PointClickCare

PointClickCare as an ONC certified health IT developer with Health IT Modules certified to Decision Support Intervention criterion at **§ 170.315(b)(11)** is obligated to apply risk management practices for all Predictive DSIs, ensuring they are subject to risk analysis and mitigation related to validity, reliability, robustness, fairness, intelligibility, safety, security, and privacy. It is our intention to respond to each criterion requested with clear and concise language, demonstrating compliance and transparency through the publishing of this document.

Intervention Risk Management (IRM) practices are a set of activities used to analyze and mitigate various kinds of risks associated with Predictive DSIs. IRM practices also include policies and controls for governance and data management related to Predictive DSIs. The Certification Program specifies that risk analysis and risk mitigation should cover topics identified in the NIST Artificial Intelligence Risk Management Framework, including validity, reliability, robustness, fairness, intelligibility, safety, security, and privacy.

A. Risk Analysis

Every Predictive DSI that PointClickCare supplies as part of its Health IT Module requires the management of risks and adverse impacts related to (1) validity; (2) reliability; (3) robustness; (4) fairness; (5) intelligibility; (6) safety; (7) security; and (8) privacy. Every Predictive DSI will therefore have these concepts integrated into the Health IT Module lifecycle. Following NIST, and ONC HTI -1 regulations, and defined for our teams into each phase of our product development lifecycle we will establish, define, and share the following:

1. Validity means confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled. Each model which PCC develops will have a determined metric which by scientific rigor, sets model validity. Metrics like F1, AUC, etc. are used. Our team will ensure the model has been tested on a diverse dataset that reflects the population consistent with the Health IT Module where it will be deployed. Model monitoring will verify that the model's predictions align with established accuracy range.

For the predictive Return to Hospital (pRTH) model, we calculate AUC & F1 to measure the model's accuracy and ability to distinguish between patients who will and will not be readmitted in the next 7 days. We validate model accuracy metrics to make sure they match pre and postproduction set range.

In determining model accuracy, we sought to validate the models' predictions with those made by humans. We recruited clinical leaders, mostly registered nurses in the roles of Director of Nursing, or Assistant Director of Nursing, and some Case Managers. We collected over 2000 human predictions and validated them against the models' predictions. The model performed better than nurses in the role of Case Manager and equal to nurses in the clinical leadership roles. We used power analysis to determine the number of human predictions needed and used

this work to establish a human baseline of prediction accuracy to compare the model's accuracy rates too. Each time the model is versioned, the team follows a model checklist including accuracy testing. Model monitoring is automated, and the team is alerted if the set threshold is reached meaning the model is not at the point of accuracy we set. To date the set threshold has not been reached, the model remains performing in the range of accuracy set.

2. Reliability is determined by performing repeated testing to ensure consistent model performance. We check for stability in predictions with similar patient profiles. The metrics used to measure predictive Return to Hospital model (pRTH) include Area Under the Curve (AUC) and F1 score. The F1 score is a metric for evaluating the accuracy of a model, balancing precision, and recall. It is the harmonic means of precision and recall, providing a single measure of a model's performance. We use AUC and FI to assure the model's ability to accurately distinguish between patients who will and will not be likely to readmit to a hospital setting in the next 7 days. We monitor the model's performance in production and test the reproducibility as part of validation testing among our different engineering environments to ensure model reliability. The Machine Learning Prediction (MLP) Testing process at PointClickCare takes necessary steps to ensure the prediction pipeline produces consistent results, including preparing test data, running predictions, and comparing results across multiple engineering environments.
3. Fairness is evaluated through a review of the model for biases based on demographics such as age, gender, and ethnicity. At PointClickCare, our teams study measures of bias at the time of model development and over time, through positive predictive accuracy by age, gender, and ethnicity. For pRTH our analysis to date includes as assessment of the model by age, gender and ethnicity.

In skilled nursing facilities, there is a measure of quality named Five-Star, a rating calculated by the Centers for Medicare & Medicaid Services (CMS). We studied the demographic analysis of the racial makeup of patients in Skilled Nursing Facilities (SNFs) broken out by five-star rating. There is a statewide forced distribution of five-star facilities such that only 10 % of SNFs in each state can be 5-star, and 20% will be 2-star or less. For more on the CMS Five Star rating program see https://www.cms.gov/medicare/health-safety-standards/certification-compliance/five-star-quality-rating-system.

In United States nursing homes the predominate race is white, and the predominate gender is women. The proportion of black patients in 1-star facilities is far higher than the proportion of black patients in five-star facilities. The rehospitalization prevalence for black patients is also higher than for white patients. We did find that model accuracy metric AUC was 0.01-0.02 points lower for black patients, even when broken out by five-star rating. The accuracy metric F1 score was even smaller among black patients likely because of the higher rehospitalization prevalence. Additional ethnic groups in the data are not well documented and fall into “other” categories due to the age of the information. Recent changes in data collection under ONC make it possible

for future monitoring of bias and fairness to be titrated to other ethnic groups and may show some regional variation in the United States SNF markets.

In an age analysis the model's accuracy is worse for younger patients, particularly those under 18, due to the small sample size and different health conditions.

4. Robustness refers to a model's ability to maintain stable performance when exposed to variations or unexpected changes in input data. Robustness can be measured using adversarial accuracy, which evaluates the model's performance when subjected to erroneous inputs. ?(@ben when we control for crazy outlier vitals?
5. Intelligibility is critical to transparency and trust. PointClickCare health IT models will provide clear explanations for the model's predictions and ensure that clinicians can understand the model's decision-making process. FAQs are available on screen for the user to access in their workflow.

The metrics include risk score, a risk group classification, one-day change indicator, listing of contributing factors and their values, trends, and original source notes of importance to the model output. User feedback will be monitored for indication of non-intelligibility.

6. Safety assessments are performed as part of initial model ideation, in development and as part of ongoing monitoring. The team includes members from the clinical advisory team and other clinicians as needed to assess the risk associated with the model idea. To answer the question what might happen if the model prediction or classification contains errors. For classification models an assessment of the potential for harm for each of the four possible model outcomes is conducted. Models with minimal risk for harm following clinical risk assessment are considered safe. Models which are not classification models, such as generative AI models use alternative assessments of clinical safety. In all cases the team works to understand and validate that the model's use does not introduce new risks to patients served by the Health IT Module.

The pRTH clinical risk assessment resulted in minimal risk of actual harm. We monitor and track false negative rates and have not had any drift in the model outside of the acceptable model accuracy range.

7. Security at PointClickCare is a number one priority. Engineering teams follow PointClickCare' s security risk assessment protocols and procedures to ensure that the model and its data are protected against unauthorized access, as we seek to maintain compliance with healthcare data security regulations. Our teams conduct routine security scans, audits, and reviews. Our teams follow PointClickCare security policies and security review for new software with PRR. Security scanning is completed by engineering with each sprint, all critical and high risks are corrected before production deployment. For pRTH there have been no security incidents to date.
8. Privacy in healthcare is a priority which ranks as a top priority right alongside security in PointClickCare' s Health IT Module. We limit our teams to building models only with data

where our customer agreements which include data sharing for model creation permissions. We verify at the start of each model training, or versioning phase of work, that the data to be used has been authorized. We monitor the percent of organizations with enhanced BAA agreements for model creation. For pRTH 100% of the organizations whose data was used to train the model have an enhanced BAA in place.

B. Risk Mitigation: The Predictive Decision Support Intervention(s) must be subject to practices to mitigate risks. No model is without risk and to the extent that these risks are known steps are put in place to mitigate such risks. The effort should include the following:

Practices are used to prioritize or establish different levels of risk.

Predictive Decision Support Intervention(s) will always come with some risk of predicting 'wrong'. This is true for humans who also make predictions about future state of outcomes, based on their own training, and the data they have collected in the present. For Classification models the team considers along with clinical advisors, SMEs, and other clinical personas what risk of actual harm and what alternative human surveillance systems would otherwise remain in place. Estimating the risk for actual harm at each potential model output.

Practices to mitigate or minimize identified risks are considered for all predictive decision support models. We acknowledge that no model will predict 100% accurately for every setting where Health IT Modules with predictive decision support intervention technology is utilized. To address this risk and mitigate potential risks PointClickCare seeks to establish a baseline for model accuracy and communicating to users when the model accuracy threshold has not surpassed a level of acceptable accuracy for their setting. This was conducted as a part of model feasibility, early in our development process. Communicating to users when the model accuracy threshold has not surpassed a level of acceptable accuracy is built into the Health IT Module and user materials.

Baseline model accuracy and accuracy for each skilled nursing facility within each organization is computed. Warning Messages are displayed along with a warning icon for those facilities below the baseline. See the section on Validation above to review the work done on human baseline.

Mitigating risk through warning message: For pRTH we studied individual skilled nursing location accuracy compared to baseline with each version of the health IT model before deployment. 94% of SNF locations were above the baseline. We identified the risk with underperforming facilities and mitigated this risk with a warning icon next to each resident prediction when the facility accuracy is below the model baseline. The note to staff reads: This prediction might include incorrect or incomplete information. It is recommended to review the clinical chart to evaluate the risk of rehospitalization.

1. Change control plans or ongoing validation and updating processes. We perform ongoing model validation. We monitor model accuracy routinely and we have alerts built into monitored dashboards. With model version updates we have conducted model accuracy testing, and model validation.
2. Processes to supersede, disengage, or deactivate deviations from intended use. For each PointClickCare Health IT Module, a limited set of users have the ability to enable or disable any predictive decision support intervention models provided with the module. We have an enablement process where such users can disable the model. pRTH is enabled by each organizations designated administrative or "super - user". Enablement by user role and security settings is standard PointClickCare Health IT Module process.

Our process utilizing a warning message mitigates the risk of unintended use. As stated, when model accuracy is not above baseline threshold at the facility level - the predictions have a warning icon and message to supersede the intended use.

3. Approaches to include the including of subject matter experts (SMEs) in measuring/validating performance. PointClickCare includes SMEs in measuring and validating predictive decision support intervention models. The intentional inclusion of SMEs has long been aligned with PointClickCare' s product development lifecycle and predictive decision support intervention model development.

For pRTH, a masters prepared nurse served as SME to the model building team and served as product manager, conducting proof of concept work and model validation feedback sessions with clinical leaders from dozens of our customer sites. In the proof-of-concept phase 2000 clinical predictions are made by nurses, physicians, nurse practitioners. The baseline accuracy of the pRTH model was computed using predictions from SME, clinical leaders from SNF locations.

C. Governance

PointClickCare has adopted the NIST framework to guide the organization in AI. We aligned the Map, Manage, Measure concepts with our existing product development lifecycle, adding key elements of work in each product phase. All staff were required to attend training on AI, including the basics of NIST, ONC and other governing works. The roles and responsibilities for compliance with RAI were established, trained, and are observed for modification as needed. The existing governate teams including established corporate security, and AI Steering committees, Clinical Advisory Committees, are all involved in the governance of RAI models.

PointClickCare' s BAA agreements with their clients defines our polices and controls regarding data rights, use and management. The only data used for model training comes from organizations where permission has been granted for use of data in models. Per certification requirements we make summary information regarding IRM requirements publicly available. PointClickCare supports this broad transparency so that users, patients, researchers, and other

interested parties can understand the steps we take to identify and mitigate risks related to Predictive DSIs we supply as part of our Health IT Modules.

As an ongoing part of certification, we will review and update transparency information, as necessary. We recognize that we are required to review and update descriptions of DSI source attributes listed in § 170.315(b)(11)(iv)(A) and (B) as needed, certainly with any updated version of a model, or when changes to source attributes occur. We acknowledge an ongoing responsibility for reviewing and updating our IRM practices for all Predictive DSIs supplied by PointClickCare.