, steatosis can be measured by either percent hepatocytes with a steatotic droplet or by percent non-fibrotic surface area with fat. The former method has been adopted by the NASH-CRN; however, this may overestimate degree of fat within the liver by counting hepatocytes with small droplet macrovesicular steatosis as being equivalent to hepatocytes with large droplets of fat that entirely fill the cell. Ballooning degeneration was not specifically defined in the NAS nor was there guidance for scoring severity of ballooning beyond the descriptors “few” or “many”. The most accepted definition of ballooning degeneration is a hepatocyte that is generally larger than the surrounding hepatocytes with a distinctive rarified cytoplasm that is irregularly stranded or clumped. Given the importance of ballooning degeneration in the diagnosis of NASH as well as its association with fibrosis, a clear definition on how to assess this feature is needed. A focus of lobular inflammation has also not been defined in the NAS. For example, SAF system defines a focus as two or more inflammatory cells (neutrophils, lymphocytes/other mononuclear cells, eosinophils, and microgranulomas) present within the sinusoids or surrounding injured ballooned or apoptotic hepatocytes. These variabilities may contribute to the variability in pathologists. You should plan to provide evidence that the system can correctly score each histological feature across a range of conditions and disease activity states, in addition to the score overall.

3.COU Considerations

Requestor's COU Statement:

A surrogate endpoint biomarker, based on Artificial Intelligence (AI), to measure treatment response based on histological change in Non-Alcoholic Fatty Liver Disease Activity Score (NAS) components (i.e., steatosis, ballooning, inflammation) and fibrosis scores in liver biopsies from baseline to follow-up in patients in clinical trials for treatment of non-alcoholic steatohepatitis (NASH).

FDA COU Recommendation:

The FDA recommends that you pursue a stepwise approach to qualification of biomarkers in this area and begin with a COU[3] with a lower burden of evidence than a surrogate endpoint, such as:

A diagnostic biomarker to assess disease activity score components (i.e., steatosis, ballooning, inflammation) and fibrosis stage in liver biopsies as part of evaluation for enrollment in non-alcoholic steatohepatitis (NASH) clinical trials.

3.1 Clarify if you intend your machine learning algorithm to be the only reading utilized to determine entry into a clinical trial or (eventually) for final determination of efficacy or if these reading will be confirmed by a clinical pathologist. If the biomarker is qualified for this COU, will patients enrolled in NASH clinical trials have a differential risk based on biomarker interpretation (e.g., will patients be triaged to different treatment arms based on biomarker status)?

3.2 Since biomarker qualification is not a regulatory endorsement of clinical use outside of clinical trials, please ensure that you are meeting the requirements of device use and regulation for any use of the Path AI technology for clinical decision making.

4. Analytical Considerations

4.1. Please address the following biomarker interpretation considerations in your QP:

- 4.1.a. Please describe the input requirements to your AI algorithm. This may include, but not limited to, the compatible image acquisition hardware and protocols, image quality, size and format of images.
- 4.1.b. Please provide a description of your algorithm in a step-by-step fashion. This may include, but not limited to, data/image preprocessing, AI architectures,

3 This recommended COU (i.e., diagnostic biomarker vs. surrogate endpoint) forms the basis for the analytical and statistical considerations that follow.

and postprocessing steps. Please provide details regarding the system operation (e.g., how your system selects regions of interest, excludes background regions, handles imaging artifacts). You should plan to assess the precision of the regions of interest selected by your system.

4.1.c. Please describe the method of training the primary algorithm and the method of tuning hyperparameters. We suggest you document and clearly describe the use of patient data in all the stages of algorithm development, training, verification, and validation. Please provide collection protocols and population characteristics of these patient data. In general, we do not recommend the use of cross-validation (including random shuffling, splitting, and then hold-out) for the final clinical validation of your algorithm as it may introduce bias and affect cross-domain generalization.

4.1.d. Include in your Qualification Plan the inter- and intra- reader concordance for the pathologists used to annotate and train the algorithm and for the validation of the Model.

4.1.e. Please describe the algorithm output information: for scores please make clear the scale (categorical or continuous); if any other information is provided by the algorithm to end users besides scores, etc. Please make clear the intended user of the algorithm.

4.2. You indicate that “The NASH algorithm consists of a result sub-system which will provide scores for inflammation, steatosis, ballooning as well as overall scores, based on the CRN- derived NAS scoring system (Figure 1), and can indicate both CRN-based and Ishak fibrosis scores as continuous values on the slide level CRN based fibrosis scoring is done separately from the NAS score and is different from Ishak score.” Inclusion of Ishak score is inconsistent with the context of use and inconsistent with the pathological assessments in current clinical practice and will change the NASH-CRN scoring; therefore it is not appropriate. We recommend you remove these features from the score and restrict the system to CRN- derived NAS scoring system.

4.3. The inclusion criteria of the trial described in the LOI was NASH with bridging fibrosis (F3, STELLAR-3) or compensated cirrhosis (F4, STELLAR-4). If the algorithm training was performed at these subsets, then it only captures a part of NASH patients and therefore it is not clear that the biomarker and assessment method is broadly representative of the population for NASH CRN scoring. Clarify if all screened patients were included in the data set and what proportions of patients did and did not have liver fibrosis, fibrosis stages and NAS scores $< $ and ≥ 4. You should ensure that the algorithm is validated for the indicated population and have a plan to demonstrate the biomarker performance in the presence of other prognostic variables and comorbidities (e.g., diabetes, obesity).

4.4. You indicated that you validated your algorithm using 161 cases reserved from the Stellar 4 clinical trial. Specimens from this same trial were used to train the algorithm. You propose on page 10 of the LOI to continue to train and test with separate subsets of cases from a large population taken from the NASH CRN database. Please be

advised it is inappropriate to use specimens from the same trial because it can lead to overfitting and false discovery. You should plan to provide a detailed description of how specimens will be selected for your validation to avoid bias and you should also plan to describe how data integrity is maintained during validation such that testing is blinded to outcome.

4.5. You indicate that additional optimization of the model, based on annotations collected from experienced liver pathologists will be performed to find tune the model using a pre-defined subset of WSIs. Please be advised, you should have a finalized "locked down" algorithm prior to validation.

4.6. Please clarify whether your study will use any control slides or slide images and if so, provide a description of these materials. Please further describe all quality assurance and quality control (QA/QC) measures intended for use during operation of the system during a clinical trial. Provide details of engineering controls, administrative controls, or other measures to ensure reliable evaluation of the biomarker. Describe detection mechanisms to identify faults and any corrective and preventive actions to be taken in the event of a deviation from expected operation.

5. Statistical Considerations

In your qualification plan, you should submit the statistical analysis plan (SAP) detailing: the study objective for clinical validation of the final model; the primary statistical hypothesis; the datasets to be used for clinical validation; the statistical analysis method and, if there are missing data, how will they be handled in the model validation.

The SAP should include what is the intent-to-diagnose population, the anticipated prevalence of the intent-to-diagnose population, and the specific diagnostic performance measures. Based on the LOI document, we have the following statistical comments for your consideration in preparation of your SAP for the QP submission.

5.1. The SAP should include a testable hypothesis regarding the expected diagnostic yield with your AI-algorithm in comparison with the standard pathology scoring by component and overall. Provide the details of the AI-algorithm and of the standard pathology scoring. Include the primary statistical analysis method. Pre-specify a clinically meaningful performance goal/OPC (Objective Performance Criteria) and what constitutes success of accurate diagnosis based on the AI-algorithm for adequate assessment of the diagnostic performance. The OPC should reflect the stated study goal.

5.2. The sample size planning in the SAP should be based on OPC and success criteria with justification and include details regarding the number of samples each expert pathologist will score and the number of experts who will score each sample.

5.3. The composition and operation of the “panel of expert board certified, liver pathologists” should be identified in the qualification plan. You indicated (page 11) that the performance of the locked down model will be compared to a consensus score using a panel of expert pathologists. Please provide more details regarding how the consensus score will be obtained including how discordance between pathologists will be resolved.

5.4. Consider how to summarize intra-observer variation and inter-observer variation for the clinical validation samples. We also recommend providing the percent agreement between the same and different experts.

5.5. The qualification plan should have reference rules for how comparative testing is conducted (e.g., what outcomes will be considered as “real”).

5.6. You propose to conduct repeatability and reproducibility studies to include a variety of scanners and tissue stained by local laboratories. Please be advised you should plan to provide robust evidence of specific scanners and inputs to provide the specifications of the input required for analysis. You should pre-specify the acceptance criteria for individual scanner performance in order to pool the data.

5.7. Provide the model performance measures to be employed for identifying the final AI algorithm. Also, provide details on how the AI-algorithm output from the optimized (finalized) model will be used.

5.8. You conducted a preliminary study with STELLAR-3 and STELLAR-4 datasets. The validation dataset should be independent of the training dataset. Therefore, you should utilize other studies for model validation. In the QP, please list the studies to be used for clinical validation assessment of the final Path-AI model. It appears that the iterative verification and validation steps you describe in Figure 3 for training the AI algorithm development framework may be using the same verification or validation datasets repeatedly. If the verification or validation data are reused, there is a risk that the AI algorithm may overfitted to these specific data. In addition, key study design elements (study population, primary objective, etc.) of each study should also be provided.

5.9 Please include a figure similar to Figure 5 in the LOI for the validation study. You should provide a rationale why you chose Spearman’s correlation and not another correlation measure. Please also add the estimated correlation with interval estimate between the AI-algorithm and central pathologist in the validation study.

Please address each of the specific considerations and recommendations and any data requests cross-referencing the numbered list above in a separate addendum to your QP submission. Please note that we do not intend to provide exhaustive comments on your LOI as some of your descriptions are high-level and lack sufficient details. We will provide full feedback when we review your QP submission.

When evaluating biomarkers prospectively in clinical trials, requesters are encouraged to submit study data using Clinical Data Interchange Consortium (CDISC) standards to facilitate review and utilization of data. Data sharing and the capability to integrate data across trials can enhance biomarker development and utilization. If sponsors plan to use the biomarker prior to qualification to support regulatory review for a specific Investigational New Drug (IND), New Drug Application (NDA) or Abbreviated New Drug Application (ANDA) development program, they should prospectively discuss the approach with the appropriate CDER or CBER division.

The BQP encourages collaboration and consolidation of resources to aid biomarker qualification efforts. Any individuals or groups (academia, industry, government) that would like to join in this effort, have information or data that may be useful can contact Dr. Wack (katy.wack@pathai.com).

Should you have any questions or if you would like a teleconference to clarify the content of this letter, please contact the CDER Biomarker Qualification Program via email at CDER-BiomarkerQualificationProgram@fda.hhs.gov with reference to DDT BMQ#000105 in the subject line. For additional information and guidance on the BQP please see the program's web pages at the link below.[4]

Sincerely,

Christopher L. Leptak -S

Digitally signed by Christopher L. Leptak -S
DN: c=US, o=U.S. Government, ou=HHS, ou=FDA, ou=People, 0.9.2342.19200300.100.1.1=1300421152, cn=Christopher L. Leptak -S
Date: 2020.09.16 09:32:51 -04'00'

Christopher Leptak, M.D., Ph.D.
Director, CDER Biomarker Qualification Program
Division of Biomedical Informatics, Research and Biomarker Development
Office of Drug Evaluation Science/Office of New Drugs
Center for Drug Evaluation and Research

Joseph G. Toerner -S

Digitally signed by Joseph G. Toerner -S
DN: c=US, o=U.S. Government, ou=HHS, ou=FDA, ou=People, 0.9.2342.19200300.100.1.1=1300136263, cn=Joseph G. Toerner -S
Date: 2020.09.16 15:12:46 -04'00'

Joseph Toerner, M.D.
Acting Division Director
Division of Hepatology and Nutrition
Office of Inflammation and Immunity/Office of New Drugs
Center for Drug Evaluation and Research

[4] https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/cder-biomarker-qualification-program

U.S. Food & Drug Administration
10903 New Hampshire Avenue
Silver Spring, MD 20993
www.fda.gov