MDIC
MEDICAL DEVICE INNOVATION CONSORTIUM

THE ALLIANCE FOR DIGITAL PATHOLOGY

MGH – MDIC – DPA - FDA

Medical Device Innovation Consortium

The Alliance for Digital Pathology

Digital Pathology & Artificial Intelligence

Monday, November 4th, 2019

8:00am – 5:30pm Eastern

Key Bridge Marriott

1401 Lee Highway

Arlington, VA 22209

This meeting will help prioritize the areas of digital pathology and AI, where MDIC can collaborate with its partners on new projects.

About this Event

MDIC, as a part of a broader alliance in digital pathology including FDA, Massachusetts General Hospital (MGH), and Digital Pathology Association (DPA), is working to prioritize the areas of Digital Pathology and Artificial Intelligence (AI) where MDIC can bring industry, users, government, insurance companies, and patients together to collaborate on several new projects.

The Alliance aims to work on:

A temporary framework to synergize and tackle larger scale projects

Harmonizing & standardizing a reference set to be used in end to end workflow by:
(1) Creating tools and datasets,
(2) Progressing and enabling market access,
(3) Creating clarity on regulatory pathways via mock submissions, and
(4) Harmonizing efforts between various stakeholders to optimize interoperability, integration and implementation.

This meeting will further the recent discussions MDIC has been having on the areas of digital pathology and AI.

We hope you will enjoy the meeting.

THE ALLIANCE FOR DIGITAL PATHOLOGY

MGH – MDIC – DPA - FDA

THE ALLIANCE IS GRATEFUL FOR BEING HOSTED BY MDIC.

MDIC, as a part of a broader alliance in digital pathology including FDA and Digital Pathology Association (DPA), is working to prioritize the areas of Digital Pathology and Artificial Intelligence (AI) where MDIC can bring industry, users, government, insurance companies, and patients together to collaborate on several new projects.

For those of you who are not familiar with the Alliance…

The Alliance is a regulatory science initiative to harmonize and standardize digital pathology processes to speed up innovation to patients.

The Alliance for Digital Pathology is a collaborative and voluntary group interested in the evolution of regulatory science as it applies to digital pathology. We seek participation from all stakeholders (industry, vendors, academic medical centers, patient advocates, regulatory bodies, associations etc) to come together and identify key elements necessary to move the field of digital pathology forward.

The purpose of the Alliance is to accomplish *concrete practical deliverables* and *relevant strategic aims* in order to sustain and expand the existing collaborative infrastructure.

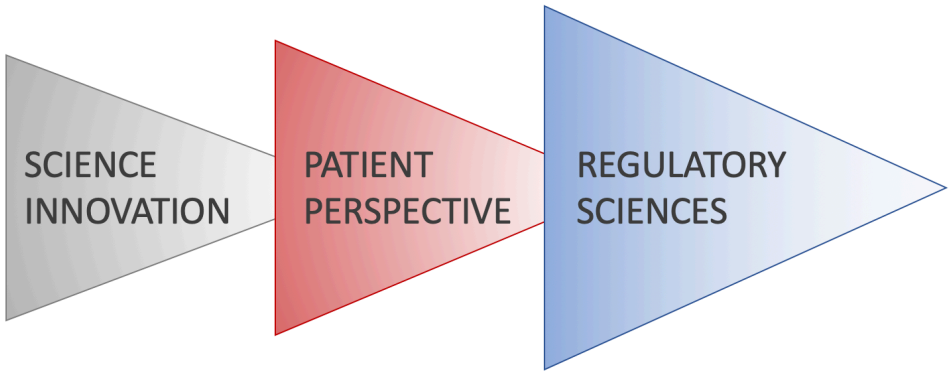
You can find more information here: https://digitalpathologyalliance.org/

Or contact us via: DigiPathAlliance@gmail.com

THE ALLIANCE FOR DIGITAL PATHOLOGY

MGH – MDIC – DPA - FDA

A regulatory science initiative to harmonize and standardize digital pathology processes to speed up innovation to patients.



The Alliance for Digital Pathology is a collaborative and voluntary group interested in the evolution of regulatory science as it applies to digital pathology. We seek participation from all stakeholders(industry, vendors, academic medical centers, patient advocates, regulatory bodies, associations etc.) to come together and identify key elements necessary to move the field of digital pathology forward.
The purpose of the Alliance is to accomplish *concrete practical deliverables* and *relevant strategic aims* in order to sustain and expand the existing collaborative infrastructure.

AIMS AND SCOPE OF THE ALLIANCE

- Tackle large-scale projects in pre-competitive space
- Develop evaluation tools, methods, and standards
- Clarify and improve regulatory pathways

AGENDA – 1

8.00-8.30 **Breakfast, Networking, & Arrival**

8.30-9.00 **MDIC Introduction and Welcome**
Pamela Goldberg
President & CEO, MDIC

9.00-9.10 **Introductions**
Joe Lennerz, MD, PhD
Medical Director, Center for Integrated Diagnostics
Associate Chief, Department of Pathology, Massachusetts General Hospital
Associate Professor, Harvard Medical School
Esther Abels, MS
VP Regulatory Affairs, Clinical Affairs, and Strategic Business Development
PathAI
Brandon Gallas, PhD
Senior Mathematician
Division of Imaging, Diagnostics, & Software Reliability, CDRH, OSEL, FDA

9.10-9.30 **Alliance Progress Update**
Joe Lennerz, MD, PhD
Esther Abels, MS

9.30-9.45 **Digital Pathology Association (DPA)**
Esther Abels, MS
Scott Blakely
Business Development Manager, Whole Slide Imaging & Digital Pathology
Hamamatsu Corporation

9.45-10.00 **Break**

AGENDA – 2

10.00-10.15 **Food and Drug Administration (FDA)**
Sara Brenner, MD, MPH
Associate Director for Medical Affairs,
Chief Medical Officer In Vitro Diagnostics, CDRH, FDA
Brandon Gallas, PhD

10.15-10.30 **Patient Centered Outcomes Research Institute (PCORI)**
Bill Lawrence, MD, MS
Senior Clinical Advisor
Office of the Chief Engagement and Dissemination Office
PCORI

10.30-10.45 **Friends of Cancer Research (FOCR)**
Laura Lasiter, PhD
Science Policy Analyst
FOCR

10.45-11.00 **American College of Radiology (ACR)**
Bibb Allen, MD, FACR
President
ACR

11.00-11.15 **Healthcare Infrastructure**
Joe Lennerz

11.15-11.50 **Overview of Projects** for Breakout Sessions

11.50-12.00 **Explain Breakout Session Structure**

12.00-1.00 **Lunch**
(self-assignment for breakout session)

AGENDA – 3

1.00-1.45	**Breakout Session 1** Participants work with group leader on one of 6 separate topics Key points and deliverables are captured on established outline
1.45-2.15	**Breakout Reports as brief presentations to Full Meeting** Each table/topic gets 1-2 min plus 2-3 min questions
2.15-2.30	**Short Break**
2.30-3.15	**Breakout Session 2** Participants switch topic (same format as Breakout session 1) Key points and deliverables are captured on established outline
3.15-3.45	**Breakout Reports as brief presentations to Full Meeting** Each table/topic gets 1-2 min plus 2-3 min questions
3.45-4.15	**Coffee Break, Networking**
4.15-5.00	**Closing Discussion, Feedback on the Meeting, Next Steps**
5.00-5.30	**Final Remarks and Adjourn**

PROJECT PROPOSAL – INDEX

Projects are assigned for discussion in 6 breakout topics

Pre-analytics – **Amanda** Lowe

Recognize the importance of pre-analytical factors clashing with the lack of standardization; develop a guideline and tools to enable generalizability of AI/ML applications and minimize the variability of human factors in this workstream

HistoQC: a QC tool for DP-slides – Andrew Janowczyk & Anant Madabhushi
Develop an open-sourced, high-throughput, quality control pipeline that can help precisely quantify characteristics and reporting of whole slide images

Pre-Analytical Tools – Amanda Lowe
Identify the deliverable tools and methods that will decrease possible error sources around the generation of tissue images and data to create a harmonized approach based on the need/requirement of the medical device seeking regulatory clearance

Slide Scanning – **Scott** Blakely

Deliverables for analytical studies will save time and expense for device manufacturers and users; standardization of slide scanning goes beyond technical performance assessment.

Scanning Automation – Zoltan Laszik, Melike Pekmezci, Cathryn Cadwell, Christopher Bowman, Grace Kim
An automated pipeline to scan high risk slides at adjusted sensitivity to improve subpar scanning fidelity to enable timely rescans and mitigate medicolegal liability

Truthing – **Sarah** Dudgeon, **Brandon** Gallas

Establish and disseminate a framework for creating truthing datasets with a demonstrated use case to unify human elements; of ground truthing

TILdataMDDT – Sarah Dudgeon, Richard Huang, Matthew Hanna, Brandon Gallas
Produce an FDA qualified MDDT dataset that medical device sponsors can use to develop and evaluate medical devices

Competitions And Data – Rajesh Dash, Ricardo Pietrobon, Brandon Gallas
A scalable Digital Pathology Competition preparation methodology to deliver challenges that deliver reproducible algorithms; the project also includes a set of evidence-based guidelines to document metrics and statistical methods

*Please see following pages for details

ML/Models/Use cases – **Matthew** Hanna

ML/AI and practically relevant use cases have the potential to unlock the full potential of digital pathology; change protocols -as currently described- will hardly capture all ramifications of adaptive algorithms in clinical practice.

Imaging Informatics Fellows @ FDA – Brandon Gallas
Create a training opportunity for a regulatory and research scientist in the field of medical imaging, informatics, and statistics. The project includes training development and training.

Evaluation Platform – Rajendra Singh, Matthew Hanna
An open, web-based framework for creating large-scale public and private data sets and evaluation tools including annotation tasks.

Surveillance Tools – Daniel Rubin,
Software tool that implements an approach to post-marketing surveillance; the software enables surveillance of commercially available AI algorithms.

Standards **Markus** Herrmann, **Mike** Isaacs

Generation of a database of standards as well as the importance of datasets as a resource and tools for algorithm development; the proposed resources could function as an incubator for method development

Remote signout – Joseph Sirintrapun & Joe Lennerz
identification of the specific prohibitive sections in CLIA88; produce a consensus whitepaper that can ultimately serve as one element of a future CLIA88 amendment.

De-Identify WSI – Vijay Narayanasamy
A fully automated de-identification software for WSI images of any file format that could help with creating WSI datasets for research, education and developing AI-based image analysis tools

MediPaTeD -Terminology Dictionary – Kingsley Ebare & Esther Abels
Create a framework for standardization of datau sed in the design, execution, analysis, regulatory submission and archival of pathology research studies using CAP cancer protocols

Payor Strategy – **Joe** Lennerz, **Esther** Abels

Include payors into the Alliance discussions and approach clinical guideline developers to ultimately understand how to integrate digital pathology in a financially sustainable and meaning manner.

Payor-Patient-Perspectives – Joe Lennerz & Esther Abels
Capture the required evidence to clearly delineate deliverables towards financial sustainability via appropriate coverage determination by payors.

The Alliance

Breakout session worksheet

Key elements, next steps, timeline

Pros for Patient, Clinical, R&D, and regulatory

Deliverable and timeline

Concerns for patients, clinical, R&D, and regulatory

Implications and efforts

BREAKOUT SESSIONS

Pre-analytics **Amanda** Lowe	**ML/Models/Use cases** **Matthew** Hanna
Slide Scanning **Scott** Blakely	**Standards** **Markus** Herrmann **Mike** Isaacs
Truthing **Sarah** Dudgeon **Brandon** Gallas	**Payor Strategy** **Joe** Lennerz **Esther** Abels

Breakout Session 1 – Pre-analytics

GUIDELINES + SURVEY

A

Pre-analytics
Amanda Lowe

DEFINITIONS
Education campaign

B

Slide scanning
Scott Blakely

GUIDELINES + SURVEY

C

Truthing data sets
Sarah Dudgeon **Hetal** Marble

VER&VAL + IOP GUIDELINES

D

Continuous learning
Esther Abels

STANDARD DATABASE

E

Data standardization
Markus Herrmann **Mike** Isaacs

TOOLKIT WITH ROI CALC

F

Practical use cases
Matthew Hanna

WHITE PAPER AI CATEGORIES

G

ML/AI – Model creation
Ashish Sharma

BEST PRACTICES guide+covrg

H

Payor strategies
Joe Lennerz

1

Key Elements, Next Steps, Timeline

1. Survey (Role/Responsibilities) + 3 months
   - a. Pathologists (control slide to imaging)
   - b. Histologists/Lab director (control slide for tissue prep)
2. Pre-analytical Prioritization and/or key requirements are identified (Tissue, Imaging, Ground Truth) + 5 months
3. Results feed to Whitepaper/Publications/Guidelines + 9 months

2

Pros for Patient, Clinical, R&D, and regulatory

1. Guides manufacturers on quality control requirements
2. Interpretative accuracy improved by controlling variability via standards
3. Improved pathologists' concordance
4. Enables the objective assessment of slides across different laboratories
5. Harmonization of efforts across other Alliance projects

3

Concerns for patients, clinical, R&D, and regulatory

1. Risk to interpretive accuracy if poor data is used
2. Poor ground truth data/imaging sets for innovation/technology development
3. Wrong treatment provided to patient if decision was based on pre-analytical mistakes
4. Garbage in-Garbage Out

4

Implications & Efforts

1. Education of impact of pre-analytical variation
2. Guidelines to promote MDDT submissions for for pre-analytical standards
3. Quality Control recommendations
4. Whitepaper
5. Guideline for control slide creation

5

Deliverable of the project and timeline

Pre-Analytical Guideline(s)
Timeline – 12 mos

6

Pre-Analytics: Summary

Need: There is a need to standardize human factors to create comparable samples from lab-to-lab for use in algorithmic/ML applications.

Problem: Currently, human factors for pre-analytic variables (e.g. control slides, staining techniques, fixing/mounting, scoring, etc.) are slightly different from lab-to-lab.

Project Focus: Create a set of standardized guidelines and tools that offer protocols, instructions, definitions, and examples to enable generalizability of AI/ML applications.

7

Title: HistoQC: a quality control tool for digital pathology slides

Author(s): Andrew Janowczyk, Anant Madabhushi

Addressed Parties: Slide scanner vendors, pathology departments, and algorithm developers

Background: Recent studies have shown cross site variability in the preparation and scanning of whole slide images (WSI) has a significant impact on the performance of deep learning approaches. Even when employing 12,000 WSIs for training [1], a 3% drop in AUC was experienced when the classifier was evaluated on images produced via a scanner different than that of the training set. It stands to reason then, an approach for the quality control and quantitative characterization of WSIs is required for improved hardware validation and classifier evaluation, robustness, and confidence.

Approach: HistoQC is an open-sourced, high-throughput, quality control pipeline which can help precisely quantify characteristics of WSIs, to aid in artifact and outlier detection [2,3]. Additionally, these metrics provide a quantitative specification of the WSIs used for training a machine learning classifier, and thus may help dictate which WSIs are most appropriate for evaluation using the classifier. For example, if a machine classifier is trained using slides with modest overstaining of hematoxylin, it may not be appropriate to employ that classifier on slides which are under-stained of hematoxylin without a rigorous evaluation.

Objective: 1) Further develop HistoQC via the implementation of additional pertinent metrics and artifact detectors. 2) enable the more facile roll-out of scanners and computational algorithms via the precise quantification of training and testing datasets (see below), easing validation of both (a) upstream technologies and (b) downstream algorithms employing WSI.

Deliverable(s): An extendible reference quality control pipeline to enable reproducible, quantitative reporting of slide presentation and artifact detection

Value proposition:

Clinical: employing HistoQC in line with clinical digital pathology workflows will enable (a) slides containing artifacts to be detected early in the pipeline, reducing delays and costs, and (b) lab managers can identify in real-time if slide production is exiting established parameter ranges (e.g., contamination of stains)

Regulatory: the validation of machine learning diagnostic and prognostic algorithms requires extensive evaluation of their robustness in the context of preanalytic sources of variance in slide preparation and digitization. HistoQC allows for both the measurement and identification of potential variabilities helping to ensure suitable validation sets are curated. Additionally, HistoQC can be employed to measure color constancy and image quality across various scanners. Precisely measuring both inter and intra scanner variability of the same slide may allow for a more rigorous specification of hardware requirements for WSI manufactures/vendors.

Research & Development: beyond helping to define appropriate ranges for the employment of machine classifiers, HistoQC can aid in the identification of batch effects which may unintentionally seriously weaken experimental conclusions [2, supplemental material].

Funding sources: ITCR-1U01CA239055 (link)

Benefit to patients (≤2) : technical advance, increased quality

References/Resources (optional):

1. Gabriele Campanella, Matthew G. Hanna, Luke Geneslaw, Allen Miraflor, Vitor Werneck Krauss Silva, Klaus J. Busam, Edi Brogi, Victor E. Reuter, David S. Klimstra & Thomas J. Fuchs, “Clinical-grade computational pathology using weakly supervised deep learning on whole slide images”, Nature Medicine Volume 25, pages1301–1309 (2019)
2. Janowczyk A., Zuo R., Gilmore H., Feldman M., Madabhushi A., "HistoQC: An Open-Source Quality Control Tool for Digital Pathology Slides", JCO Clinical Cancer Informatics, 2019
3. https://github.com/choosehappy/HistoQC

Title: Pre-Analytical Considerations in Digital Pathology and Recommendation for Guidelines to Improve Practices for the Generation of Tissue-based Images and Data

Author(s): Amanda Lowe, Visiopharm

Addressed Parties: Pathologists, Anatomic Pathology Laboratory personnel, United States Food and Drug Administration (FDA), College of American Pathologists (CAP), National Society of Histotechnicians (NSH) and relevant industry associations (DPA, API, STP, etc.) Patient Interest Groups, Medical Device and Innovation Consortium (MDIC)

Background: Human and animal tissue samples support decisions in diagnostics, research, drug development, business, and the regulatory approvals of drugs, medical devices and software as a medical Device (SaMD). It is essential to understand the error sources of generating tissue samples and the resulting images and data, the impact of error sources and how to mitigate them.

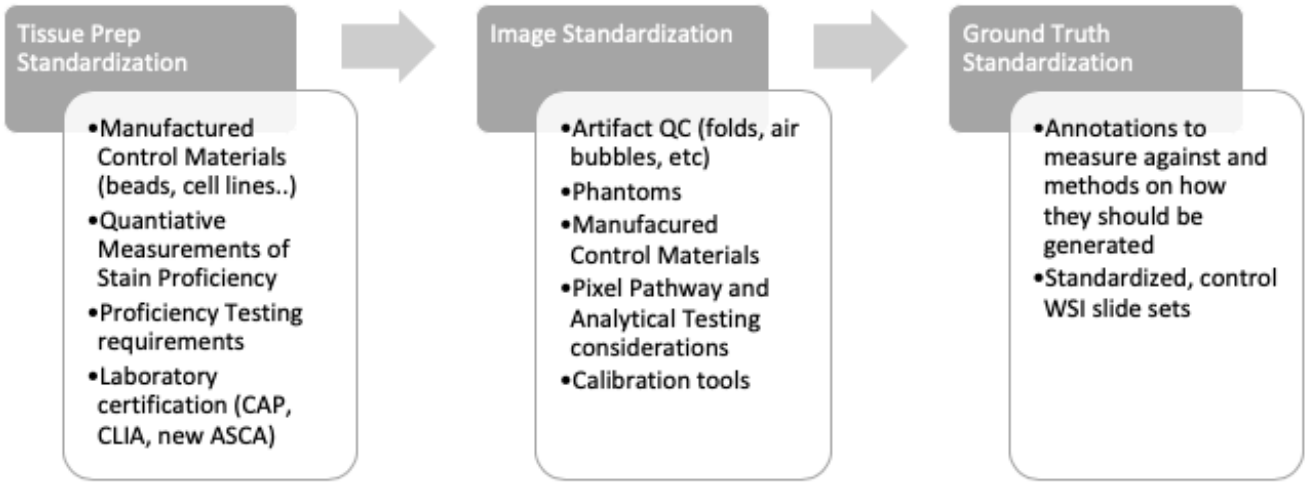
Although broadly relevant, the goal of this specific project is to generate guidelines for key recommendations about pre-analytical standardization, including imaging, that will promote pathways for regulatory clearance based on high quality images of tissue samples.

The possible error sources around the generation of tissue images and data include tissue processing artifacts, fixation, stain quality, inter and intra reader variability, imaging characteristics and more.

Approach & Objectives:

The first objective is to create a survey that will be sent to pathologists and laboratory professionals, with the goal of obtaining quantitative relevant data (ideally supported by CAP, NSH, DPA for widespread distribution), that will document key concerns, challenges, priorities for the generation of tissue-based images and data.

There are three main areas that should be discussed for creation of the survey and guidelines for pre-analytical standardization. Ideally tools from the various areas could be combined to create a harmonized approach based on the need/requirement of the medical device seeking regulatory clearance.



The second deliverable will be to identify a committee to build the guidelines with stakeholder with and various expertise. The final deliverable will be to build the guidelines in collaboration with the support of one or more key associations (CAP, NSH, DPA, STP, etc)

The guidelines will lower the barrier for medical device innovation, improve tissue data quality for diagnostics and research, and improve the quality of submission of higher quality data for submissions to the FDA.

Value proposition:

Clinical – Standardization and removal of potential error sources in the tissue preparation and imaging process will improve that interpretative accuracy of the pathology diagnosis.

Regulatory – The deliverables should lower the barrier for medical device innovation, improve tissue data quality, while reducing the regulatory burden with high quality, relevant submissions to the FDA.

R&D – Tissue and imaging standardization is not only relevant to clinical diagnostics but should have widespread impact into research and development (academic, biopharma, etc.) that all work with animal or human tissue. These deliverables can improve all aspects of tissue-based data.

Funding sources: None

Benefit to patients (≤2) : ☒technical advance, ☒increased quality, ☒outcome, ☐access, ☐affordability

References/Resources (optional):

1. https://www.visiopharm.com/images/publications/Posters/20170501_Qualitopix_-_Automatic_Quality_Assessment_of_the_Estrogen_Receptor_Poster_Final.pdf
2. Medical Device Development Tools (MDDT): https://www.fda.gov/medical-devices/science-and-research-medical-devices/medical-device-development-tools-mddt
3. Gabriele Campanella, Matthew G. Hanna, Luke Geneslaw, Allen Miraflor, Vitor Werneck Krauss Silva, Klaus J. Busam, Edi Brogi, Victor E. Reuter, David S. Klimstra & Thomas J. Fuchs, "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images", Nature Medicine Volume 25, pages1301–1309 (2019)
4. Janowczyk A., Zuo R., Gilmore H., Feldman M., Madabhushi A., "HistoQC: An Open-Source Quality Control Tool for Digital Pathology Slides", JCO Clinical Cancer Informatics, 2019
5. Accreditation Scheme Conformity Assessment (ASCA): https://www.fda.gov/regulatory-information/search-fda-guidance-documents/accreditation-scheme-conformity-assessment-asca-pilot-program

The Alliance

Breakout session worksheet

Key elements, next steps, timeline

Pros for Patient, Clinical, R&D, and regulatory

Deliverable and timeline

Concerns for patients, clinical, R&D, and regulatory

Implications and efforts

Breakout Session 2 – Slide scanning

GUIDELINES + SURVEY A Pre-analytics **Amanda** Lowe	DEFINITIONS Education campaign B Slide scanning **Scott** Blakely	GUIDELINES + SURVEY C Truthing data sets **Sarah** Dudgeon **Hetal** Marble	VER&VAL + IOP GUIDELINES D Continuous learning **Esther** Abels
STANDARD DATABASE E Data standardization **Markus** Herrmann **Mike** Isaacs	TOOLKIT WITH ROI CALC F Practical use cases **Matthew** Hanna	WHITE PAPER AI CATEGORIES G ML/AI – Model creation **Ashish** Sharma	BEST PRACTICES guide+covrg H Payor strategies **Joe** Lennerz

1

Key Elements, Next Steps, Timeline

1. Standardization of definitions, convene committee
   6 month (dedicated)
2. Technical target(s), survey for suitability
   1.5 years
3. Standard tissue target

2

Pros for Patient, Clinical, R&D, and regulatory

1. Common parameters that characterize any scanner
2. Enables the objective assessment of the common parameters in 1
3. Parameters as key variables for quality with relevance for patient care, clinical, R&D and regulatory

3

Concerns for patients, clinical, R&D, and regulatory

1. Could be perceived as limiting innovation.
   Education effort to explain can mitigate.
2. Unanticipated parameter(s) may turn out to be critical

4

Implications & Efforts

1. Increases uniformity of scanner evaluations.
2. Levels the competitive towards a comparative environment.
3. Should simplify the evaluation time and design time for new scanners (known endpoints)
4. Implications & efforts needs to be determined

5

Deliverables of the project and timeline

Work on DEFINITIONS
Education campaign
~12-15 months

6

Slide Scanning: Summary

Need: An outlined approach for scanner evaluation, which clearly delinates the necessary features for analytical studies.

Problem: Current scanner evaluation methods are too costly (>1M USD). Regulatory science does not dictate which features need to be examined, and each sponsor creates their lists ad-hoc, leading to iterative studies, increased labor on regulatory and vendor bodies, and increased costs.

Project Focus: A survey or other consensus-driven method to create an examination "checklist" for technical and analytical scanner evaluations.

7

Title: To develop automated pipelines to improve and mitigate the impact of subpar scanning fidelity.

Author(s): Laszik Z, Pekmezci M, Cadwell C, Bowman S, Kim G

Addressed Parties: "Who are you looking for in terms of collaborators, supporters, stakeholders?"

Industry, academia, developers, others. The nature of the project requires involvement by manufacturers. Our baseline data were generated on images from a Philips UFS platform; therefore, Philips IMS would be the ideal environment to execute the project. However, we are open to work with other vendors, developers, and collaborators. The proposed pipelines could be used to generate objective QC data ("standards") to assess certain performance metrics of scanners and image management solutions.

Background: "What is the current challenge? What is your problem statement?"

High fidelity scanning of histology glass slides is one of the prerequisites for successful adoption of digital pathology for primary diagnosis. Due to selective scanning of region of interest, fat-rich, highly-fragmented and/or faintly-stained tissues may result in loss of fidelity in whole slide digital images versus glass slides. To assess the scope of the problem and to generate baseline values we evaluated scanning fidelity on a select set of mastectomies [n=6; 429 slides] and Cavitron Ultrasonic Surgical Aspirator [CUSA]-resected brain tumors [n=40; 296 slides]) from the UCSF surgical pathology files from 2016-2018. Subpar scanning fidelity defined as tissues missing on whole slide digital images was identified in 163/429 (38%) and 160/296 (54%) slides of mastectomies and CUSA specimens, respectively. Although the diagnosis was not compromised in any of these cases, subpar scanning fidelity might pose challenges for the adoption of digital pathology for primary diagnosis for "high risk" specimens. There is an unmet need to improve scanning fidelity and automatically identify whole slide images with subpar scanning fidelity for quality patient care/patient safety and to mitigate potential medicolegal liability.

Approach & Objectives: "How are you addressing the problem?" Idea, Concept, Design

Objectives: **1.** To develop an automated pipeline to scan high risk slides at adjusted sensitivity to improve subpar scanning fidelity (SF) in order to maintain high quality patient care and to mitigate medicolegal liability. **2.** To develop an automated pipeline to identify images with subpar scanning fidelity to promulgate timely rescans and to mitigate medicolegal liability.

Materials, Design and Deliverables: The annotated whole slide image library and corresponding glass slides of UCSF mastectomies and CUSA specimens will serve as test environment for both objectives. Success/deliverables of objective #1 will be measured against baseline values as shown above. Success/deliverables for objective #2 (i.e., pick–up rate of slides with subpar fidelity) will be measured against visual assessment of the digital images by pathologists for subpar scanning fidelity.

Deliverable(s): What deliverable(s) will your project produce? See above.

Value proposition:
"How will the proposed project be valuable from each of these categories?" Address each in detail:

- Clinical: Successful completion of the projects will improve subpar scanning fidelity and streamline clinical workflow via promulgation of timely rescans.
- Regulatory: Successful completion of the projects will mitigate medicolegal liability.
- R&D: This is a R&D project to promote the spread of DP for primary diagnosis.

Funding sources: if existing, please mention source or propose a funding source: None; also see above.

Benefit to patients (≤2) : x☐technical advance, x☐increased quality, ☐outcome, ☐access, ☐affordability

References/Resources (optional): Unpublished data from UCSF; some data submitted for publication.

The Alliance

Breakout session worksheet

Key elements, next steps, timeline

Pros for Patient, Clinical, R&D, and regulatory

Deliverable and timeline

Concerns for patients, clinical, R&D, and regulatory

Implications and efforts

Breakout Session 3 – Truthing data sets

GUIDELINES + SURVEY A Pre-analytics **Amanda** Lowe	DEFINITIONS Education campaign B Slide scanning **Scott** Blakely	GUIDELINES + SURVEY C Truthing data sets **Sarah** Dudgeon **Brandon** Gallas	VER&VAL + IOP GUIDELINES D Continuous learning **Esther** Abels
STANDARD DATABASE E Data standardization **Markus** Herrmann **Mike** Isaacs	TOOLKIT WITH ROI CALC F Practical use cases **Matthew** Hanna	WHITE PAPER AI CATEGORIES G ML/AI – Model creation **Ashish** Sharma	BEST PRACTICES guide+covrg H Payor strategies **Joe** Lennerz

1

Key Elements, Next Steps, Timeline

1. What is the actual ground truth?
2. What to do when it doesn't exist?
3. How many samples do you need in a dataset and what distribution?
4. How would a digital slide-set be created and made available?
5. What information is needed?

2

Pros for Patient, Clinical, R&D, and regulatory

1. There is a clear application to adaptive algorithms
2. "leave room" for people to improve their programs
3. A standardized framework makes a lot of questions disappear – people can follow rules and generate data fast
4. Ground truth datasets improve speed to market for vendors, and faster access to treatment/algorithms for patients

3

Concerns for patients, clinical, R&D, and regulatory

1. If a company/developer does not have an approved scanner, or wants to use their own scanner, they will then create a dataset that is severely limited in its use.
2. Collecting consent for long-time period monitoring can be challenging and may not be fair to the patient
3. Each dataset could be too narrow per use case and could cause an increased burden for regulatory oversight

4

Implications & Efforts

1. Actual glass slides would be challenging to use as a ground truth dataset because these may get lost or broken
2. Slide images taken on one of the FDA-approved scanners would provide path to universal applicability
3. Patient treatment and comorbidities must be accounted for when assessing datasets

5

Deliverables of the project and timeline

Framework/guidelines for getting at ground truth regardless of application
Survey/Whitepaper: how to obtain ground truth data
~12-15 months

6

Truthing data sets: Summary

Need: Ground truth in data sets used for validation of AI/ML or other digital pathology tools/applications.

Problem: The ground truth is difficult to establish given the physical nature of the samples, variability in "truth" by physicians and outcomes – generalizability is difficult within these confines, since there is no "True" ground truth.

Project Focus: Establish and disseminate a framework for creating truthing datasets with a demonstrated use case to unify human elements to ground truthing.

7

Title: Tumor Infiltrating Lymphocytes (TILs) Annotated Whole Slide Imaging (WSI) Dataset for Analytical Validation of Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD)

Author(s):

Richard Huang, Sarah Dudgeon, Matthew Hanna, Brandon Gallas, Representatives of the High-throughput Truthing Collaborators

Addressed Parties:

Medical centers: to provide glass slides of specimens with TILs, and whole slide imaging scanning capabilities

FDA regulators: to provide regulatory guidance on mock submission and actual submission of the MDDT qualification process

Background:

Currently, there is no standardized set of data that is used across algorithm developers to analytically validate their algorithms for regulatory submission. Therefore, the FDA has to review multiple sets of data for the analytical validation of different algorithms. This is very burdensome for the reviewers.

Approach & Objectives:

The FDA has a Medical Device Development Tools (MDDT) program as a way for the FDA to qualify tools that medical device sponsors can use in the development and evaluation of medical devices. Qualification means that the FDA has evaluated the tool and concurs with available supporting evidence that the tool produces scientifically-plausible measurements and works as intended within the specified context of use. Our goal is to develop a qualified MDDT dataset that can be used by all algorithm developers to analytically validate their algorithms within a specific context of use.

Deliverable(s):

FDA qualified MDDT dataset.

Value proposition:
Clinical: Until algorithms reach the market, they cannot help real patients. Therefore, by making the review process more efficient, the algorithm approvals could happen faster without sacrificing patient safety. This would allow patients faster access to the best algorithms on the market.

Regulatory: This would significantly reduce FDA reviewers' burden when reviewing algorithms analytically validated using this MDDT, which would allow them to better use their limited resources to address other critical concerns during the algorithm review process.

R&D: This would allow different algorithm developers to analytically validate their algorithms against a common dataset, which would allow the algorithms to be better compared against each other. Furthermore, algorithm developers can be confident that the dataset used in their analytical validation would be accepted by the FDA as being valid.

Funding sources: (TBD)

Benefit to patients (≤2) : ☒technical advance, ☐increased quality, ☐outcome, ☒access, ☐affordability

References/Resources (optional):

Medical Device Development Tools (MDDT): https://www.fda.gov/medical-devices/science-and-research-medical-devices/medical-device-development-tools-mddt

Machine learning competitions as an aid to FDA regulatory approval in Digital Pathology.

A project proposal to the Alliance for Digital Pathology.

Author(s): Rajesh Dash, M.D.; Ricardo Pietrobon, M.D., Ph.D.; Brandon Gallas, Ph.D.

Addressed Parties. United States Food and Drug Administration (FDA), College of American Pathologists (CAP), Duke University (Duke), Data Science startup (SporeData Inc.), patient representatives (Friends of Cancer Research), Medical Device and Innovation Consortium (MDIC)

Background. The paradigm for regulatory approval of machine learning algorithms applied to Digital Pathology is challenging. First, whereas medical devices are physically stable, algorithms display concept drift (1), degrading quickly in the face of changes in the patient population, ever-evolving scanner technologies, technician experience in slide preparation, among other factors. Second, given the current pace of advances in machine learning, the current regulatory process required for sequential approval of improved algorithms is cumbersome and expensive for small companies, as recognized by FDA (2), thus imposing a barrier to innovative startups. Third, traditional documentation regarding model predictive performance – e.g., area under the curve, precision and recall – does not necessarily match model performance in the real world (3). Fourth, the datasets used for algorithm training and validation are often from single institutions, making the resulting Artificial Intelligence models less applicable to patients across the United States. Last, staff members at the FDA do not have the time and resources to test the algorithms' reproducibility. All of the factors mentioned above place patients at risk.

Approach & Objectives. Competitions involving machine learning address each of the previously raised issues in a novel way. In these competitions, data are made publicly available in a de-identified, labeled format, so that competitors can attempt to train and validate machine learning models with the highest predictive performance. The testing, however, is performed against a separate dataset only available to the competition organizers. Besides receiving a financial award, winners are invited to submit a manuscript describing their algorithm strategy, the resulting articles frequently being published in journals with high visibility (4). Competitions often lead to major advances in the field by, paradoxically, going back to the public demonstrations that used to characterize the peer-reviewed mechanism in the 1600s (5). In this project, we propose the use of competitions involving Machine Learning as a mechanism to augment the FDA's regulatory approval, monitoring, and upgrade processes. To accomplish this project, we will establish partnerships to accomplish three aims. First, we will combine human and machine learning methods to scale up the preparation of ground truth datasets (FDA-qualified Medical Device Development Tool or MDDT, https://ncihub.org/groups/eedapstudies/wiki/HighthroughputTruthingYear2) for future Digital Pathology Competitions and as additional publicly available validation resources. Second, we will explore methods to allow for better comparison across different machine learning models. This aim will involve a review of currently employed statistical methods and various metrics with a clearer qualification of those that represent theoretical measures of performance vs. those more likely to correlate with real-world patient outcomes. Third, we will conduct a set of evaluations of the competition, including user experience, health economics, and quality assessment of the implementation science. Findings will be summarized in an FDA report evaluating the viability of Digital Pathology Competitions as a regulatory mechanism. Common to all areas, we will take a patient-centered approach where patient-scientists will contribute to the design, conducting, interpretation, and knowledge dissemination of our project.

Deliverable(s). (1) A scalable Digital Pathology Competition preparation methodology that combines human tagging assisted by machine learning, aimed at the creation of Digital Pathology Challenges that deliver reproducible algorithms with high predictive performance across a range of digital images coming from diverse organizations and patient populations. Included in this aim is the provision of information technology infrastructure, e.g., open-source containers to increase the reproducibility of our methods. Last, we will also address security and privacy issues, e.g., by providing validated software to de-identify whole slide images (WSI). (2) A set of evidence-based guidelines to document various metrics, metric types, and associated statistical methods to help ensure that the comparison across algorithm results is meaningful. Based on

previous work from our group (6–8), we will also identify those that relate closely to patient outcomes, those focused on guiding clinical decisions, and those that represent largely theoretical measures of performance. Of importance, since the statistical tests will compare the results of classification and regression rather than the algorithms themselves, these methods will be robust across future machine learning methods. (3) Pilot results from a Digital Pathology Challenge conducted in partnership with the CAP to ensure quality review of data, subject matter expert resource availability, and concomitant alignment with evolving data and technology standards in this area. The CAP has mature processes in place for evaluation of the analytic quality of laboratory testing. We will leverage this experience to establish a set of evaluations of the competition, including user experience, health economics, and quality assessment of the implementation science. These results will serve as the basis for the preparation of an FDA report proposing alternative paths toward regulatory approval of Digital Pathology algorithms. We will generate at least two publications from each aim. Risks associated with our project include, first, accessing data for aim 1, which we will mitigate through the Duke repository currently managed by our Principal Investigator (R.D.). The main risk associated with aim 2 is protracted dissemination of these methods among the research and vendor community, which we will mitigate by releasing our guidance and code under an open-access license and statistical language and in working closely with the CAP. Vendor adoption of guidance and best practices will be more likely in settings where performance assessment for FDA approval and ongoing clinical accreditation and proficiency testing requirements are well aligned. Third, the main risk associated with the Machine Learning Challenge is the lack of interest from the community, which we will mitigate by offering financial compensation as well as the use of popular Challenge channels such as Kaggle (https://www.kaggle.com). We do not foresee the need for stopping rules at any of these stages.

Value proposition. **Clinical**: This proposal aims to deliver algorithms that are more reliable, reproducible, and, therefore, potentially safe to patients having their tissue samples evaluated through machine learning algorithms. **Regulatory**: The end product of this project will establish the foundation for a regulatory workflow that will potentially change the current FDA workflow. The proposed workflow is faster, reproducible, and agile enough to match the dynamic and ever-changing nature of machine learning algorithms. For example, we could develop an algorithm change protocol following the principles outlined in the Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning-Based Software as a Medical Device (SaMD, https://www.fda.gov/media/122535/download). We could then pursue an MDDT qualification of the protocol that would outline a streamlined retraining and evaluation pathway using the Machine Learning challenges discussed above. This MDDT would be enhanced over time to increase its representativeness across diverse patient populations, scanning devices, staining methods, among other characteristics. Aside from the FDA's approval, the CAP could also develop accreditation, proficiency testing, and certification programs so that health systems could leverage existing pathways for ongoing assurance of algorithm performance or quality. **R&D**: By creating a regulatory approval pathway that is easier to access, Digital Pathology will attract a more significant number of innovators from the startup ecosystem as well as from established medical device companies. This regulatory innovation will ultimately enhance society's ability to offer solutions that can improve diagnostic methods, decrease pathologist burden with mundane tasks, and ultimately improve the health care provided to our patients. Of importance, project achievements in Digital Pathology could also spread to other diagnostic areas involving AI algorithms. We welcome input and actively seek participation from MDIC.

Funding sources. We do not ask for funding from the Alliance. We propose to write a patient-centered grant proposal to be submitted to NIH, AHRQ, or PCORI. An award will allow us to not only provide support for the methodological development described above but also to access large datasets from academic and industry partners. A full budget will accompany the first draft of the proposal's Approach section.

Benefit to patients. Increased quality.

References/Resources:

1. Widmer G, Kubat M. Learning in the presence of concept drift and hidden contexts. Machine learning. 1996;23(1):69–101.

2. Food U, Administration D, others. Proposed regulatory framework for modifications to artificial intelligence/machine learning (ai/ml)-based software as a medical device (samd)-discussion paper and request for feedback. 2019.

The Alliance

Breakout session worksheet

Key elements, next steps, timeline

Pros for Patient, Clinical, R&D, and regulatory

Deliverable and timeline

Concerns for patients, clinical, R&D, and regulatory

Implications and efforts

Breakout Session 4 – ML/Models/Use cases

GUIDELINES + SURVEY A Pre-analytics **Amanda** Lowe	DEFINITIONS Education campaign B Slide scanning **Scott** Blakely	GUIDELINES + SURVEY C Truthing data sets **Sarah** Dudgeon **Hetal** Marble	VER&VAL + IOP GUIDELINES D ML-Continuous learning **Esther** Abels
STANDARD DATABASE E Data standardization **Markus** Herrmann **Mike** Isaacs	TOOLKIT WITH ROI CALC F Practical use cases **Matthew** Hanna	WHITE PAPER AI CATEGORIES G ML/AI – Model creation **Ashish** Sharma	BEST PRACTICES guide+covrg H Payor strategies **Joe** Lennerz

1

Key Elements, Next Steps, Timeline

1. Use current two guidances for changes (SaMD&MedDev) + locked AI to be changed + timeline
2. Use CADe and CADx + use RWD + timeline
3. Use cases, start with QC/aid of pathologist, already cleared IVDs (timeline)
4. Use SaMD, incl different scanners
5. Use differences btw radiology & DP
6. Need Pre sub and Mock sub
7. Need MDDT to create:

   a validated reference dataset (incl clinical outcome data)

2

Key Elements, Next Steps, Timeline

1. Open implementation/ROI toolkit+ vendor and site stakeholders for common parameters, 12 months
2. Common dictionary for real world pathology
3. Control slide/intermachine concordance

3

Key Elements, Next Steps, Timeline

Identify first tier categories from

1. AI
2. Clinical Use
3. Volume of Data needed

4

Pros for Patient, Clinical, R&D, and regulatory

1. Global access
2. Reduce time to market
3. Reduce costs
4. Reduce submission risk
5. Increase accuracy and precision in Dx
6. Increasing Precision medicine (right Dx and Rx for patient)

5

Pros for Patient, Clinical, R&D, and regulatory

1. Implementation calculator for ROI, making a business case, creating standardization across organizations/institutions; provide regulatory insights into apples to apples for each lab; increase adoption of digital pathology, increase clarity, maximize lab turnaround time
2. Providing better information to vendors for common terms
3. Pros for patient, clinical, R&D and regulatory

6

Pros for Patient, Clinical, R&D, and regulatory

1. Clean guidance on how AI algorithms are trained and evaluated

7

Concerns for patients, clinical, R&D, and regulatory

1. Pathologist will not review the image, HCP adoption, R&D and Class III
2. Feature picked by AI is different vs pathologist clinical (need RWD and clinical outcome data), R&D and regulatory
3. Reimburse Concerns for patient, clinical, R&D and regulatory

8

Concerns for patients, clinical, R&D, and regulatory

1. Potentially forcing change and introducing errors for laboratories who need to abide by new rules
2. Potential to change hospital lab workflows
3. Concerns for patient, clinical, R&D and regulatory

9

Concerns for patients, clinical, R&D, and regulatory

1. Ethical concerns?
2. Performance issues?

10

Implications & Efforts

1. Clarity on regulatory pathway – retrain and how often to release
2. Continuous learning implementing in RWD
3. Retraining vs self learning – using different data, need quality data – pre specify changes

11

Implications & Efforts

1. Get data out from multiple labs and compare data in, standardize those data points (private, academic, reference labs, industry)
2. Conforming to laboratory guidelines.
3. Implications & efforts

12

Implications & Efforts

Performance is affected by how algorithms are:

1. Trained
2. Evaluated and Validated
3. Their intended use is established
4. Clinical Guidelines are presented

13

Deliverables of the project and timeline

- Validation, Verification, and Interoperability Guidance
  ~12-15 months
- Digital pathology implementation toolkit with ROI calculator
  ~12 months
- A white paper outlining AI "categorizations" [that may go beyond/expand Class II/III distinction]
  ~8-12 months

14

ML Continuous Learning: Summary

Need: A better understanding of how to verify and validate continuous learning algorithms, currently none are approved by FDA.

Problem: There is currently too much uncertainty around the least burdensome approach for verification and validation as well as testing for continuous learning algorithms.

Project Focus: Guidance from FDA, with initial input from stakeholders in this group, on general principles for verification and validation testing for increased efficiency and access

15

Practical Use Cases: Summary

Need:

- For health systems to implement digital pathology solutions, a clear business case is necessary

Problem:

- Different groups approach the business case differently, with varying instruments, vernacular, technical validation standards, etc.

Project Focus:

- Toolkit to unify language, calibration, lab workflows, and other parameters to input into an ROI calculator measuring business impacts of digital pathology adoption in a standard way

16

Model Creation: Summary

Need: Guidance more specific to each unique type of algorithm that each serve a specific purpose.

Problem: The current guidance is a monolith that groups all of these diverse algorithms and purposes together, which is no longer representative of the future landscape.

Project Focus: A whitepaper outlining AI categorizations which enable a more directed guidance toward streamlined v&v testing.

17

Title: Imaging Informatics Fellows @ FDA

Author(s): Brandon D. Gallas (Proposal for discussion, The FDA has not made any decisions)

Addressed Parties: AMIA, FDA, Academia, Industry, undergrad and graduate students, and groups in the enterprise of medical imaging, informatics, statistics, and regulatory affairs.

Background: The FDA and industry are challenged by finding talent that can lead regulatory submissions and reviews and talent to do research and development that informs those submissions and reviews. The purpose of this project is to train fellows on 1) regulatory policy and practice, including reviews of actual devices, and 2) measurement science to support device assessment. The scope of the project includes medical imaging, informatics, and statistics. Fellows will also be trained in these areas.

Approach & Objectives: The idea is to create a training opportunity for a regulatory and research scientist in the field of medical imaging, informatics, and statistics. There are two pieces for this project. First, training development and training.

Training development (year 1): A team of subject matter experts will be identified to develop a syllabus designed around, for example, CDRH's "New Reviewer Training", AMIA's "Clinical Informatics Board Review Course (CIBRC)", API's ACGME-accredited Clinical Informatics Fellowships, STARD, Standards for Reporting of Diagnostic Accuracy Studies [1], and specific subject matter topics in medical imaging, informatics, statistics, and regulatory affairs.

Training (year 2): A steering committee will be identified to receive applications for the fellowship. Applications will include the person applying and a sponsoring mentor. Besides requisite training and interest, the application will be evaluated in terms of support from the mentor (research, funding, facilities). Preference will be given to fellows from organizations contributing to the *training development* and fellows that can, themselves, contribute to the *training development*.

To give context to the training, during both training development and training, participants will shadow FDA staff during critical moments of submission reviews occurring for medical imaging AI software.

Deliverable(s): A training syllabus and trained fellows.

Value proposition:
Regulatory: The training materials will be useful to train FDA reviewers of AI software. The trained fellows will be good (hard to find) candidates to be hired by the agency. They will also populate industry and disseminate the expertise learned from this training; this will improve submissions to the agency.

R&D: The training materials will give industry clarity on what the FDA expects in submissions of medical imaging AI software. The trained fellows will be good (hard to find) candidates for hire that will disseminate the expertise learned from this training; this will improve submissions to the agency.

Clinical: Better submissions lead to faster commercialization of medical imaging AI software that is expected to help clinicians make better decisions to improve public health.

Funding sources: Volunteers to develop the training materials and sponsors of fellows. These participants will benefit from intimate interactions with FDA and FDA reviews.

Benefit to patients (≤2) : ☐technical advance, ☒increased quality, ☐outcome, ☐access, ☐affordability

References/Resources (optional):

1. Bossuyt, P. M.; Cohen, J. F.; Gatsonis, C. A.; Korevaar, D. A. & group, S. T. A. R. D. (2016), 'STARD 2015: updated reporting guidelines for all diagnostic accuracy studies.', *Annals of translational medicine* **4**, 85.

Title: Building a Framework/Platform for creating data sets and evaluation tools: Key for Generalizability Evaluation and Monitoring

Author(s): Rajendra Singh, MD; Matthew Hanna, MD
Addressed Parties: Academia, FDA, MDIC, Data scientists

Background: Local data set curation is often the most facile for developing machine learning models due to the proximity of the data, however it has been shown that development site data alone often does not generalize to deployment site validation data. There is a need for an open access platform to share data across the world, and collaboratively generate annotated data across a multitude of use cases. Additionally, many organizations are attempting to gather open access data or participate in grand challenges in order to supply their armamentarium of machine learning tools. The key to building true clinically relevant models and algorithms that can predict patient outcomes, management, or prognosis is having access to a large and varied patient data. The only hope of building large data sets is when multiple institutions and organizations find a way to share their data as no single institution would have enough diverse data that would help build tools that can be used worldwide. The innate variation of each institution by itself could cause modeling failures when isolated models are built and transferred to other institution data.

Approach & Objectives: Gaining access to high quality big data sources, especially open access, will need to have shared data governance, accuracy, and dependability. Open access platforms with deidentification or anonymization will need to provide these principles to support such deliverables. Web based platforms will allow for collaborative annotations made on the data, which will also need to be verified; in order to produce viable models for real practice.

Deliverable(s): An open, web-based platform for creating public and private datasets has been developed with collaborative tools for annotation tasks. This annotation platform can import and export XML annotation data, as well as annotated images (with the option of RGB normalization) of labelled data. Open data collections can also be organized for public annotated data collection.

Value proposition: Datasets generated from anatomic pathology practice, research, or education can be shared, trained, and validated in crowd sourced datasets with high variance for reliable modeling. Development such an innovative, interactive, and intuitive workflow will be the first step in fulfilling the role of pathology becoming the center of personalized medicine. Development of machine learning models based on such diverse and large-scale datasets will ensure appropriate performance when clinically deployed and validated on non-development site data. Regulatory bodies may use this platform as a framework for discrete data of annotations that has transparency with available importing/exporting and analysis of labelled data. The platform will allow the use of common datasets (and potential MDDT) to be used to explore generalizability and "Algorithm Change Protocols" as described in the FDA white paper, "Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning [AI/ML] Based Software as a Medical Device (SaMD)." Researchers will also benefit by having an open network of annotators to assign tasks for specific intended model development with easy access of off-loading generated data for subsequent model development.

Funding sources: Microsoft has provided credits for storage of data and use of the Azure cloud

Benefit to patients (≤2) : This platform will ultimately enable the regulatory body a framework for dataset curation, annotation, and discrete data for the research community to gain access to high quality datasets with enough variation to withstand clinical variability; and ultimately ensure patients are provided with clinical grade trained machine learning models. These models may provide increased quality of care to patients, or better diagnostic and prognostication for their pathology.

References/Resources (optional): AI.pathpresenter.net

Author(s): Daniel L. Rubin, MD, MS, Stanford University

Addressed Parties: We are looking for industrial collaborators, supporters, and stakeholders. We already have identified collaborators at FDA and at the American College of Radiology.

Background: There is explosive growth in the number of AI algorithms for decision support evaluated by FDA for clearance. The current regulatory approach has limitations in that the data used for evaluating algorithms may not be broadly representative of patient data in actual practice. It is difficult to assemble a sufficiently broad dataset to be used for evaluating AI algorithms for FDA clearance. Consequently, **post-marketing surveillance** is crucial to ensuring that FDA-cleared AI tools are effective, to identifying failures, and to providing a means of providing vendor feedback of such failures so that they can improve their products. Tools are needed to capture the ground truth for cases on which AI is applied (e.g., the radiologist's agreement or disagreement with AI for an image detection task, or the pathological diagnosis for an AI disease classification task). Our lab has already been developing foundational methods to enable post-marketing surveillance of AI diagnostic tools, and in this proposal, we wish to refine and apply those methods to meet those needs through partnership with industry and ACR to demonstrate feasibility.

Approach & Objectives: We recently developed infrastructure to enable collecting data needed for post-marketing surveillance data on AI diagnostic algorithms: the open source ePAD platform (http://epad.stanford.edu) provides a web based image viewer and annotator that permits radiologists to view images and create of modify annotations on images (e.g., view and edit annotations created by AI algorithms). It is also standards-based, using DICOM standard objects for saving outputs of AI algorithms that anticipates future support by commercial PACS that will be able to incorporate our methods into routine radiology workflow. We will use ePAD as the testbed for our approach, with a commercial partnership we already have with an AI vendor of a CT brain hemorrhage AI product, though our methods will generalize to many other AI products. We will extend the ePAD platform to enable collecting radiologist-derived ground truth and create a registry of AI failures (using the radiologist and medical record). We will work with industry to codify standards in the data objects that are produced by AI algorithms to facilitate interoperability with systems deployed in hospitals such as ePAD or PACS that will permit our methods that collect AI algorithm performance metrics to be widely deployed for a post-marketing surveillance approach. We will evaluate our work using historical clinical data at our institution. We have partnered with the ACR to incorporate ePAD into their AI-LAB to collect performance metrics of AI algorithms in practice and to submit those to an AI registry of performance metrics to establish national benchmarks.

Deliverable(s): We will provide working software that implements our approach to post-marketing surveillance and we will illustrate its use by doing surveillance of a commercially available AI algorithm.

Value proposition:
"How will the proposed project be valuable from each of these categories?" Address each in detail:

- Clinical: Post-marketing surveillance of AI algorithm performance is crucial, since literature is showing that AI algorithms often do not perform as well on real-world data as on the data used to train them, including data that is the basis for FDA clearance of the algorithms. The risk to a practitioner is reliance on an AI product that may not work as well as expected, with potential patient harm.
- Regulatory: FDA has stated the need for post-marketing surveillance of AI algorithms, and is looking for guidance as to how this can be accomplished. This proposal will deliver a potential solution to this regulatory challenge. In addition, our methods will permit continual evaluation as vendors improve their algorithms, providing a way to address the regulatory challenge of oversight of AI performance as algorithms are versioned.
- R&D: Our work helps to identify standards that will improve data interoperability. Our proposal also enables catalyzing R&D on developing improved AI algorithms since we can identify AI failures at local sites that can be fed back to AI vendors to improve their algorithms.

Funding sources: none currently

Benefit to patients (≤2) : ☒technical advance, ☒increased quality, ☐outcome, ☐access, ☐affordability

References/Resources (optional):

The Alliance

Breakout session worksheet

Key elements, next steps, timeline

Pros for Patient, Clinical, R&D, and regulatory

Deliverable and timeline

Concerns for patients, clinical, R&D, and regulatory

Implications and efforts

Breakout Session 5 – Data standardization

GUIDELINES + SURVEY A Pre-analytics **Amanda** Lowe	DEFINITIONS Education campaign B Slide scanning **Scott** Blakely	GUIDELINES + SURVEY C Truthing data sets **Sarah** Dudgeon **Hetal** Marble	VER&VAL + IOP GUIDELINES D Continuous learning **Esther** Abels
STANDARD DATABASE E Data standardization **Markus** Herrmann **Mike** Isaacs	TOOLKIT WITH ROI CALC F Practical use cases **Matthew** Hanna	WHITE PAPER AI CATEGORIES G ML/AI – Model creation **Ashish** Sharma	BEST PRACTICES guide+covrg H Payor strategies **Joe** Lennerz

1

Key Elements, Next Steps, Timeline

1. Interoperability (e.g. integration with LIS, scanner agnostic algorithms)
2. Standardization at different levels (e.g., images, reports) and linking/tying information together
3. Integrate into medical record

2

Pros for Patient, Clinical, R&D, and regulatory

1. R&D: Vendor neutral interpretation of data
2. Patients: sharing data between institutions in a format anyone can open and read, defined content for identification and matching
3. De-identification but still linked

3

Concerns for patients, clinical, R&D, and regulatory

1. Why has standard not yet been adopted?
2. Consent issues if data is accessible more easily?
3. Can standard evolve and adapt fast enough to new requirements?
4. Democratization is long and complicated?

4

Implications & Efforts

1. Circular dependency: manufacturers are waiting for customers and users are waiting for products
2. Starting somewhere: baseline features/elements that are required but still leaving flexibility (basic functionality versus perfect solution)
3. Taking action beyond white paper (including academic medical centers and pathologists)

5

Deliverables of the project and timeline

Database of standards; datasets as a resource/tool for algorithm development
~12-15 months

6

Data Standardization: Summary

Need:

- Interoperability between digital phases of the healthcare enterprise, consisting of various data types (EMR, images, specimen processing, LIS information, etc.) and data sensitivity, for “plug and play” integration and better patient handling

Problem:

- Various components of health data are stored differently, do not integrate well with legacy systems, and new formats are developed with every new system, perpetuating lack of conformity

Project Focus:

- Reference database of standard datasets to help medical device developers understand what mixed, multi-modal data could look like if it allowed forward and reverse compatibility, thus promoting conformity and standardization across the field

7

Title: A Framework for Remote Sign-out in CLIA20

Author(s): Joseph Sirintrapun & Jochen Lennerz

Addressed Parties: Domain experts including pathologists, vendors, patient-advocacy groups, representatives of regulatory entities, and legal counsel.

Background: To be able to fully leverage the advantages of digital pathology, we need to evaluate the constraints on the regulatory governance of laboratory tests. In particular, CLIA88 as it applies to digital pathology and sign-out activities. CLIA88 in its current form creates a barrier for remote review and clinical diagnostics such that any potential remote site requires a separate CLIA license. The regulatory framework does currently not reflect the technological advances in the field of digital pathology. We consider CLIA88 in its current form a key limitation to prevent remote sign-out and diagnosis.

Approach & Objectives: To unlock the full potential of digital pathology for remote diagnosis, we propose identification of the specific prohibitive sections in CLIA88, review and collection by domain experts including pathologists, vendors, patient-advocacy groups, representatives of regulatory entities, and legal counsel. A key strategic element of this approach is to integrate the patient perspective and accommodate the legal and policy-related issues to move the field of digital pathology forward.

Deliverable(s): The group will produce a consensus whitepaper that can ultimately serve as one element of a future CLIA88 amendment.

Value proposition:
Clinical: The ability to review and sign-out cases remotely will provide diagnostic access in underserved regions as well as improve the quality of diagnoses via increased transparency (=consensus) and safety.

Regulatory: The whitepaper will serve as a guidance document to enable future CLIA revisions to reflect the current scientific progress.

R&D: By decreasing the current regulatory barrier, the produced content will increase utilization of digital pathology and thereby incentivize further R&D efforts.

Funding sources: N/A

Benefit to patients (≤2) : ☒technical advance, ☐increased quality, ☐outcome, ☒access, ☐affordability

References/Resources (optional):

Hanna MG, Reuter VE, Samboy J, England C, Corsale L, Fine SW, Agaram NP, Stamelos E, Yagi Y, Hameed M, Klimstra DS, **Sirintrapun SJ**. Implementation of Digital Pathology Offers Clinical and Operational Increase in Efficiency and Cost Savings. Arch Pathol Lab Med. 2019. PMID: 31173528

Hanna MG, Reuter VE, Hameed MR, Tan LK, Chiang S, Sigel C, Hollmann T, Giri D, Samboy J, Moradel C, Rosado A, Otilano JR 3rd, England C, Corsale L, Stamelos E, Yagi Y, Schüffler PJ, Fuchs T, Klimstra DS, **Sirintrapun SJ**. Whole slide imaging equivalency and efficiency study: experience at a large academic center. Mod Pathol. 2019 Jul;32(7):916-928. PMID: 30778169

Title: De-Identification of WSI images using Artificial Intelligence

Author(s): Vijay Narayanasamy

Addressed Parties: Pathologists, Digital Pathology Manufacturers, Informatics/AI Engineers

Background: Creating Pathology WSI databases requires expensive, time consuming and inefficient manual de-identification by re-scanning slides after applying a new slide label to hide the original label with patient identifiers. This significantly limits the availability of WSI for creating large datasets that could be useful for research and to develop AI-aided image analysis tools.

Approach & Objectives: We plan to develop Software that will

a. detect and open all WSI file-formats, extract associated texts (annotations, properties, headers, filenames etc.) and extract associated images (label, macro, thumbnail).
b. extract text present in image files (slide label image) using Computer Vision algorithms
c. analyze text to detect and remove PHI from all extracted text using NLP algorithms
d. create de-identified WSI

Deliverable(s): What deliverable(s) will your project produce?

A fully automated de-identification software for WSI images of any file format that could help with creating WSI datasets for research, education and developing AI-based image analysis tools.

Value proposition:
"How will the proposed project be valuable from each of these categories?" Address each in detail:

- Clinical – Helps to create a vast WSI image datasets for Similar Cases, AI-based image analysis
- Regulatory – Helps to comply with HIPAA, GDPR and other privacy laws and IRB policies. Enables creation of WSI bases Real-World evidence
- R&D – Helps to get WSI data de-identified for clinical trials and multi-site studies. Publicly available large datasets can help to do secondary research without expense of collecting new data

Funding sources: if existing, please mention source or propose a funding source

None. Applying for NIH SBIR funding

Benefit to patients (≤2) : ☒technical advance, ☒increased quality, ☐outcome, ☐access, ☐affordability

References/Resources (optional):

Author(s): Kingsley Ebare & Esther Abels

Addressed Parties: Industry, Academia, Biopharma, Regulatory Bodies who work with clinical studies and artificial intelligence in digital pathology

Background: Algorithms have shown to support drug development and decisions in diagnostics. Diagnostic Algorithms have been developed and are in clinical use in the clinical field for a few decades. Pathology is the one of the last disciplines in healthcare to become digitized. To develop and use algorithms to its full potential, the metadata is crucial input. One of the essential elements of training algorithms is the input using clinical data, such as clinical features and clinical outcome data. In the drug industry, data is standardized using Clinical Data Interchange Standards Consortium (CDISC) operational data model and MedDRA. However, in the medical device field and especially in the field of pathology, the standardized use of medical terminology is lacking.

Approach & Objectives: The objective is to develop a dictionary with standardized data so coding can be applied to improve data collection, safety information, retrieval, evaluation, statistical analyses, and presentation to improving scientific reporting and to reduce the regulatory burden relevant for submissions to Regulatory Authorities. It will also allow for sharing regulatory information globally for In Vitro Diagnostic devices used in Digital Pathology.

There are three main areas that require coding for algorithm development: 1. Training 2. Verification and 3. Validation.

Deliverables: Create a framework for standardization of data used in the design, execution, analysis, regulatory submission and archival of pathology research studies using CAP cancer protocols as input. Also develop a dictionary to code pathologic features as well as medical pathologic diagnoses.

Value proposition: Standardized pathology data is currently missing despite the availability of some tools like synoptic reporting for cancers. Providing this to the industry and regulatory authorities would be helpful in training, generating and presenting evidence for algorithms used in clinical diagnostic decision making. Also, with increasing demand for interoperability, pathology data standardization will allow for efficient data exchange, easier aggregation and analysis and in the long run decreased cost of conducting research.

Clinical – Standardization and removal of potential error sources will improve clinical interpretation in a harmonized way for pathology diagnosis.

Regulatory – Standardization of data facilitates statistical analyses and allows meta analyses to be used for showing devices are safe and performing to its intended use. In addition, it facilitates the analyses and cross analyses with data collected during drug trials, for example in the field of Companion Diagnostics

R&D – Tissue, imaging and metadata standardization will allow harmonized usage in training and as such allows for a general training approach of algorithms which will have a widespread impact into research and development.

Funding sources: None

Benefit to patients (≤2) : ☒technical advance, ☒increased quality, ☒outcome, ☒access, ☐affordability

References/Resources (optional):

https://www.meddra.org/how-to-use/support-documentation/english (accessed Oct 27, 2019)

https://www.cdisc.org/standards (accessed Oct 28, 2019)

The Alliance

Breakout session worksheet

Key elements, next steps, timeline

Pros for Patient, Clinical, R&D, and regulatory

Deliverable and timeline

Concerns for patients, clinical, R&D, and regulatory

Implications and efforts

Breakout Session 6 – Payor strategies

GUIDELINES + SURVEY A Pre-analytics **Amanda** Lowe	DEFINITIONS Education campaign B Slide scanning **Scott** Blakely	GUIDELINES + SURVEY C Truthing data sets **Sarah** Dudgeon **Hetal** Marble	VER&VAL + IOP GUIDELINES D Continuous learning **Esther** Abels
STANDARD DATABASE E Data standardization **Markus** Herrmann **Mike** Isaacs	TOOLKIT WITH ROI CALC F Practical use cases **Matthew** Hanna	WHITE PAPER AI CATEGORIES G ML/AI – Model creation **Ashish** Sharma	BEST PRACTICES guide+covrg H Payor strategies **Joe** Lennerz

1

Key Elements, Next Steps, Timeline

1. Include payors understand pain points
2. Educate medical directors of health plans
3. Population health analysis
4. Focus on individual diseases
5. Review non-pathology approaches
6. Payment not necessarily payor related: approach bundled payments (e.g. DRG)

2

Pros for Patient, Clinical, R&D, and regulatory

1. Efficiency gain of digital pathology may be interpreted as cost-savings by payors
2. Workforce gap: Reduces total number of pathologists (17%) with anticipated growth in field (40%)
3. Economic impact study
4. Time to diagnosis

3

Concerns for patients, clinical, R&D, and regulatory

1. Reduced rates of reimbursements
2. We need evidence that digital pathology decreased cost and/or improves outcomes
3. IHC may be replaced = no bill (e.g. H. Pylori)
4. Road to revenue to monetize ML/AI
5. Who is going to pay for clinical utility studies
6. High implementation cost

4

Implications & Efforts

1. How many people are doing this?
2. Importance of data: Lab, Payor, Patient
3. More independent laboratories (agile players)
4. Cost of storage; maintenance cost
5. Value of data – as opposed to reimbursement

5

Deliverables of the project and timeline

Approach payors and clinical guideline developers
~4-6 months

6

Payor Strategies: Summary

Need:

- Payors need conclusive clinical outcomes and utility data
- ROI for digital pathology solutions

Problem:

- Clinical utility studies are costly, but necessary
- How to effectively integrate digital pathology into existing clinical decision-making guidelines

Project Focus:

- White paper for best practices to understand coverage determination
- White paper for best practices to incorporate digital pathology into NCCN guidelines

7

Title: Payor- and Patient-Perspectives in Digital Pathology

Author(s): Joe Lennerz & Esther Abels

Addressed Parties: Payors, Patient-Representatives, Consultants, Policy Makers, Health Care Providers

Background: Currently, the lack of clearly demonstrated added value continues to stall larger-scale clinical adoption. The return on investment remains unclear and financially sustainable clinical integration is lacking. The promise of ML/AI is relying on large-scale clinical implementation; however, the practice, clinical and economic impact as well as clinical utility remains to be demonstrated.

Approach & Objectives: There is a clear disconnect between industry, researchers, payors, health care provider, government and patients. We need patient- and outcome centered studies that focus on demonstration of added value(s). While this demand is not new, (unfortunately) existing efforts largely ignore the patient, provider and payor perspectives. The objective of this project is to capture these positions and opinions.

Deliverable(s): We are aiming capture opinions of Provider- Payor- and Patient Representative, as well as required evidence for a financial interest from payors several Consultants in the space as well as the FDA to draft a position (white) paper that captures these vantage points in clear and concise manner.

Value proposition: Capturing the payor and patient-perspectives and alignment with clinical provider economics aspects is currently missing. Providing this to the industry would be helpful in charting our strategies for financially sustainable clinical integration.

Clinical: Demonstration of clinical utility for digital pathology is very important; however, require larger-scale adoption. Providing a position/perspective that entails the patient and provider perspective is currently missing.

Regulatory & reimbursement: The aim is to provide distinct vantage points that will take into account the patient and payor perspective and required evidence to clearly delineate what is currently missing and the field of digital pathology needs to deliver.

R&D: Once clinically integrated, continuous streams of digital image and metadata will be available for research and continued development

Funding sources: We can

Benefit to patients (≤2) : ☐ technical advance, ☐ increased quality, ☒ outcome, ☒ access, ☒ affordability

References/Resources (optional):

Abels & Pantanowitz. Current state of the regulatory trajectory for whole slide imaging devices in the USA. J Pathol Inform 2017, 8:23

Lennerz et al., Health Care Infrastructure for Financially Sustainable Clinical Genomics. J Mol Diagn 2016, 18(5)

The Alliance

Breakout session worksheet

Key elements, next steps, timeline

Pros for Patient, Clinical, R&D, and regulatory

Deliverable and timeline

Concerns for patients, clinical, R&D, and regulatory

Implications and efforts