

## **ABBREVIATIONS**

### **1. COVER PAGE**

## 2. SPECIFIC AIMS

Early diagnosis of Autism Spectrum Disorder (ASD) and the timely intervention is widely recognized as critical for achieving improved cognitive, behavioral and social outcomes.<sup>[1]</sup> While a diversity of factors are implicated,<sup>[2]–[5]</sup> the etiology of Autism is still unknown. With no laboratory tests for ASD, and despite advances from widespread adoption of screening with standardized checklists at 18 and 24 months of age, the median age of diagnosis remains over 4 years. Starting with a positive initial screen, a clinical diagnosis of ASD is a frustrating multi-step process spanning 3 months to 1 year, often delaying time-critical intervention. One obvious source of these delays is the vast number of false positives encountered in the current initial screening tools. For example, the M-CHAT/F, the most widely used screen,<sup>[1], [6]</sup> produces about over 85 false positives out of every 100 flagged for diagnostic evaluation, significantly inflating wait times<sup>[5]</sup> especially in rural and underserved communities. Further, current screening tools are sensitive to language barriers and cultural issues, and are particularly ineffective for children with milder symptoms with average or above-average cognitive abilities until about school age,<sup>[1], [7]</sup> often due to a “wait and see” approach adopted at the primary care.

In this study, we plan to develop and validate the efficacy of machine inferred **digital biomarkers** for autism, mined automatically from past medical encounters. Using individual diagnostic codes already present in individual patient files, we plan to engineer a reliable risk estimator (**ASD Co-morbid Risk: ACoR**) enabled by novel machine learning algorithms, and compare its effectiveness againsts existing screening tools such as M-CHAT/F. Our rationale is informed by the extensively documented comorbidities of ASD ranging from dysregulation of immune pathways such as eczema, allergies, asthma, as well as ear and respiratory infections, gastrointestinal problems, developmental issues, severe headaches, migraines, and seizures.<sup>[8], [9]</sup> While ASD presentation is highly variable, sophisticated pattern recognition on the longitudinal history of diagnostic codes is expected to reveal uncharted associations that allow precise screening for at-risk patients. Orthogonal to questionnaire based detection of behavioral signals, the proposed tool potentially reduces socio-economic, ethnic and demographic biases to elicit more objective and stable results — with zero administrative burden on clinicians and parents. With a team comprising machine learning experts, pediatric clinicians and developmental experts with extensive history of research and service in the field, the ACoR score can significantly improve outcomes by either boosting sensitivity of current screening or slashing false positives by half.

Thus, the principal aims this study are the following:

- **Aim 1: Reduce false positives in current screening protocols.** The current high false positive rate exacerbates post screen wait-time, increasing diagnostic delay. To evaluate the *Hypothesis: ACoR reduces upto 50% of false positives*, we will track the cases in which MCHAT-F triggers a flag, but ACoR does not. Our aim to evaluate the positive predictive value (PPV) of ACoR, under high specificity conditions (> 95%). Additionally, evaluate if ACoR replicates high sensitivity observed in preliminary studies without losing specificity.
- **Aim 2. Evaluate the statistical relationship between the ACoR score and M/CHAT-F, and formalize a joint or conditional operational protocol.** We will characterize statistical association, if any, between the test scores. *Hypothesis: The uncertainties or errors in the two tests are statistically independent*. Additionally, we will evaluate our ability to boost performance by conditioning the sensitivity-specificity trade-offs on the M-CHAT/F score of individual patients.
- **Aim 3. Evaluate the effectiveness of ACoR in a demographically diverse population with a range of socio-economic confounders.** *Hypothesis: A questionnaire-free approach has the potential to mitigate biases that arise from limitation of language, cultural barriers, and demographic diversity, e.g. disproportionately failing to diagnose children with average to above-average intelligence in diverse populations,<sup>[10]</sup> and under-reporting of symptoms by parents or primary care-givers due to cultural differences.<sup>[11]</sup>*
- **Aim 4. Characterize heterogeneity of ASD presentation by relating it to patterns in medical history, and predictive co-morbidities.** Heterogeneous presentation is a key barrier in the mechanistic understanding of ASD pathobiology. *Hypothesis: We can characterize the distinct classes and/or hierarchies of co-morbidities, by leveraging our ability to disambiguate them from individual medical histories*. This will shed light into the potentially intrinsic classes of the underlying disease processes, and refine/inform intervention design.

Thus, we are proposing to exploit observed co-morbidities in children who ultimately meet the criteria for ASD to develop a risk estimation pipeline, and predict future clinical diagnosis under 2 years of age. Orthogonal to checklists, we aim to reduce the median diagnostic age for ASD, by reducing the long post-screen wait times,<sup>[5]</sup> by significant boosts in positive predictive value, reduction in false positives, and increased sensitivities at little or no loss of specificity, and at no additional administrative burden or resource utilization.

### 3. RESEARCH STRATEGY

**3.1. Significance:** Autism spectrum disorder is a developmental disability associated with significant social, communication, and behavioral challenges. The prevalence of ASD has risen dramatically in the United States from 1 in 10,000 in 1972 to 1 in 59 children in 2014, with males diagnosed at nearly four times the rate of females.<sup>[12], [13]</sup> There is a current lack of consensus on whether increased awareness and recent changes in diagnostic practices<sup>[1]</sup> can fully explain this trend.<sup>[14]</sup> Nevertheless, with possibly over 1% of individuals affected worldwide,<sup>[15]</sup> ASD is a human condition with potentially serious negative impacts on individuals, families, and communities. Early detection can and does improve outcomes,<sup>[1]</sup> and is of paramount importance when designing interventions, and is aligned with the envisioned goal of this initiative, as supported by the National Advisory Mental Health Council (NAMHC) (<https://www.nimh.nih.gov/funding/grant-writing-and-application-process/concept-clearances/2018/early-screening-for-autism-spectrum.shtml>).

Even though ASD may be reliably diagnosed as early as the age of two,<sup>[13]</sup> children frequently remain undiagnosed until after the fourth birthday.<sup>[16]</sup> At this time, there are no laboratory tests for ASD, so a careful review of behavioral history and social interactions is necessary for a clinical diagnosis.<sup>[1], [17]</sup> Starting with being flagged by an initial screen based on standardized checklists presented to parents at the ages between 1.5 and 2 years, a confirmed ASD diagnosis is a multi-step process that very often spans 3 months to 1 year. Most of this time is spent waiting to see qualified providers who can carry out the evaluation necessary for a clinical diagnosis. This extended wait is stressful to families, and impacts patient outcomes by delaying entry into time-critical intervention programs. While lengthy evaluations,<sup>[2]</sup> cost of care,<sup>[3]</sup> lack of providers,<sup>[4]</sup> and lack of comfort in diagnosing ASD by primary care providers<sup>[4]</sup> are all responsible to varying degrees,<sup>[18]</sup> one obvious factor responsible is the number of false positives produced in the initial ASD-specific screening tools in use today. For example, the M-CHAT/F, the most widely used screen,<sup>[1], [6]</sup> produces about over 85 false positives out of every 100 people flagged for further diagnostic evaluation, contributing to extended queues.<sup>[18]</sup> The impact from an excessive number of false positives is exacerbated by the current limited access to care and sparse availability of resources except near urban academic centers.<sup>[18], [19]</sup>

The standardized questionnaires attempt to measure risk by direct observation of behavioral symptoms, as reported by untrained observers (parents). Hence the current screening tests are only as good as the ability of the questions to discern and disambiguate behavior in infants and toddlers on casual observation, and on the ability of parents and caregivers to correctly interpret and answer the items without bias. This has lead to possibility of under-diagnosis in diverse communities as reflected by the lower apparent prevalence among African-American and Hispanic children. Also, children with average or higher-than-average cognitive abilities seem to have been under-diagnosed as reported in large scale population studies.<sup>[1]</sup> Borderline cases are typically problematic to screen for due to the possibility of subjective interpretation that is built into questionnaire based risk assessment. Responses to checklists are clearly confounded by a host of socio-economic (SES) variables, potential interpretive biases, and cultural differences. The heterogeneity of presentation also causes issues, since a potential plurality of symptom classes makes it harder for clinicians to recognize borderline cases, or on-the-fly combine observed co-morbidities with scores from standardized screening tools.

In this study, we operationalize a documented aspect of ASD symptomology in that it has a wide range of co-morbidities occurring at much higher rates than in the general population.<sup>[1]</sup> The ACoR methodology we propose in this grant can address the aforementioned complicated challenges of ASD screening, by distilling incipient predictive patterns of elevated risk from past medical history of individual patients, gleaned by machine learning algorithms from large de-identified databases of retrospective patient records. Powered by novel stochastic learning algorithms, we reverse-engineer sparse noisy uncertain diagnostic code sequences into actionable signatures; in effect giving us a fundamentally new approach to ASD screening and risk evaluation.

**Potential Impact of ACoR In Science and Health:** An automated screening capability deployable at point of care with little or no specialized training, requires no behavioral observations, and is functionally independent of the current tools can transform care. That such predictions of neuropsychiatric disorders might be possible from analyzing historical patterns is suggested by parents of children with ASD often noticing developmental problem very early; while vision and hearing problems are not uncommon in the first year, differences in social, communication, and fine motor skills have been reported to be evident from about 6 months of age.<sup>[20]-[22]</sup>

Abbreviations Used	
ASD	Autism Spectrum Disorder
MCHAT-F	Modified Checklist for Autism in Toddlers with Followup
ADOS / ADOS-2	Autism Diagnostic Observation Schedule
ABA	Applied Behavior Analysis
ACoR	Autism Comorbid Risk

**Significance of Specific Aim 1:** Diagnostic delay partially arises from families waiting in queue for diagnostic evaluations.<sup>[18]</sup> We expect ACoR to be able to cut down the number of false positives significantly, thus reducing the number of children currently flagged for diagnostic evaluation, which potentially cuts down the wait-time, thereby reducing diagnostic delays. Under Specific Aim 1, our goal is to track the fraction of cases where ACoR correctly signals no-risk while MCHAT-F produces a positive flag.

**Significance of Specific Aim 2:** In our preliminary studies, we established that ACoR outperforms M-CHAT/F by considering the average reported performance of M-CHAT/F in a recent study<sup>[23]</sup> on a large cohort with near-universal screening carried out at the Children's Hospital of Philadelphia (CHOP). However, not having observed the ACoR and the M-CHAT/F scores jointly for individual patients, our preliminary studies lack objective assessment of statistical dependence between the two scores. While the nature of the methodologies suggest functional independence, specific Aim 2 will investigate and establish this rigorously.

Functional independence from existing tools implies we can combine the scores; especially leveraging the population stratification induced by the M-CHAT/F scores as reported by the CHOP study to significantly boost combined screening performance. In particular, since patients in the lower M-CHAT/F score bracket have a smaller chance of an ASD diagnosis compared to the high risk upper brackets, we can tailor the sensitivity/specificity trade-offs in the ACoR to maximize either the global PPV or the global sensitivity without losing specificity. Our preliminary results suggest that the expected gains are substantial, with the possibility of doubling the PPV, or increasing the sensitivity by over 50% while keeping the specificity above 95%. Specific aim 2 will investigate the viability of the preliminary results in a pediatric primary care setting.

**Significance of Specific Aim 3:** While still lacking the certainty of a diagnostic blood test, use of subtle patterns emergent in the diagnostic history to estimate risk might help reduce the subjective component in questionnaire-based screening tools, resulting in reduced effect of potential language and cultural barriers in diverse populations.<sup>[1]</sup> With a significant portion of the cohort expected to be African Americans and Hispanics in our primary care clinic, our comparative investigations will be able to explicitly answer these questions.

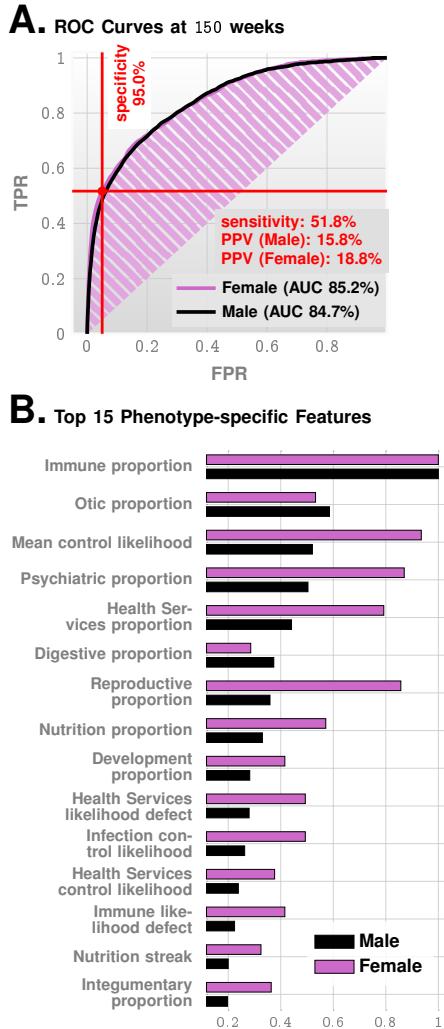
**Significance of Specific Aim 4:** Despite unprecedented advances in charting ASD heritability and genetic burden,<sup>[24], [25]</sup> etiology is still unclear.<sup>[26], [27]</sup> Efforts to identify causal biomarkers have had limited success, with 100 – 1000 genes contributing to ASD risk,<sup>[24], [28]–[30]</sup> and still, genetics have accounted for a limited number of cases.<sup>[31]</sup> Sources of environmental risk are estimated to range from maternal infection and inflammation, diet, and household chemical exposures, to autoimmune conditions and localized perinatal inflammation of the central nervous system.<sup>[26], [27], [32]–[37]</sup> Indeed, heterogeneity of presentation might indicate a plurality of etiologies with converging pathophysiological pathways. Specific Aim 4 will aim to unravel clues to mechanistic drivers by categorizing the heterogeneous presentation by past diagnostic pattern in individuals.

### 3.2. Innovation:

**Paradigm Shift in ASD Screening:** Despite extensive documentation of co-morbidities, a risk estimator that makes reliable predictions for individuals — based purely on co-morbidity patterns — has never been reported to our knowledge. The sparsity of diagnostic codes in individuals, the absence of physiological disorders that would consistently signal the eventual emergence of ASD symptoms, combined with the heterogeneity of ASD presentation, make such an endeavor challenging. In this study we leverage our preliminary work on the formulation of a *first-of-its-kind* framework to make predictions based on models of statistically curated patterns of diagnostic code sequences automatically learned from sufficiently large databases of electronic health records (EHR), that achieves an out-of-sample AUC exceeding 80% for either sex from just over 2 years of age. The machine learning tools that make this possible are also fundamentally novel, designed to address the specific issues in handling sparse, noisy categorical diagnostic sequences. **Most recent ML advances aim to model the physician, nothing ne wis learnt**

**3.3. Approach:** We leverage the ACoR methodology from our preliminary studies, towards prospective application in a primary care setting. We describe the formulation of the ACoR score with a brief description of our preliminary results.

**Source of Electronic Patient Records in Preliminary Studies:** Of the two independent sources of patient records used in our preliminary study, the primary source used to train our models is the Truven Health Analytics MarketScan® Commercial Claims and Encounters Database for the years 2003 to 2012<sup>[38]</sup> (referred to as the Truven dataset). We extracted histories of patients within the age of 0 – 5 years, and excluded



**Fig. 1.** Panel A: ROC curves. Panel B: feature importance inferred by our prediction pipeline. The most import feature is related to immunologic disorders, and we note that in addition to features related to individual disease categories, we also have the mean control likelihood (rank 3), which may be interpreted as the average likelihood of the diagnostic patterns in the control category vs the positive category.

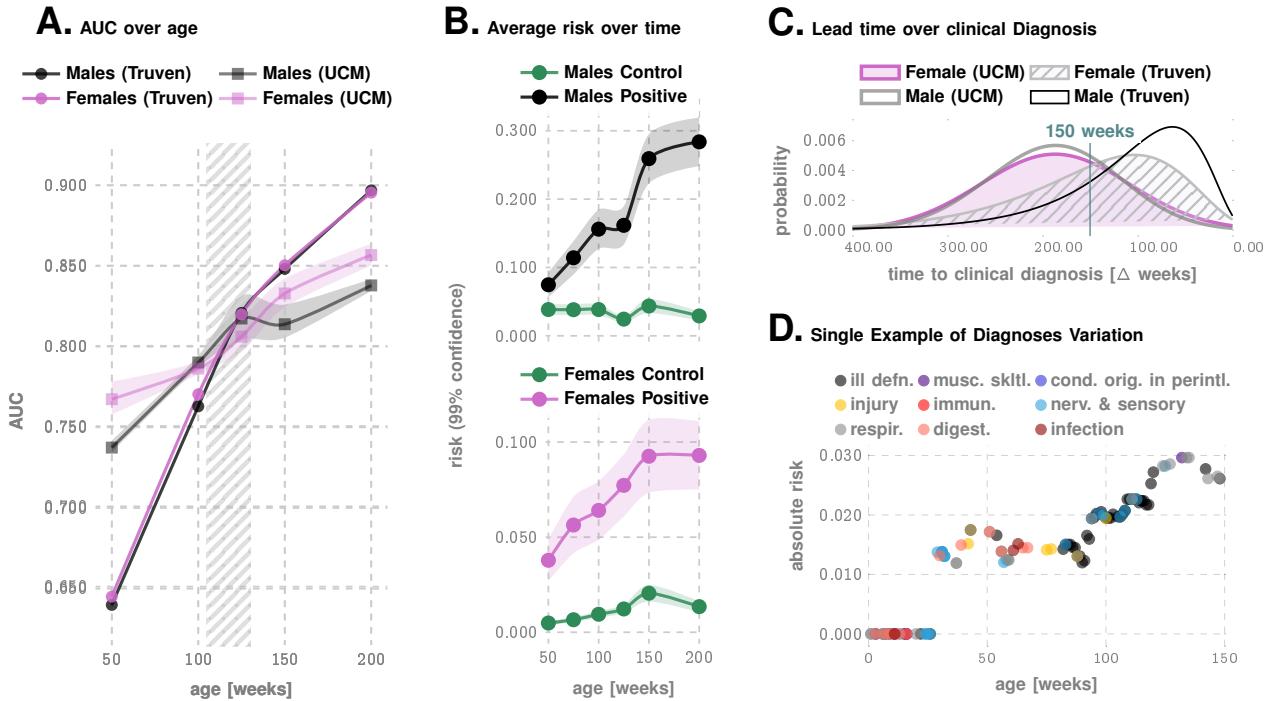
Once we have defined these diagnosis phenotypes, each patient is represented by 51 sparse stochastic time series of events, which are compressed into specialized Hidden Markov Models known as Probabilistic Finite Automata.<sup>[39], [40]</sup> These models are inferred separately for each phenotype, for each sex, and for the control and the positive cohorts, to identify the distinctive average patterns emerging at the population level. Thus, we infer  $51 \times 2 \times 2 = 204$  PFSA models in total in this study. Our inference algorithm assumes no fixed structure, and is able to work with non-synchronized variable length data streams. Variation in either the structure or the parameters of these inferred models across positive and control groups quantify the divergence of comorbidity patterns with increasing risk. Given these models, and the history of a specific patient, we can then quantify the likelihood of this patient belonging to the control models as opposed to the positive group. Calculation of this difference known as the sequence likelihood defect (SLD)<sup>[?]</sup> is the key insight driving our approach. In addition to the phenotype specific Markov models, we use a range of engineered features that reflect various aspects of the patient-specific diagnostic histories, ultimately computing 701 features for each patient. These features are finally used to train a standard gradient boosting classifier<sup>[?]</sup> aiming to map individual patients to a raw risk score. 75% of our patients are randomly selected for training with the rest held-out as a validation set. We measure our performance using standard metrics including the Area Under the receiver-operating characteristic curve (AUC), sensitivity, specificity and the Positive Predictive Value (PPV). We also report accuracy, which is the probability of a correct prediction (positive or control).

patients who do not satisfy the following criteria: 1) At least one code of any available phenotypes is present, 2) Lag between first and last available record for a patient should be at least 15 weeks. These exclusion criteria ensure that we are not considering patients with too few observations to either train on. For training, we analyzed over 4M children ( $n = 4.4M$ ), with 30M diagnostic records (16,649,548 for males and 14,318,303 for females with 9,835 unique diagnostic codes).

While the Truven database is used for both training and out-of-sample cross-validation with held-back patient data, our second independent dataset (referred to as the UCM dataset) consisting of de-identified diagnostic records for children treated at the University of Chicago Medical Center between the years of 2006 to 2018, aids in further cross-validation. We considered children between the ages of 0 – 5 years, and applied the same exclusion criteria as the Truven dataset.

Predicting future ASD diagnosis is a binary classification problem: we classify time-stamped sequences of diagnostic codes into positive and control categories, where the “positive” category refers to patients eventually diagnosed with ASD. Overall we analyze  $n = 729,018$  patients, with 729,018 patients in the positive group and 729,018 patients in the control group (See CONSORT diagram in Fig. ??c), considering approximately 42 million diagnostic codes, with over 46K unique codes in total for both sexes. We do not pre-select any diagnostic code based on its suspected comorbidity with ASD.

**Modeling & Prediction:** The significant diversity of diagnostic codes, along with the sparsity of codes per patient (XX diagnostic per patient on average) makes this a difficult learning problem. We proceed by partitioning the disease spectrum into 51 broad categories, *e.g.* infectious diseases, immunologic disorders, and endocrinial disorders. Some of these categories comprises a relatively large number of diagnostic codes aligning roughly with the categories defined within the ICD framework.<sup>[?]</sup> The remaining categories represent groups of one or more codes that might have some known or suspected association with pulmonary disorders. Each of the diagnostic categories yield a single time series over weeks (each week being identified as having a value ‘0’ for no code corresponding to the diagnostic category, or ‘1’ if some code is present, and ‘2’ if a diagnostic code from any of the other categories is present). We refer to the individual diagnostic categories as a broad phenotype.



**Fig. 2.** Variation of Inferred Risk. Panel A illustrates AUC achieved as a function of patient age, for the Truven and UCM datasets: we achieve  $> 80\%$  AUC for either gender from shortly after 2 years. Panel B illustrates risk variation with time for the control and the positive cohorts. Panel C shows the distribution of the prediction horizon: the time to a clinical diagnosis after inferred relative risk crosses 90%. Panel d illustrates the risk progression of a specific, ultimately autistic male child in the Truven database.

Beyond the demonstrated predictive performance, calculation of the ACoR score offers insights into the ADRD comorbidities that have predictive value. Estimating the relative importance of the features used is crucial for sanity checks, as well as for insights into the underlying causal mechanisms. We compute the relative importance of the features by estimating the mean change in the raw risk via random perturbation of a particular feature: this is the “feature importance” shown in Fig. ??c for the different diagnostic categories, which illustrates that cardio-vascular, metabolic, ophthalmological disorders as important diagnostic categories (expected) modulating the ACoR score, along with contributions from all across the disease spectrum.

Our features are based on data already available in the past medical records. We do not demand results from specific tests, or look for specific demographic, bio-molecular, physiological and other parameters; we use what we get in the diagnostic history of patients, which presents un-structured sequence of labels pertaining to the ICD and the prescription codes, and is typically prone to noise, coding errors and sparsity. Our ability to effectively work with uncurated data and achieve high out-of-sample predictive performance showcases the immediate clinical applicability with zero additional burden to patients and providers.

TABLE 1

Standalone ACoR performance (For Comparison  
M-CHAT/F Performance:  
sensitivity=38.8%, specificity=95%, PPV=14.6%  
between within  $\approx 112$  weeks)

week	spec.	sens.	PPV	sex	dataset
100	0.92	0.39	0.14	F	UCM
100	0.95	0.39	0.19	M	UCM
100	0.93	0.39	0.13	F	Truven
100	0.91	0.39	0.10	M	Truven
112	0.93	0.39	0.16	F	UCM
112	0.95	0.39	0.20	M	UCM
112	0.96	0.39	0.22	F	Truven
112	0.95	0.39	0.17	M	Truven
150	0.94	0.39	0.19	F	UCM
150	0.98	0.39	0.34	F	Truven
150	0.97	0.39	0.26	M	Truven
150	0.97	0.39	0.26	M	UCM

**Standalone Predictive Performance:** The standalone performance in preliminary studies is summarized in Figs. 1 and 2. We achieve an out-of-sample AUC of 82.3% for males and 82.5% for females at 125 weeks of age for the Truven dataset. In the UCM dataset, our performance is comparable: 83.1% and 81.3% for males and females respectively at 125 weeks of age. The good agreement of the out-of-sample performance on these independent datasets lends strong support for our claims. The specificity, sensitivity, PPV trade-offs are shown in Table 1. We enumerate the top 15 predictive features in Fig. 1B. We also computed the county-specific performance of the risk pipeline for the Truven dataset, and we got nearly uniform performance across the country for both genders. Fig. 2A illustrates the variation of the AUC with increasing age of the subjects plotted with 99% confidence bounds: the increase is very nearly linear, with a change of gradient near the 150 week mark. We find that while the AUC gradients are different in the two datasets,

they tend to match up in later ages. The differences in the early ages are possibly due to differences in patient statistics: a larger number of patients in Truven at the earlier ages with a relatively smaller number of observations on average.

**TABLE 2**  
Boosted Sensitivity, specificity and PPV at 26 months with ACoR  
Conditioned on M-CHAT/F Scores

M-CHAT/F Outcome				perf. (Truven)			perf. (UCM)		
0-2 NEG	3-7 NEG	3-7 POS	> 8 POS	speci-ficity	sensi-tivity	PPV	speci-ficity	sensi-tivity	PPV
specificity choices									
0.48	0.87	0.97	0.99	0.98	0.432	0.331	0.98	0.355	0.289
0.38	0.54	0.94	0.98	0.95	0.736	0.203	0.95	0.628	0.178

We plot the absolute or raw risk over time for males and females for the out-of-sample control and positive cohorts in Fig. 2B. Notably, in these risk plots, averaged over the population, we see disambiguation early, right from 50 weeks. Also, we see a saturation of the risk after 150 weeks, which corresponds to the median diagnosis age in the database (approx. 150 weeks).

Guthrie *et al.*<sup>[23]</sup> from Children's Hospital of

Philadelphia (CHOP) has demonstrated that as a nearly universal screening tool (n=20,375) M-CHAT/F has a sensitivity of 38.8%, specificity of 94.9% and PPV of 14.6%. Comparing the performance metrics at different age groups and sexes (See Table 1), we conclude that our approach produces a strictly superior PPV (exceeding M-CHAT/F PPV by at 14% (14.1-33.6%) when sensitivity and specificity are held at comparable values around the age of 26 months ( $\approx$  112 weeks).

In this study, we leverage the population stratification induced by an existing independent screening test (M-CHAT/F) to improve combined performance. Here a combination refers to the conditional choice of the sensitivity/specificity trade-offs for our tool in each sub-population such that the overall performance is optimized with respect to whether we wish to maximize the PPV or the sensitivity at a specified minimum level of specificity. This ultimately yields an overall performance significantly superior to M-CHAT/F alone, with a PPV close to or exceeding 30% across datasets ( $>$  33% fro Truven,  $>$  28% for UCM), or a sensitivity close to or exceeding 50% for the high recall (HR) operating point ( $>$  58% for Truven,  $>$  50% for UCM), when we restrict specificities to above 95% (See Table 2). Importantly, designing rules for conditional operation only require average population characteristics, *i.e.*, an estimate of ASD prevalence in the sub-populations defined by the relevant brackets of M-CHAT/F scores, and the prevalence of these score brackets in the general population.

Our predictive ability is predicated on the diagnostic history of individual patients being not random, hiding signatures to future neuropsychiatric outcomes. A single random patient from the Truven database is illustrated in Fig. 2D, revealing that infections and immunological disorders are experienced early to a much higher degree compared to other diseases, and diseases of the nervous system and sensory organs, as well as ill-defined symptoms dominate the latter period. This suggests deeper interrogation of the structure of co-morbid patterns (Specific Aim 4). In our preliminary studies, we found excellent disambiguation of ASD from other intellectual disabilities, which we would further evaluate.

**Research Design:** ACoR methodology can screen instantaneously every child in primary care, for whom past medical history is available, with zero administrative and resource burden. To achieve the specific goals outlined in the specific aims of this study, we will gather data from both the child and the primary caregiver of all patients in the participating UCM primary pediatric primary care clinic, and test the sensitivity and specificity of ACoR against the most commonly used screening tool M-CHAT/F, by an ADOS-2 based (near) gold-standard evaluation of patients. Using these data and the inferred statistical dependency (or the lack thereof) properties between the screening tools, ACoR can tailor the selection of sensitivity/specificity trade-offs to the particular informant and to the age of the child, with the view to optimizing global characteristics such as maximizing the PPV or the sensitivity of the screening process. Additionally, the ACoR algorithm will identify categories of heterogeneity that can lead to mechanistic insights into ASD pathobiology.

The key steps in this project are as follows (See Fig. 3):

- 1. Pediatric clinic team will administer M-CHAT/F to incoming children with 16-30 months.
- 2. The University of Chicago Research Informatics Support team will work with the PI and his team to compute the ACoR score corresponding to individual consenting patients
- 3. On being flagged by M-CHAT/F as high risk, or if the ACoR score indicates high risk and teh M-CHAT/F is borderline, the patients will be scheduled for ADOS-2 evaluation overseen by Dr. Smith and his team

- 4. The evaluation scores will be analyzed by the PI and his team to address the specific aims.

The limited scope of this project implies that we need to be aware of the number of ADOS-2 referrals generated due to ACoR, particularly since ADOS-2 evaluations involve significant resource and cost.

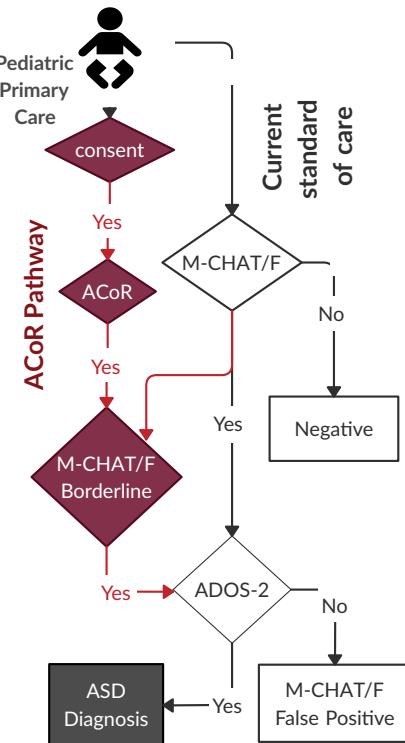


Fig. 3. Patient processing

M-CHAT/F and hence schedule some children for ADOS, who do not have autism, and might cause some stress in parents and families if they are . The potential societal benefit gained in lieu of this discomfort is the validation of the expected performance boost for ASD screening at the population level PPV by up to 100%, or the sensitivity by 50%.

**Procedures:** Eligible patients at the Department of Pediatrics, University of Chicago (patients who present for a well-child visit or any other non-emergency reason) will be asked for consent for access to their past medical history for carrying out the ACoR screen. If there is a flag either in M-CHAT/F or in ACoR, the attending pediatrician will inform parents of a potential elevated risk of ASD diagnosis, and offer to schedule for an ADOS-2 evaluation. The ADOS-2 evaluation triggered by ACoR flags will be at no cost to the patient. All study procedures and consent forms will be approved by the University of Chicago Institutional Review Board. For all assessments, basic demographic information, recruitment site, medications and diagnoses assigned by the current clinical treatment team, will be obtained from the parent/caregiver and medical record.

**Data Management:** Data collection forms for demographic and clinical history data, database design and data management procedures will be designed, created and conducted at the University of Chicago under the direction of Dr. Smith and Prof. Chattopadhyay. Demographic and clinical history data will be collected and entered into an HIPAA compliant secure database. Monthly reports will be generated to monitor data timeliness, completeness, and accuracy as well as subject flow through the study.

**Cohort Selection:** Participants will be approximately 5000 children per year (producing approximately 300 ADOS-2 referrals) who will be evaluated via both the MCHAT-F screening during wellness visits at the 1 year, 1.5 year and 2 year mark, via the standard questionnaire completed by their primary caregivers, and the ACoR algorithm applied to their diagnostic history on file. Inclusion criteria: 1) Child is between 16 and 30 months, 2) Child has diagnostic history on record with at least 5 diagnostic codes, and the first code is at least from 15 weeks in the past. Exclusion criteria: Diagnostic history only consists of health service contact codes.

**3.4. Study Interventions and Measures:** No intervention is planned, with main outcomes being efficacy and applicability of ACoR compared to MCHAT-F.

**Risk To Patients:** The design of the study guarantees that patients suffer no negative impact from the added ACoR screen. Indeed, pursuant to available resources, patients who are flagged will be expedited for ADOS evaluation which eliminates or reduces their wait-times. For some borderline cases, which would have been missed by M-CHAT/F, might get flagged by ACoR, and be scheduled for ADOS, which they would not have had to do with just M-CHAT/F. But this is a positive outcome. Only when the M-CHAT/F is borderline, and the ACoR signifies a high risk, may a patient be scheduled for ADOS-2 who otherwise will not have the referral. Thus there is a small possibility that ACoR might have some false positives that are different from that of

## REFERENCES

- [1] Hyman, S. L., Levy, S. E., Myers, S. M. et al. Identification, evaluation, and management of children with autism spectrum disorder. *Pediatrics* **145** (2020).
- [2] Kalb, L. G. et al. Determinants of appointment absenteeism at an outpatient pediatric autism clinic. *Journal of Developmental & Behavioral Pediatrics* **33**, 685–697 (2012).
- [3] Bisgaier, J., Levinson, D., Cutts, D. B. & Rhodes, K. V. Access to autism evaluation appointments with developmental-behavioral and neurodevelopmental subspecialists. *Archives of pediatrics & adolescent medicine* **165**, 673–674 (2011).
- [4] Fenikilé, T. S., Ellerbeck, K., Filippi, M. K. & Daley, C. M. Barriers to autism screening in family medicine practice: a qualitative study. *Primary health care research & development* **16**, 356–366 (2015).
- [5] Gordon-Lipkin, E., Foster, J. & Peacock, G. Whittling Down the Wait Time: Exploring Models to Minimize the Delay from Initial Concern to Diagnosis and Treatment of Autism Spectrum Disorder. *Pediatr. Clin. North Am.* **63**, 851–859 (2016).
- [6] Robins, D. L. et al. Validation of the modified checklist for autism in toddlers, revised with follow-up (m-chat-r/f). *Pediatrics* **133**, 37–45 (2014).
- [7] Jashar, D. T., Brennan, L. A., Barton, M. L. & Fein, D. Cognitive and adaptive skills in toddlers who meet criteria for autism in dsm-iv but not dsm-5. *Journal of autism and developmental disorders* **46**, 3667–3677 (2016).
- [8] Tye, C., Runicles, A. K., Whitehouse, A. J. O. & Alvares, G. A. Characterizing the Interplay Between Autism Spectrum Disorder and Comorbid Medical Conditions: An Integrative Review. *Front Psychiatry* **9**, 751 (2018).
- [9] Kohane, I. S. et al. The co-morbidity burden of children and young adults with autism spectrum disorders. *PLoS ONE* **7**, e33224 (2012).
- [10] Christensen, D. L. et al. Prevalence and characteristics of autism spectrum disorder among children aged 4 years—early autism and developmental disabilities monitoring network, seven sites, united states, 2010, 2012, and 2014. *MMWR Surveillance Summaries* **68**, 1 (2019).
- [11] Burkett, K., Morris, E., Manning-Courtney, P., Anthony, J. & Shambley-Ebron, D. African american families on autism diagnosis and treatment: The influence of culture. *Journal of Autism and Developmental Disorders* **45**, 3244–3254 (2015).
- [12] Baio, J. et al. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *MMWR Surveill Summ* **67**, 1–23 (2018).
- [13] Data & statistics on autism spectrum disorder — cdc (2019). URL <https://www.cdc.gov/nccbdd/autism/data.html>.
- [14] King, M. & Bearman, P. Diagnostic change and the increased prevalence of autism. *Int J Epidemiol* **38**, 1224–1234 (2009).
- [15] Elsabbagh, M. et al. Global prevalence of autism and other pervasive developmental disorders. *Autism Res* **5**, 160–179 (2012).
- [16] Schieve, L. A. et al. Population attributable fractions for three perinatal risk factors for autism spectrum disorders, 2002 and 2008 autism and developmental disabilities monitoring network. *Ann Epidemiol* **24**, 260–266 (2014).
- [17] Volkmar, F. et al. Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry* **53**, 237–257 (2014).
- [18] Gordon-Lipkin, E., Foster, J. & Peacock, G. Whittling down the wait time: exploring models to minimize the delay from initial concern to diagnosis and treatment of autism spectrum disorder. *Pediatric Clinics* **63**, 851–859 (2016).
- [19] Althouse, L. A. & Stockman, J. A. Pediatric workforce: A look at pediatric nephrology data from the american board of pediatrics. *The Journal of pediatrics* **148**, 575–576 (2006).
- [20] Kozlowski, A. M., Matson, J. L., Horovitz, M., Worley, J. A. & Neal, D. Parents' first concerns of their child's development in toddlers with autism spectrum disorders. *Dev Neurorehabil* **14**, 72–78 (2011).
- [21] Herlihy, L., Knoch, K., Vibert, B. & Fein, D. Parents' first concerns about toddlers with autism spectrum disorder: Effect of sibling status. *Autism* **19**, 20–28 (2015).
- [22] Chawarska, K., Macari, S. & Shic, F. Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. *Biological psychiatry* **74**, 195–203 (2013).
- [23] Guthrie, W. et al. Accuracy of Autism Screening in a Large Pediatric Network. *Pediatrics* **144** (2019).
- [24] Satterstrom, F. K. et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *bioRxiv* (2019). <https://www.biorxiv.org/content/early/2019/04/24/484113.full.pdf>.
- [25] Sandin, S. et al. The Heritability of Autism Spectrum DisorderReassessing the Heritability of Autism Spectrum DisordersLetters. *JAMA* **318**, 1182–1184 (2017). URL <https://doi.org/10.1001/jama.2017.12141>. [https://jamanetwork.com/journals/jama/articlepdf/2654804/jama\\_sandin\\_2017\\_id\\_170037.pdf](https://jamanetwork.com/journals/jama/articlepdf/2654804/jama_sandin_2017_id_170037.pdf).
- [26] Ohja, K. et al. Neuroimmunologic and Neurotrophic Interactions in Autism Spectrum Disorders: Relationship to Neuroinflammation. *Neuromolecular Med.* **20**, 161–173 (2018).
- [27] Gadysz, D., Krzywdziska, A. & Hozyasz, K. K. Immune Abnormalities in Autism Spectrum Disorder-Could They Hold Promise for Causative Treatment? *Mol. Neurobiol.* **55**, 6387–6435 (2018).
- [28] Sanders, S. J. et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215–1233 (2015).
- [29] Gaugler, T. et al. Most genetic risk for autism resides with common variation. *Nat. Genet.* **46**, 881–885 (2014).
- [30] Werling, D. M. The role of sex-differential biology in risk for autism spectrum disorder. *Biol Sex Differ* **7**, 58 (2016).
- [31] Abrahams, B. S. & Geschwind, D. H. Advances in autism genetics: on the threshold of a new neurobiology. *Nat. Rev. Genet.* **9**, 341–355 (2008).
- [32] Yamashita, Y. et al. Anti-inflammatory Effect of Ghrelin in Lymphoblastoid Cell Lines From Children With Autism Spectrum Disorder. *Front Psychiatry* **10**, 152 (2019).
- [33] Shen, L. et al. Proteomics Study of Peripheral Blood Mononuclear Cells (PBMCs) in Autistic Children. *Front Cell Neurosci* **13**, 105 (2019).
- [34] Theoharides, T. C., Tsilioni, I., Patel, A. B. & Doyle, R. Atopic diseases and inflammation of the brain in the pathogenesis of autism spectrum disorders. *Transl Psychiatry* **6**, e844 (2016).
- [35] Young, A. M. et al. From molecules to neural morphology: understanding neuroinflammation in autism spectrum condition. *Mol Autism* **7**, 9 (2016).
- [36] Croen, L. A. et al. Family history of immune conditions and autism spectrum and developmental disorders: Findings from the study to explore early development. *Autism Res* **12**, 123–135 (2019).
- [37] Zerbo, O. et al. Immune mediated conditions in autism spectrum disorders. *Brain Behav. Immun.* **46**, 232–236 (2015).

- [38] Hansen, L. The truven health marketscan databases for life sciences researchers. *Truven Health Ananlytics IBM Watson Health* (2017).
- [39] Chattopadhyay, I. & Lipson, H. Abductive learning of quantized stochastic processes with probabilistic finite automata. *Philos Trans A* **371**, 20110543 (2013). URL <http://rsta.royalsocietypublishing.org/content/371/1984/20110543.full.pdf>.
- [40] Chattopadhyay, I. & Lipson, H. Data smashing: uncovering lurking order in data. *Journal of The Royal Society Interface* **11** (2014). URL <http://rsif.royalsocietypublishing.org/content/11/101/20140826>.