National Institute of Diabetes and Digestive and Kidney Diseases

Central Repository

NIDDK-CR Resources for Research

Data Science and Meet the Expert Webinar Series

August 28, 2025

National Institute of Diabetes and Digestive and Kidney Diseases

Central Repository

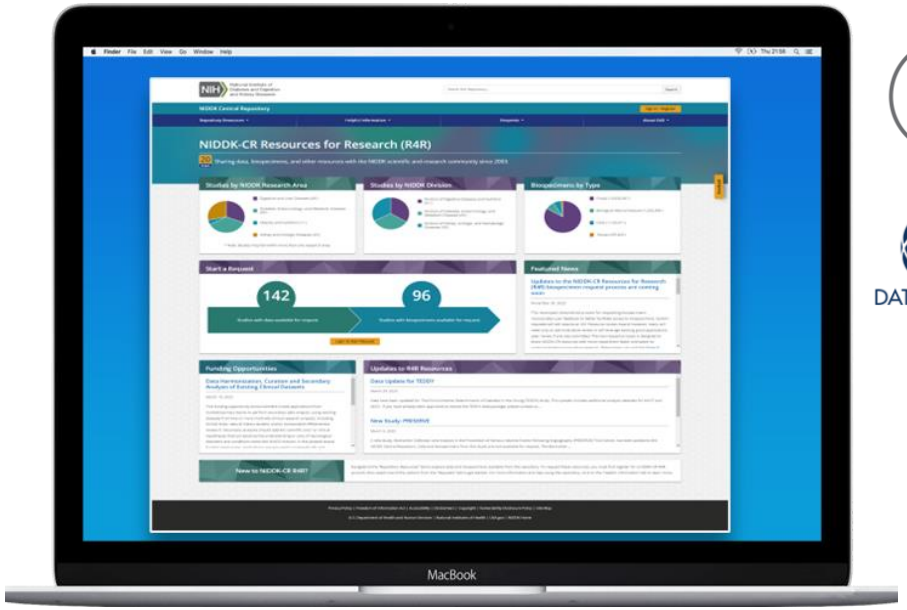
NIDDK Central Repository Overview

Mission

Established in 2003 to **facilitate sharing of data, biospecimens, and other resources** generated from studies supported by NIDDK and within NIDDK's mission by making these **resources available for request to the broader scientific and research community**.

- Supports receipt and distribution of data and biospecimens in a manner that is ethical, equitable, and efficient
- Enables investigators not involved with the original work to test new hypotheses without the need to collect new data or biospecimens
- Promotes FAIR (Findable, Accessible, Interoperable, and Reusable) and TRUST (Transparency, Responsibility, User focus, Sustainability, and Technology) principles

Recorded past tutorials, webinars, and other educational resources can be found on the NIDDK-CR website



CORE TRUST SEAL

WORLD DATA SYSTEM

Imaging Data Files
15.8 M

Clinical Datasets
>8,600
from 194 clinical studies

Biospecimens
>16 M

Registered Users
7,026

Weekly Users
>5,000

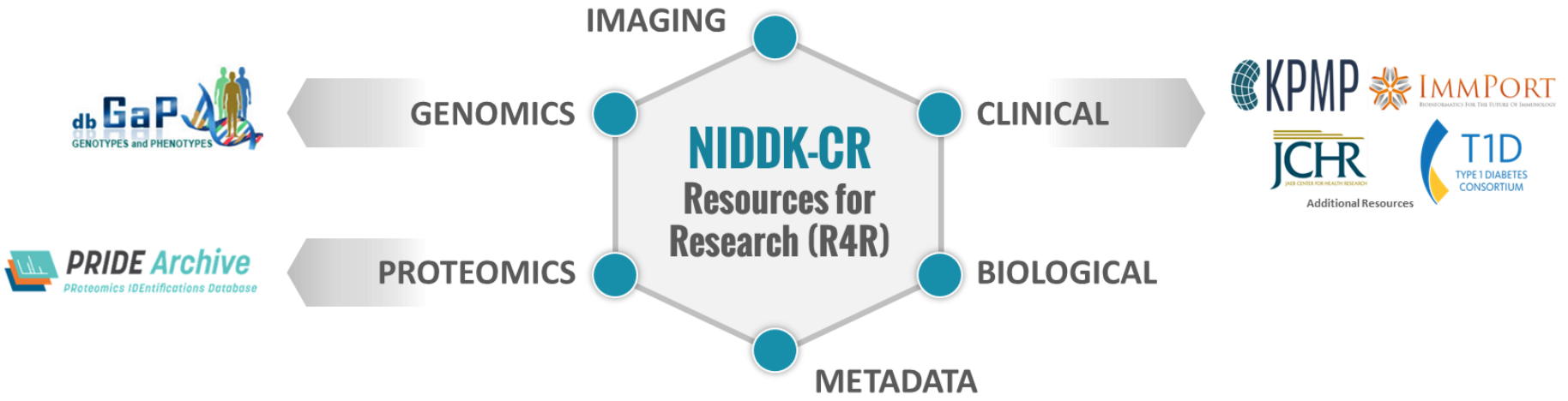
Public Releases
>875

National Institute of Diabetes and Digestive and Kidney Diseases

Central Repository

NIDDK Data Sharing Ecosystem

The NIDDK-CR is a part of the broader NIH-funded biomedical data ecosystem and plays a key role in NIH's FAIRness and TRUSTworthiness goals. The NIDDK-CR houses a broad range of data types for secondary research, provides access to biospecimens, and direct links to other repositories with additional resources such as genomics data.



FAIRsharing.org standards, databases, policies

DataCite FIND, ACCESS, AND REUSE DATA

re3data.org REGISTRY OF RESEARCH DATA REPOSITORIES

DK net

Google Dataset Search / Schema.org

NIH U.S. National Library of Medicine ClinicalTrials.gov

Vivli CENTER FOR GLOBAL CLINICAL RESEARCH DATA

PLANNING PHASE

figshare

NIH HEAL INITIATIVE

National Institute of Diabetes and Digestive and Kidney Diseases

Central Repository

Future Functionality: Analytics Workbench

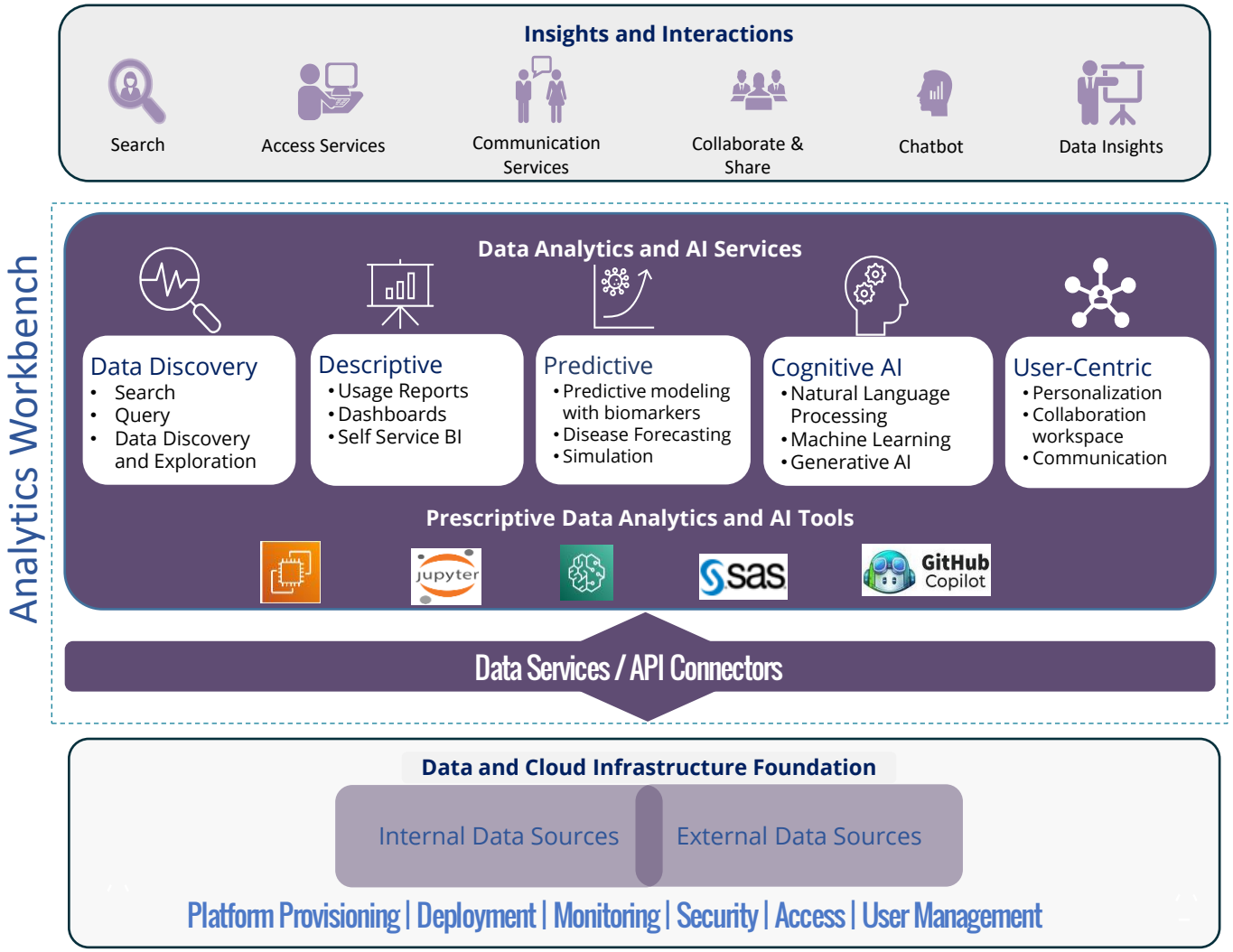
Streamlining *end-to-end data science lifecycle and discovery of data-driven biomedical insights.*

Innovation and ease of use

A cloud-based analytics environment where researchers and data scientists can access a suite of integrated analytics tools and cloud computing resources to participate in data challenges and AI innovation.

Expected Benefits of Analytics Workbench:

- **Promote Collaboration**
- **Support AI Innovation**
- **Minimize Data Movement**
- **Improve User Experience**
- **Discover Data Insights**
- **Advance NIDDK Research Mission**



National Institute of Diabetes and Digestive and Kidney Diseases

Central Repository

NIDDK-CR Data Science Centric Challenge Series

Goals of NIDDK-CR Data-science centric challenge series

- Develop tools, approaches, models and/or methods to increase data interoperability and usability for artificial intelligence (AI) and machine learning (ML) applications
- Augment and enhance existing data for future secondary research, including data-driven discovery by AI/ML researchers
- Discover innovative approaches to enhance the utility of datasets for AI/ML applications

Visit our website for more information on our data-centric movement and to learn more about our past data-challenges

National Institute of Diabetes and Digestive and Kidney Diseases

Central Repository

Secondary Data Science and Meet the Expert Webinar Series

About the Series

- Aims to accelerate data science and AI-driven biomedical research by fostering collaboration between biomedical researchers and experts in the field
- Monthly webinar held on the **last Thursday of each month**

Upcoming Webinars

- Today – Impact and Innovations from use of NIDDK-CR Resources
- September 25th - Building Real-World Data (RWD) Linkages with a Focus on Quality and Reusability

Learn more about the webinar series, register for future webinars, and access past webinars materials and recordings

National Institute of Diabetes and Digestive and Kidney Diseases

Central Repository

Meet the Experts

Dr. Adam E Gaweda is an Associate Professor in the Department of Medicine / Division of Nephrology and Hypertension, University of Louisville, Louisville KY USA. He earned his Master's of Engineering degree in Electrical Engineering from Czestochowa University of Technology (Poland) in 1997 and his PhD degree in Computer Science and Engineering from University of Louisville in 2002. He also holds a Master's Degree in Clinical Investigative Sciences (University of Louisville, 2009). Dr. Gaweda's doctoral research focused on explainable data-driven Artificial Intelligence and Machine Learning models. After joining the UofL School of Medicine in 2002, Dr. Gaweda collaborated with Drs. Aronoff, Brier, and Jacobs, on application of AI/ML methods to personalized therapy of patients with Chronic Kidney Disease. This research resulted in one of the first AI-guided Clinical Decision Support Systems for managing anemia in dialysis patients, which has been clinically implemented at the UofL dialysis unit since 2013 and is now commercially available in the US. Dr. Gaweda's current research uses innovative AI/ML approaches to enhance the understanding of pathologic processes and facilitate discovery of new therapeutic pathways for complex diseases.

National Institute of Diabetes and Digestive and Kidney Diseases

Central Repository

Meet the Experts

Dr. Juliet Emamaullee is an Associate Professor of Surgery and Immunology (Clinical Scholar) at the University of Southern California Keck School of Medicine and an attending liver and kidney transplant surgeon at Keck Hospital and Children's Hospital-Los Angeles. She is also the Associate Chief, Division of Clinical Research, Department of Surgery. Dr. Emamaullee completed her PhD and MD degrees at the University of Alberta, followed by residency training in general surgery at Emory University and an abdominal organ transplant/HPB surgery fellowship at the University of Alberta. She is a surgeon-scientist with an NIH-funded translational immunobiology lab, exploring immunological phenotypes associated with liver transplant recipients. Dr. Emamaullee's areas of expertise include computational biology, Fontan-associated liver disease, and living donor liver transplantation.

National Institute of Diabetes and Digestive and Kidney Diseases

Central Repository

Meet the Experts

Dr. Prasanna Santhanam is an Associate Professor of Clinical Medicine and Oncology at Johns Hopkins University School of Medicine, where he has been a staff physician since 2017. He specializes in the diagnosis and treatment of endocrine disorders with a particular focus on thyroidology, thyroid neoplasms, and metabolic bone disease.

Dr. Santhanam earned his MBBS from BJ Medical College and his Doctor of Medicine from NHL Municipal Medical College in Gujarat, India. Following his doctoral studies, he completed residencies in both Medicine and Therapeutics and Internal Medicine, further honing his expertise.

An active contributor to advancements in his field, Dr. Santhanam has participated in several committees, including the American Thyroid Association (ATA) and is involved with the American Board of Internal Medicine (ABIM) as a member of the Endocrinology, Diabetes, and Metabolism Item-Writing Task Force. His extensive research contributions are evidenced by numerous high-impact publications and successful grant applications, establishing him as a leader in endocrine research, particularly in biochemical, metabolic, and molecular imaging related to metabolic syndrome, thyroid disease, and neuroendocrine tumors.

Additionally, Dr. Santhanam serves as Co-founder of AI-Metab, LLC, a company focused on developing AI-based solutions for research and analysis in endocrinology, particularly in metabolic imaging and cardiometabolic health.

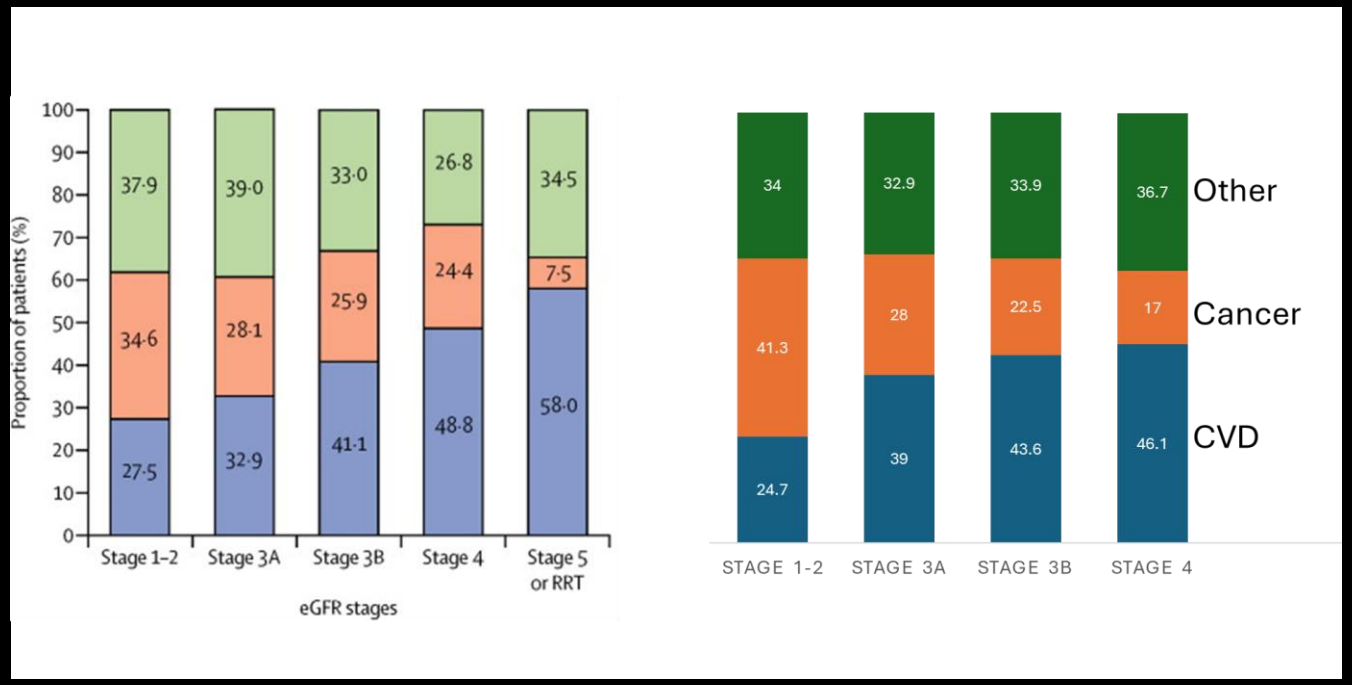
OPENING THE BLACK BOX: TOWARD UNDERSTANDING PATHOPHYSIOLOGY OF UREMIC VASCULOPATHY USING EXPLAINABLE AI

Adam E Gaweda, PhD
Associate Professor
Department of Medicine
University of Louisville
Louisville KY

OUTLINE

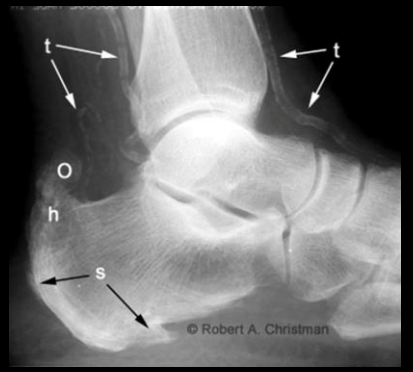
- Uremic Vasculopathy in Chronic Kidney Disease
- Data Scientific Challenges in Biomedical Computing
- Solving Data Scientific Challenges Using Human Inspired Computing

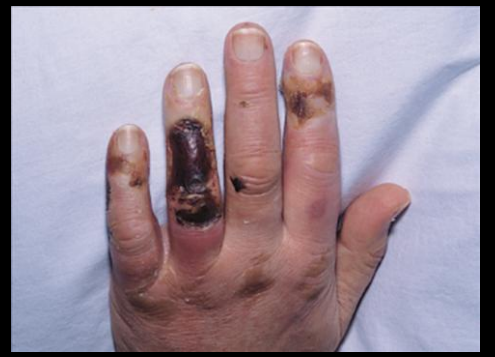
UREMIC VASCULOPATHY



Gansevoort, Lancet, 2013

Thompson, JASN, 2015





UREMIC VASCULOPATHY

- Evolutionary development of progressive endothelial, smooth muscle, and connective tissue dysfunction that lays the groundwork for eventual medial calcification.

- Early stage uremic vasculopathy
  - Lipid metabolism
  - Inflammatory processes
  - Endothelial dysfunction

- Late stage uremic vasculopathy
  - Mineral metabolism

UREMIC VASCULOPATHY

Effective diagnosis, treatment, and prevention of uremic vasculopathy requires better understanding of the pathophysiologic processes behind it and their contributions at different stages of Chronic Kidney Disease.

DATA SCIENTIFIC CHALLENGES

- Complex, multifactorial nature of physiologic processes
- High-dimensional data:
  - Increased likelihood of missing data
- Interactions between (groups of) parameters
  - Negative feedback loops

WHY ARTIFICIAL INTELLIGENCE

Translating human problem-solving and cognitive skills into the digital domain

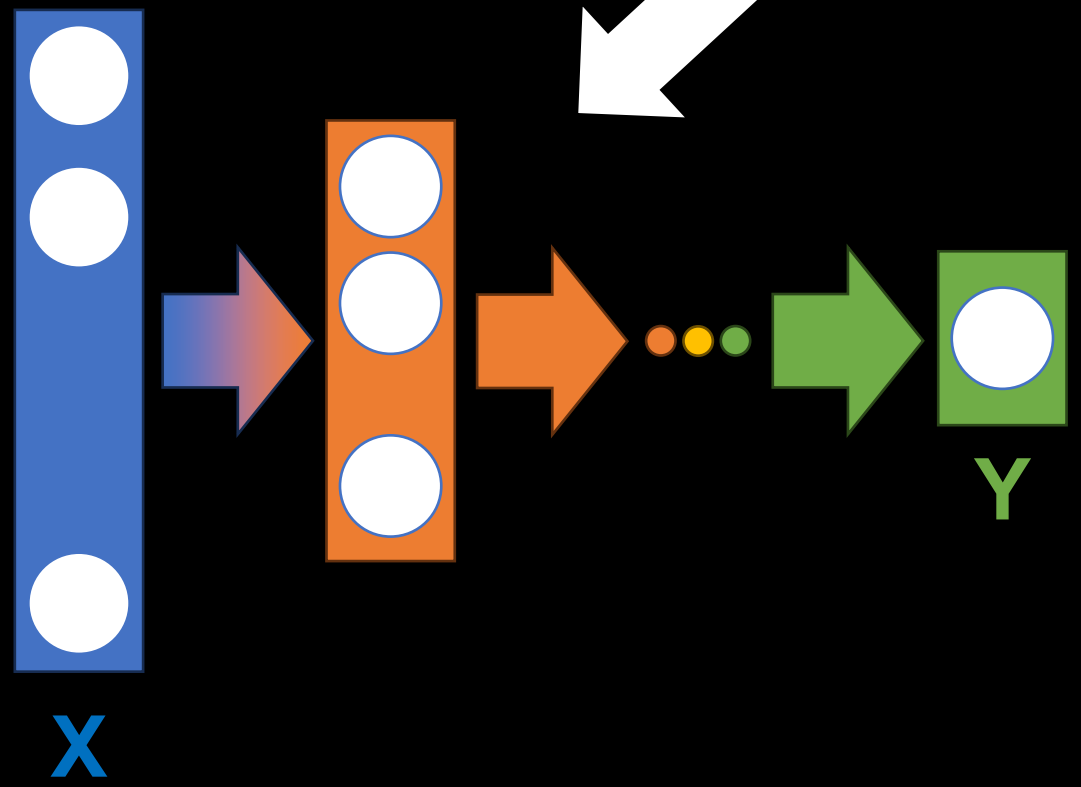
- Learning / Adaptation
- Pattern Recognition and Abstraction / Reasoning
- Planning / Decision Making

MACHINE LEARNING

- Supervised Learning
- Unsupervised Learning
- Self-Supervised Learning

SUPERVISED LEARNING

GIVEN X AND Y:
FIND MAPPING FROM X TO Y



Linear Regression / Logistic Regression

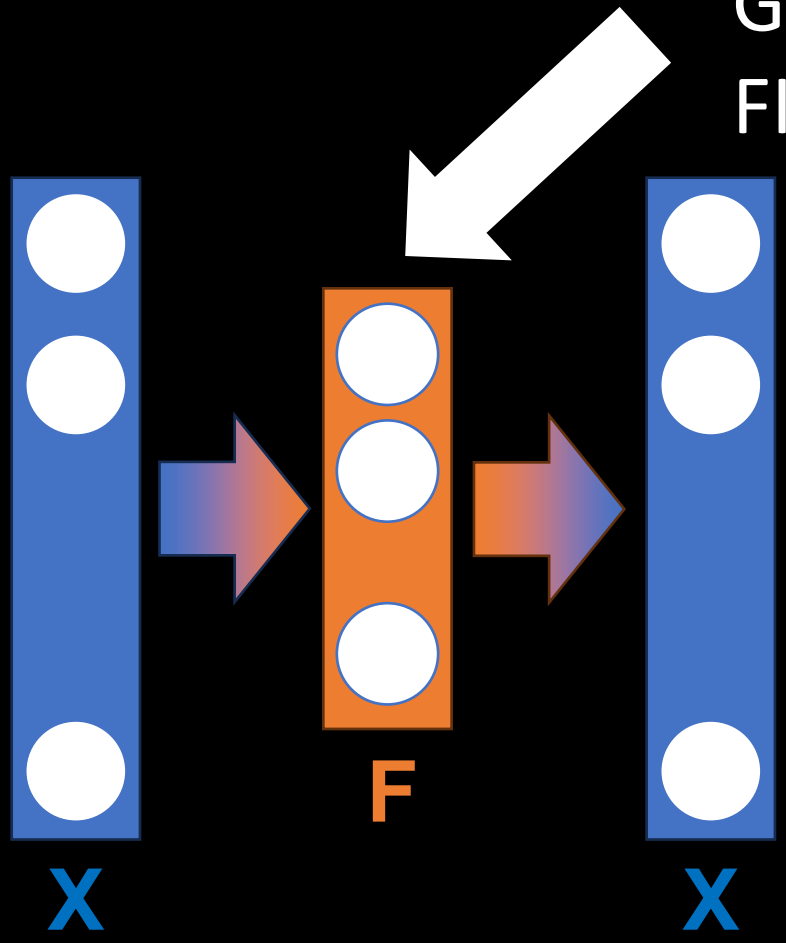
Classification and Regression Trees

Bayesian Networks

(Deep) Neural Networks

UNSUPERVISED LEARNING

GIVEN SIMILARITY METRIC: FIND PATTERNS

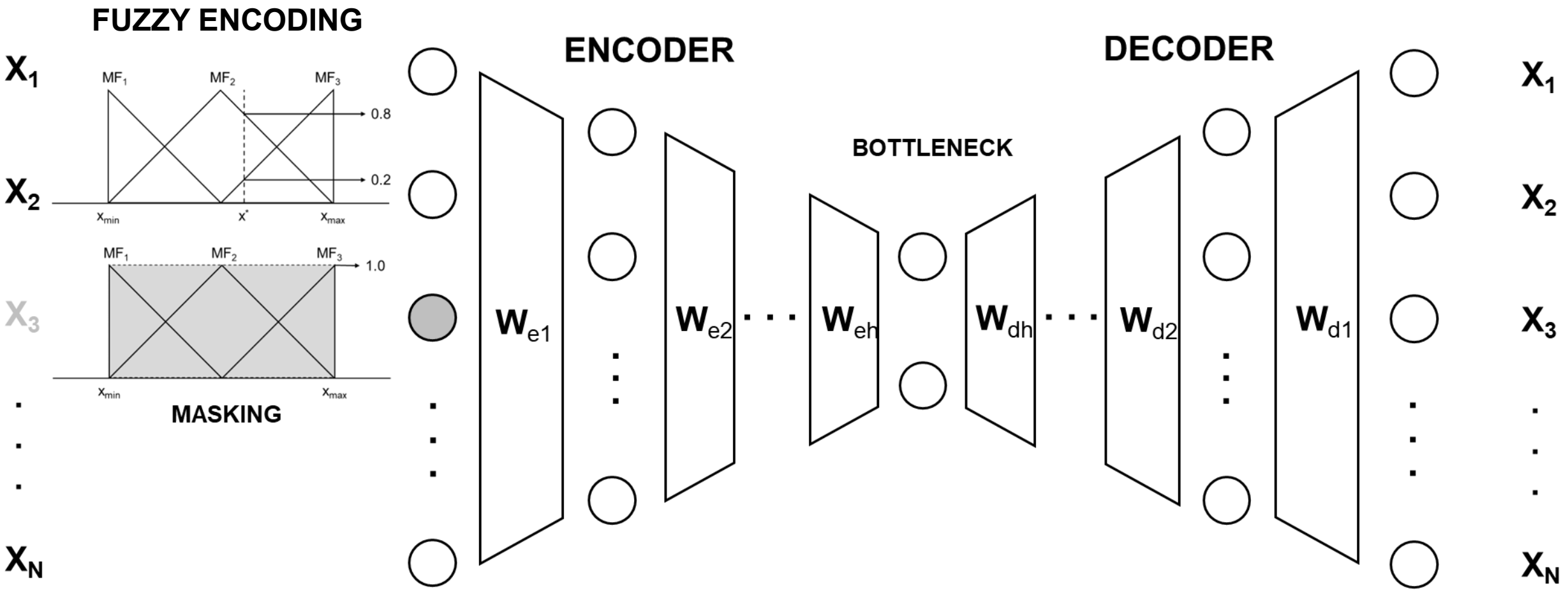


Clustering
(Hierarchical, K-means, Fuzzy C-means)

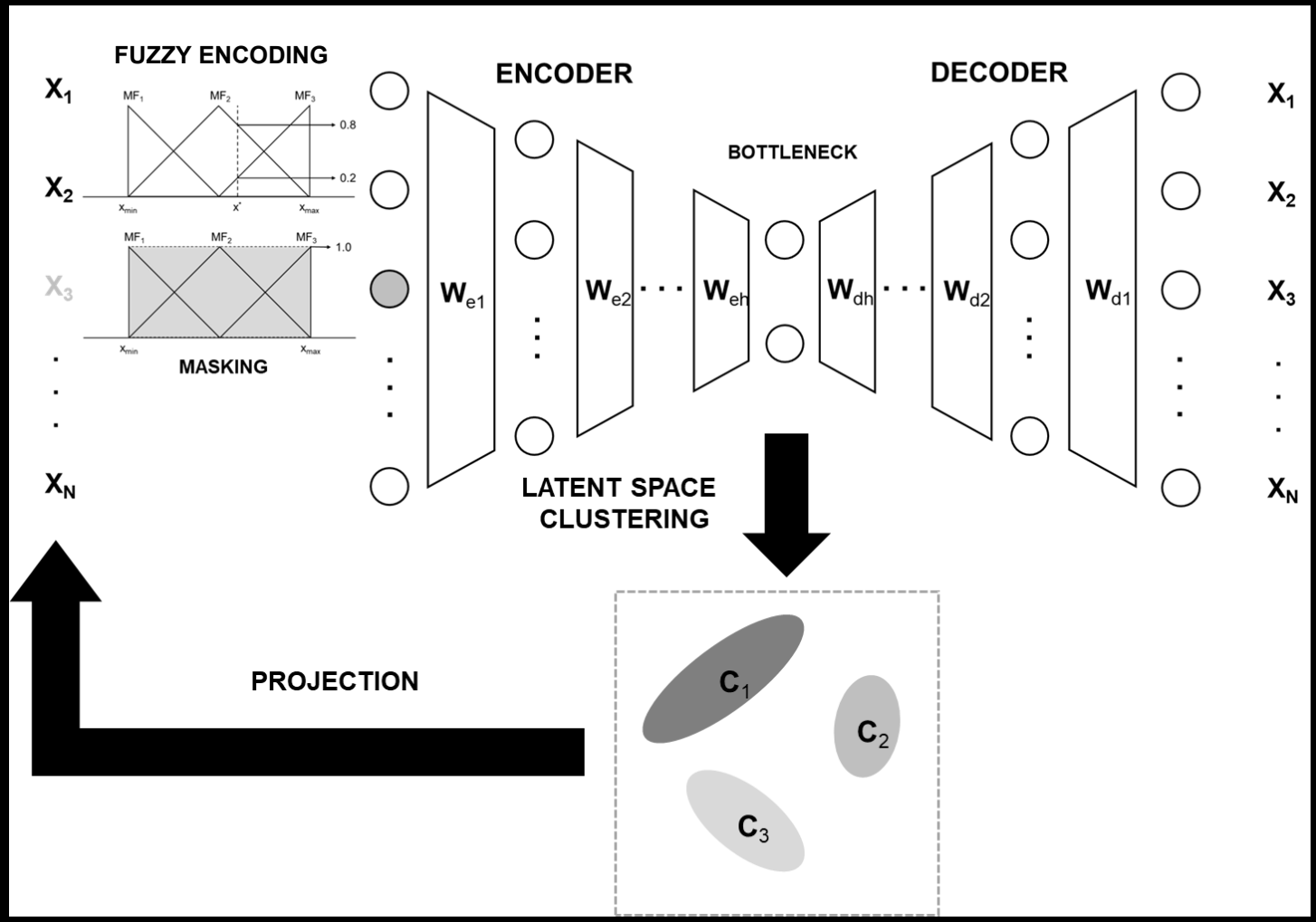
Self Organizing Maps

Autoencoders
(Self-Supervised Learning)

AUTOENCODER SELF-SUPERVISED LEARNING



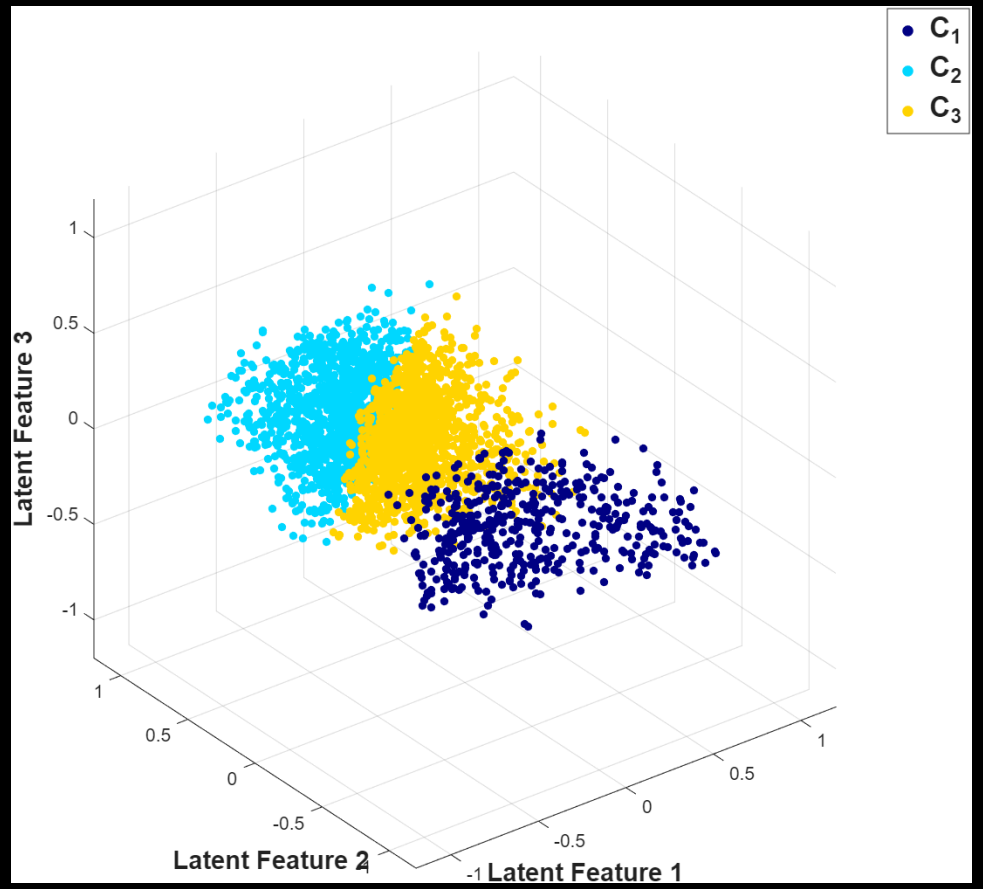
OPENING THE "BLACK BOX"

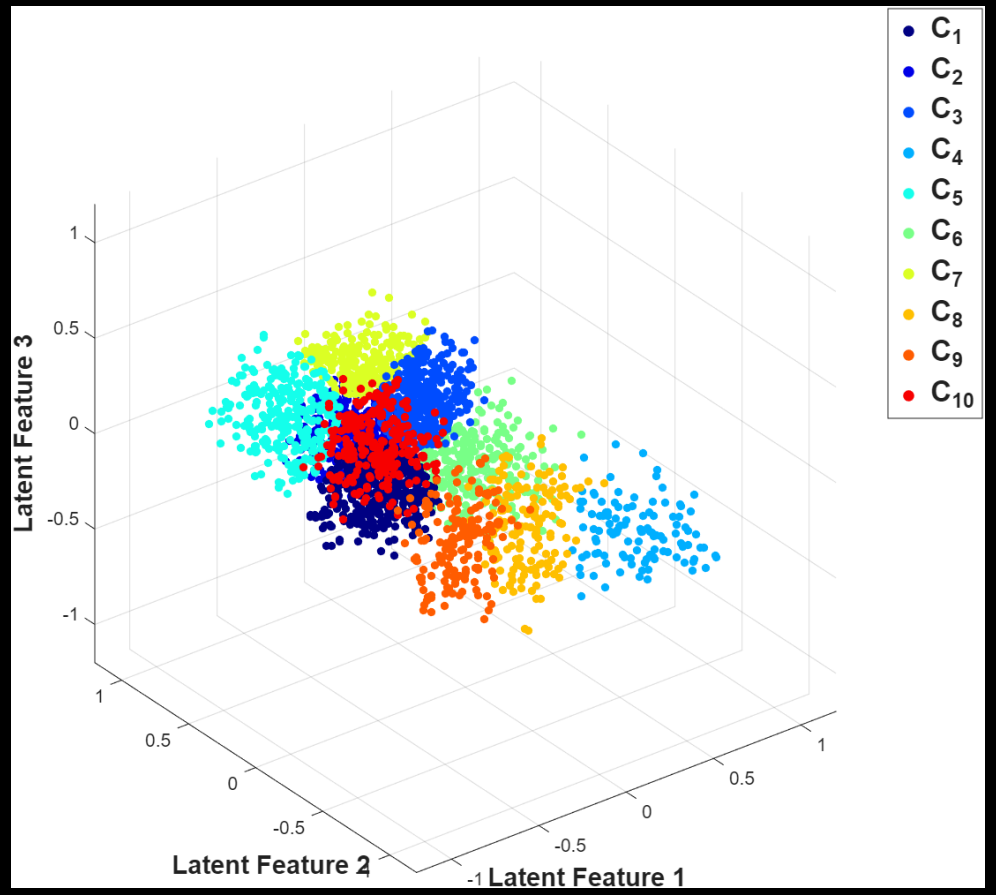


CRIC COHORT
SELECT BIOMARKERS USED IN ANALYSIS

Inflammation	Lipid Metabolism	Endothelial Modulators	Mineral Metabolism
HS CRP	Triglycerides	ADMA	Alkaline Phosphatase
suPAR	Total Cholesterol	SDMA	Ca (serum, intake, urine)
TNF Receptor 1	HDL	ARG	Mg (serum, intake)
TNF Receptor 2	LDL	SICAM-1	PO_4 (serum, intake, urine)
TNF Alpha	APOA1	Fractalkine	PTH
IL6	APOB	MCP-1	Calcitriol
IL10	LP(a)	TGF β	FGF23
		Fibrinogen	Fetuin-A
		FGF23	
		Fetuin-A	

LATENT SPACE CLUSTERING





OPENING THE "BLACK BOX" EXAMPLE

IMPORTANT FACTORS

HIGH GFR (> 50)	MEDIUM GFR (30 – 50)	LOW GFR (<30)
Sex (Male / Female) PO_4 (intake, urine) HDL, LDL, Total Cholesterol APOA1, APOB Basophils Cinnamoylglycine Kynurenic acid (urine) Tiglylglycine Xanthosine (urine) Statin Use	Sex (Male / Female) Uric acid Ca (serum) HDL, LDL, Total Cholesterol APOA1, APOB	PO_4 (serum) PTH, FGF23 LDL, Total Cholesterol APOA1, APOB B2M, BNP, BTP, CXCL12, IL6, MPO Fetuin-A, Fibrinogen, Fractalkine Adipic acid, Cinnamoylglycine (serum, urine) Dimethyluric acid Hippurate (serum, urine) P-cresol sulfate (serum, urine) Tiglylglycine, Xanthosine

CONCLUSIONS

- Uremic Vasculopathy is a complex, multi-factorial disease process that evolves over the course of Chronic Kidney Disease with significant impact on morbidity and mortality.
- Better understanding of the pathophysiology of uremic vasculopathy will improve diagnosis, treatment, and prevention.
- Complex, multi-factorial nature of this phenomenon poses challenges that call for advanced data scientific methods.
- Human inspired computing (AI) offers solutions to address these challenges.

ACKNOWLEDGMENTS

COLLABORATORS:

Eleanor D Lederer

Carlos Alvarez

Christine Vu

FUNDING:

Department of Veterans Affairs

The study is funded by VA Merit Review grant CSR&D 0001614 (E.D.Lederer)
This work reflects the views of the authors and does not represent the views of the Department of Veterans Affairs.

Data from the CRIC study were supplied by NIDDK Central Repository.
This work was not performed in collaboration with the CRIC Investigators
This work does not necessarily reflect the opinions or views of the CRIC Investigators, NIDDK Central Repository, or NIDDK.

Creation of the CHLA Pediatric Acute Liver Failure (CHALF) Score to accurately predict survival with native liver

Juliet Emamaullee MD PhD

Associate Professor of Surgery (Clinical Scholar)
Associate Chief, Division of Clinical Research
Director of Research, Abdominal transplant surgery (CHLA)
Department of Surgery
University of Southern California

@DrEmamaullee

Disclosures

In relation to this presentation, I declare the following, real or perceived conflicts of interest:

- This project has been funded, in part, by our regional OPO OneLegacy Foundation, the California Institute of Regenerative Medicine, and institutional funding from Children's Hospital Los Angeles

- I am the recipient of a K08 Award from the National Cancer Institute

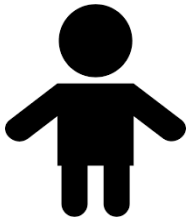
Pediatric Acute Liver Failure (PALF):
One of the most challenging patient populations in pediatric liver transplantation

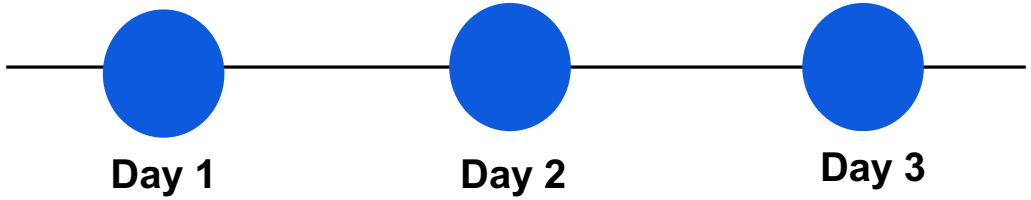
Etiology:

- Indeterminate (40%)
- Determinate (60%)

Diagnostic Criteria:

- Liver injury (AST/ALT)
- Liver function impaired (INR)
- Encephalopathy





Wait or transplant?

Outcomes:

- Survival with native liver (70%)
- Liver transplant (15%)
- Death (15%)

EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

Children's Hospital LOS ANGELES

NIDDK PALF Study Group (PALFSG)

Completed

A Multi-Center Group to Study Acute Liver Failure in Children

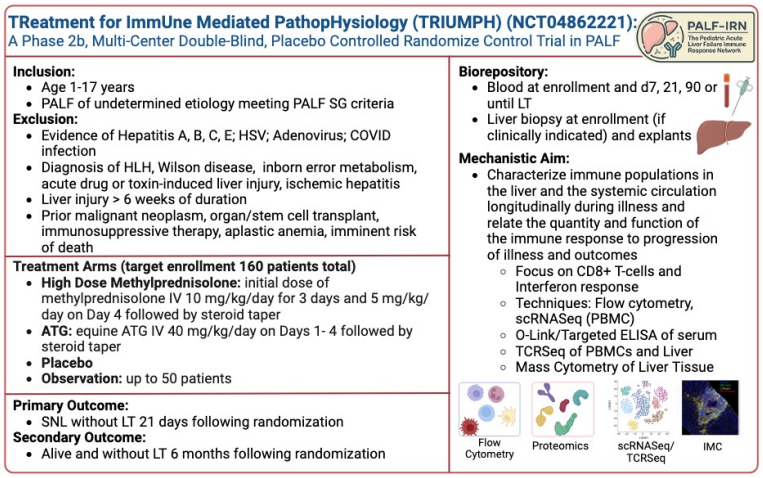
ClinicalTrials.gov ID NCT00986648

Sponsor University of Pittsburgh

Information provided by University of Pittsburgh (Responsible Party)

Last Update Posted 2016-01-12

Download Save



Study Details | Researcher View | No Results Posted | Record History

On this page

- Study Overview
- Contacts and Locations
- Participation Criteria
- Study Plan
- Collaborators and Investigators
- Publications
- Study Record Dates

Study Overview

Brief Summary

The PALF study group began with 20 sites and now continues with 12 sites (11 in the United States and 1 in Canada) in the new funding period. The primary objective of the Pediatric Acute Liver Failure (PALF) study is to collect, maintain, analyze, and report clinical, epidemiological, and outcome data in children with ALF, including information derived from biospecimens.

Detailed Description

The PALF study group will collect clinical, epidemiological and outcome data on children with ALF. This

Study Start 2000-01

Primary Completion (Actual) 2015-12

Study Completion (Actual) 2015-12

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

A Case Series of Children with Acute Hepatitis and Human Adenovirus Infection

Authors: L. Helena Gutierrez Sanchez, M.D., Henry Shiau, M.D., Julia M. Baker, Ph.D., Stephanie Saaybi, M.D., Markus Buchfellner, M.D., William Britt, M.D., Veronica Sanchez, Ph.D., +22 , and Hannah L. Kirking, M.D. Author Info & Affiliations

Published July 13, 2022 | N Engl J Med 2022;387:620-630 | DOI: 10.1056/NEJMoa2206294 | VOL. 387 NO. 7

Can we predict whether patients with PALF recover from their disease or need a liver transplant?

(and in doing so, reveal facets of PALF biology?)

EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

Children's Hospital LOS ANGELES

Scoring Systems to Predict Clinical Trajectory in PALF

- **PELD Score**
  - Designed to predict waitlist mortality in children with chronic liver disease
  - bilirubin, INR, albumin, growth failure, and age
  - Does not reliably predict death in PALF
- **Kings College Hospital Criteria**
  - Designed to predict mortality in ALF
  - Prothrombin time >100 s (INR > 6.5) OR any 3 of the following (irrespective of grade of HE):
    - Age <10 or >40 years, etiology: non-A/non-B hepatitis, drug-induced
    - Duration of jaundice to HE >7 days, Prothrombin time >50 (INR > 3.5), Serum bilirubin >300
  - Does not reliably predict death in PALF
- **Liver Injury Units (LIU) Score**
  - Designed to predict mortality in PALF
  - Peak serum total bilirubin, peak PT/INR, and peak ammonia
  - Good predictive value (c-index of 0.81) but requires peak values
  - Admission LIU (aLIU) less predictive (c-index 0.76)

EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

McDiarmid et al, Transplantation. 2002 Jul 27;74(2):173-81.
Liu et al, J Hepatol. 2006 Jan;44(1):134-41.
Sundaram et al, J Pediatr. 2013 Feb; 162(2): 319–323.e1

Children's Hospital LOS ANGELES

Objective

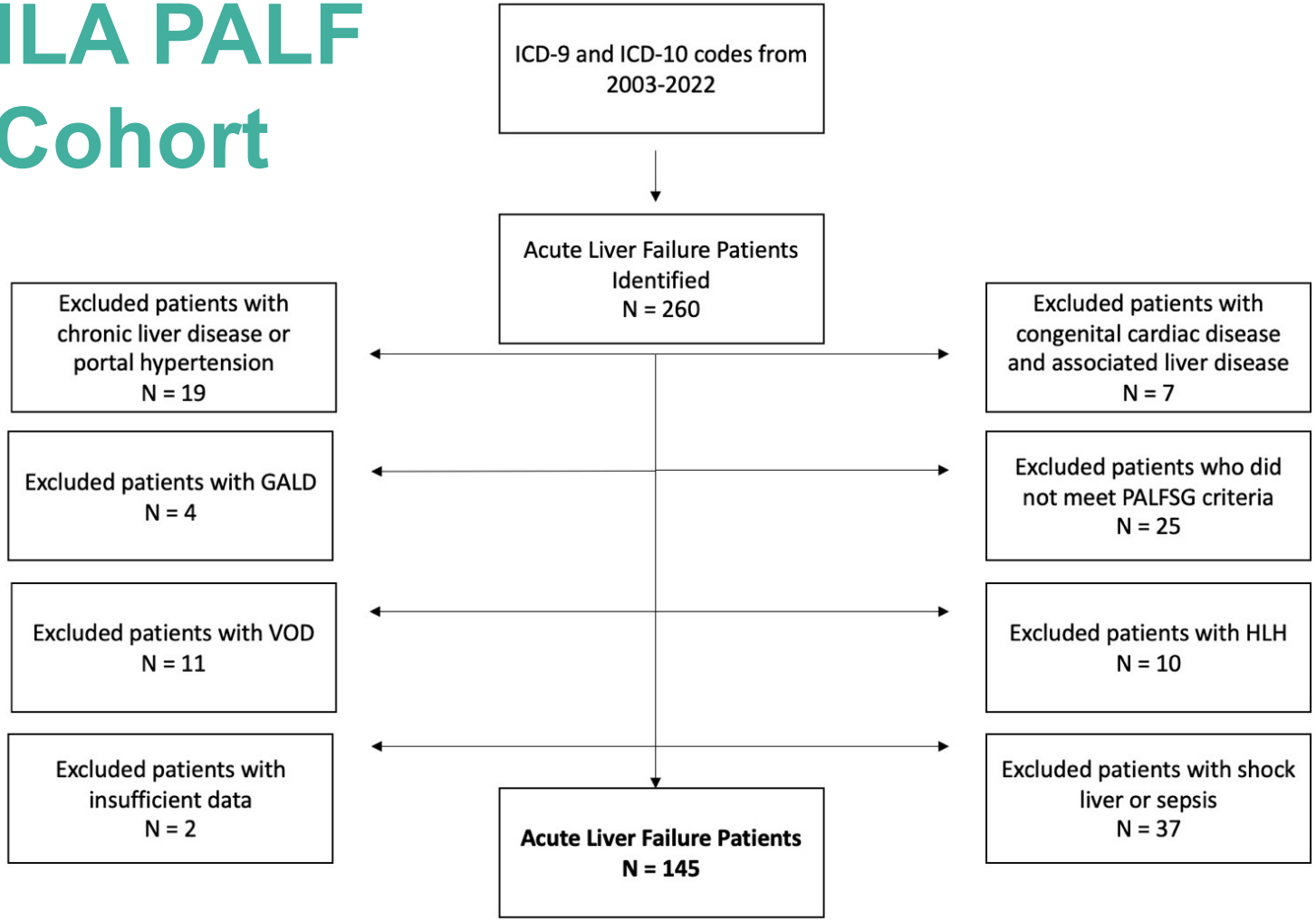
The aim of this study was to create a model based on **admission** clinical variables from a large, diverse transplant referral center population that accurately predicts clinical outcomes in PALF

- Avoid subjective (i.e. HE) and time-based (i.e. duration of jaundice) criteria
- Avoid 'peak' lab values
- Use common clinical demographic variables/lab values
- Generate easily interpreted 'score' that doesn't require statistical understanding
- Facilitate referral to a transplant center

EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

Children's Hospital LOS ANGELES

CHLA PALF Cohort



EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

Children's Hospital LOS ANGELES

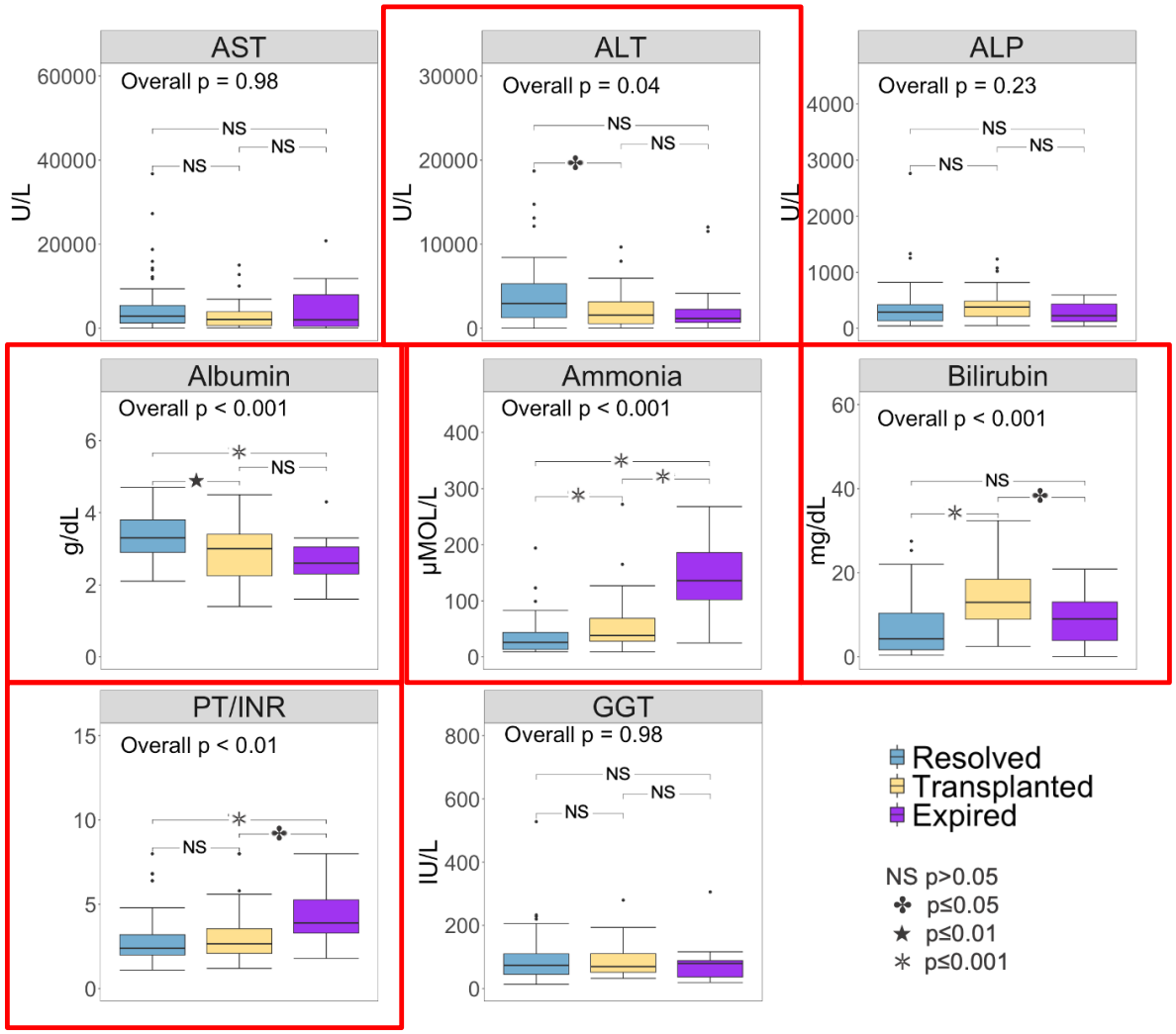
CHLA PALF Cohort

	Overall	SNL	LT	Expired	P-value
Number of patients, N (%)	145	85 (59%)	45 (31%)	15 (10%)	
Listed for transplant, N (%)	82 (57%)	34 (40%)	45 (100%)	3 (20%)	**0.03 ***
Age at Admission, years (median [IQR])	6 [2 ,14]	8 [2, 15]	6 [3, 12]	1 [0, 2]	**< 0.01 ****
Female, N (%)	84 (58%)	50 (59%)	26 (58%)	8 (53%)	0.85
Race/Ethnicity, N (%)					**0.01 ****
White		22 (26%)	9 (20%)	2 (13%)	
Hispanic		35 (41%)	31 (70%)	5 (33%)	
Black		4 (5%)	2 (4%)	2 (13%)	
Asian/Pacific		6 (7%)	-	1 (8%)	
Other		13 (15%)	1 (2%)	5 (33%)	
Unknown		5 (6%)	2 (4%)	-	
Diagnosis, N (%)					**0.001 ****
Indeterminant	52 (37%)	24 (28%)	21 (47%)	7 (47%)	
Drug-induced		32 (38%)	3 (7%)	1 (7%)	
Autoimmune Hepatitis		12 (14%)	10 (22%)	1 (7%)	
Infection		7 (8%)	9 (20%)	4 (27%)	
Other		10 (12%)	2 (4%)	2 (13%)	

EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

Children's Hospital LOS ANGELES

Evaluation of Individual Variables by Clinical Outcome



EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

Children's Hospital LOS ANGELES

Approach to building a predictive model of survival with native liver in PALF

- **Variables without statistical significance from the initial analysis were excluded from the predictive modeling.**
  - A correlation matrix created from the remaining variables further excluded variables to avoid issues of multicollinearity.
  - Variables were log-transformed, and data was randomly split into training and testing sets at a 75/25 ratio to evaluate internal model performance.
- **Statistical modeling was carried out to predict the outcome of a patient based on clinical variables, using demographics as confounder variables.**
  - Modeled as a two-class comparison: SNL vs. no SNL to prevent class imbalance
  - Several techniques evaluated: multinomial logistic regression (MLR), random forest (RF), XGBoost, support vector machine (SVM), linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA)

EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

Children's Hospital LOS ANGELES

Validation Cohort: NIH PALF Study Group

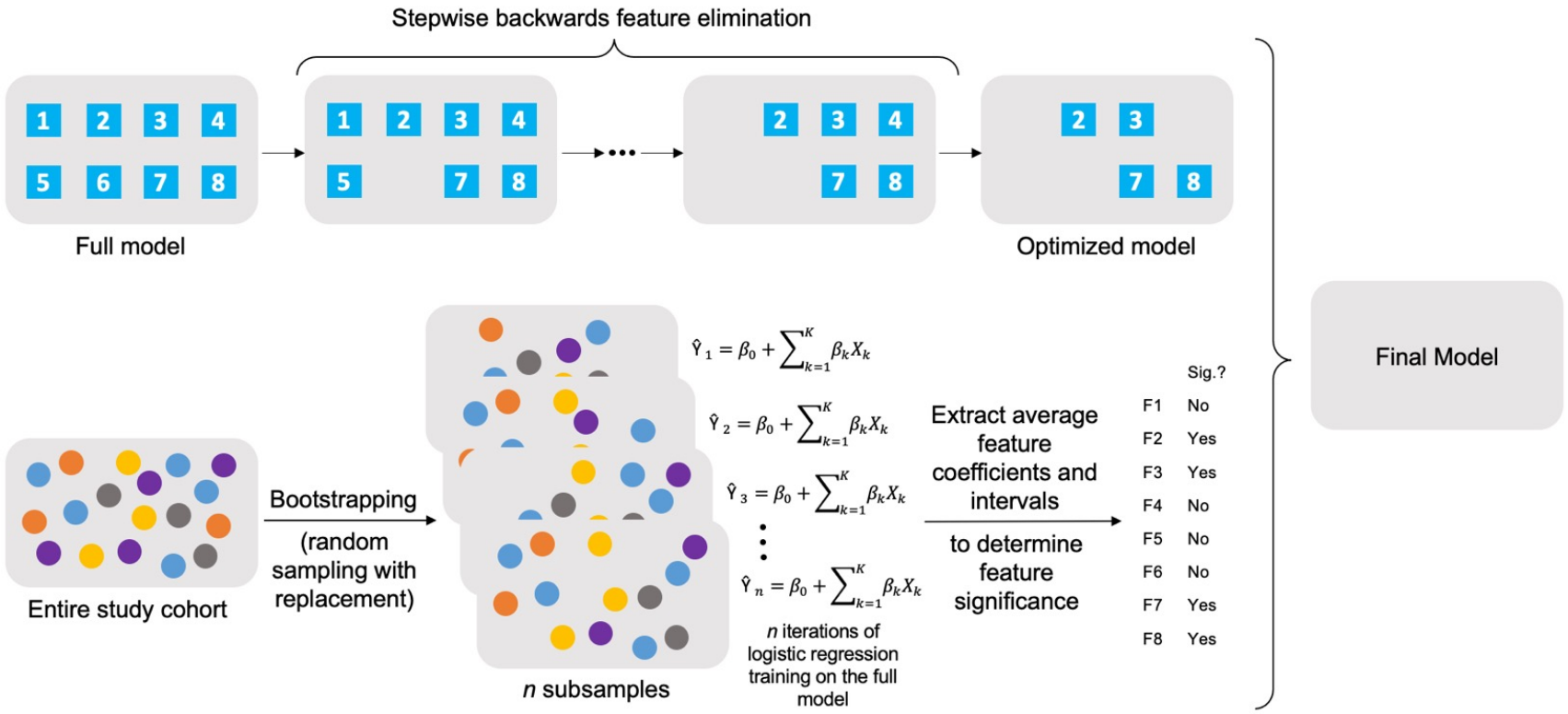
https://repository.niddk.nih.gov/studies/palf/

	Resolved	Transplanted	Expired	p-value
Number of patients, N (%)	248 (50%)	171 (35%)	73 (15%)	
Waitlisted, N (%)	57 (23%)	171 (100%)	36 (49%)	≤ 0.001 ***
Age at Admission, years (median [Q1-Q3])	5 [1 – 14]	5 [2 – 10.5]	2 [0 – 9]	≤ 0.001 ***
Female, N (%)	140 (56%)	76 (44%)	27 (37%)	≤ 0.01 **
Race/Ethnicity, N (%)				
White	150 (61%)	100 (59%)	36 (50%)	
Hispanic	34 (14%)	35 (20%)	20 (27%)	
Black	18 (7%)	21 (12%)	7 (10%)	
Asian/Pacific	27 (11%)	10 (6%)	9 (12%)	
Other	6 (2%)	4 (2%)	-	
Multi-race	13 (5%)	1 (1%)	1 (1%)	
Admission liver enzyme results, median [Q1-Q3]				
AST	1,942 [578 – 4,531]	1,777 [422 – 2,922]	1,181 [178 – 3,326]	0.02 *
ALT	1,720 [462.5 – 4,097]	1,405 [303.5 – 2,447.5]	764 [111.5 – 1,898]	≤ 0.001 ***
Albumin	3 [2.6 – 3.4]	2.9 [2.5 – 3.3]	2.7 [2.2 – 3.1]	≤ 0.01 **
Bllirubin	4.2 [1.7 – 12.1]	16.4 [10.8 – 20.7]	11.6 [6.2 – 19]	≤ 0.001 ***
PT/INR	2.4 [1.9 – 3.4]	2.8 [2.1 – 4.2]	2.9 [2 – 4]	0.03 *
GGT	66.5 [45 – 118.2]	62 [41 – 104]	59 [40.5 – 122]	0.54
Ammonia	53 [33 – 80.8]	68 [45 – 109.5]	86 [43 – 125]	≤ 0.001 ***

EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

Children's Hospital LOS ANGELES

Variable selection for the final predictive model



Reporting of our multivariable prediction model followed the TRIPOD checklist for prediction model development and validation

EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

Creation of the PALF predictive model

- Admission variables in the final model:
  - Albumin (OR 18.1; $p<0.01$)
  - Ammonia (OR 0.43; $p<0.01$)
  - Total Bilirubin (OR 0.45; $p<0.001$)

- The model performed well in predicting SNL (C-statistic 0.83)

- The area under the curve (AUC; 0.83 training, 0.78 validation), accuracy (0.75 training, 0.70 validation) and performance (F1 score 0.79 training, 0.65 validation) confirm the model's predictive accuracy

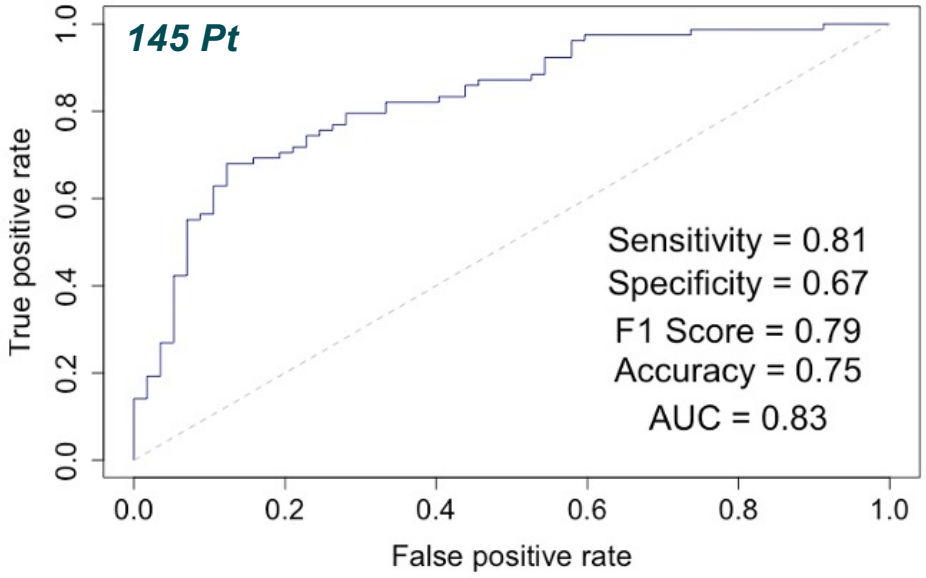
EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

Children's Hospital LOS ANGELES

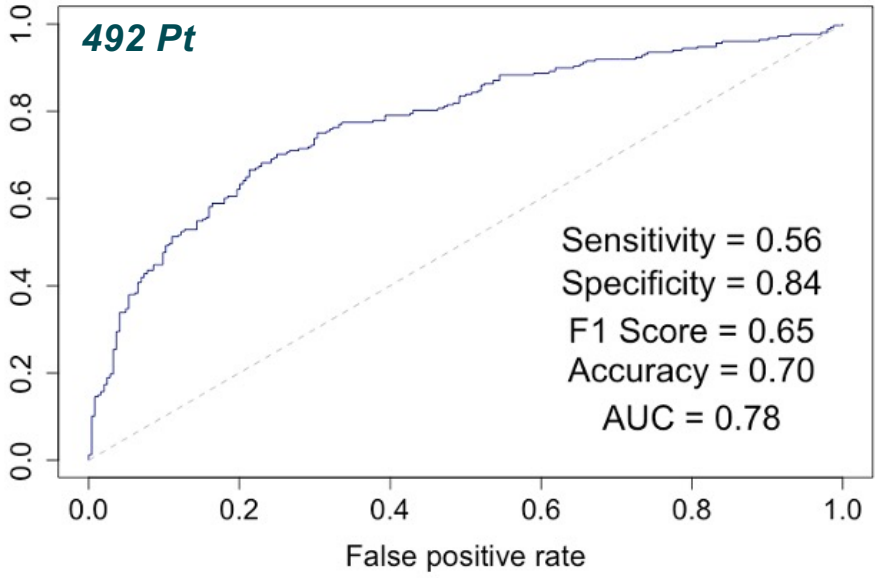
Creation of the PALF Predictive Model

$$\text{Probability SNL} = \frac{e^{(1.8846 + 2.773\ln(\text{albumin}) - 0.8193\ln(\text{total bilirubin}) - 0.8615\ln(\text{ammonia}))}}{1 + e^{(1.8846 + 2.773\ln(\text{albumin}) - 0.8193\ln(\text{total bilirubin}) - 0.8615\ln(\text{ammonia}))}}$$

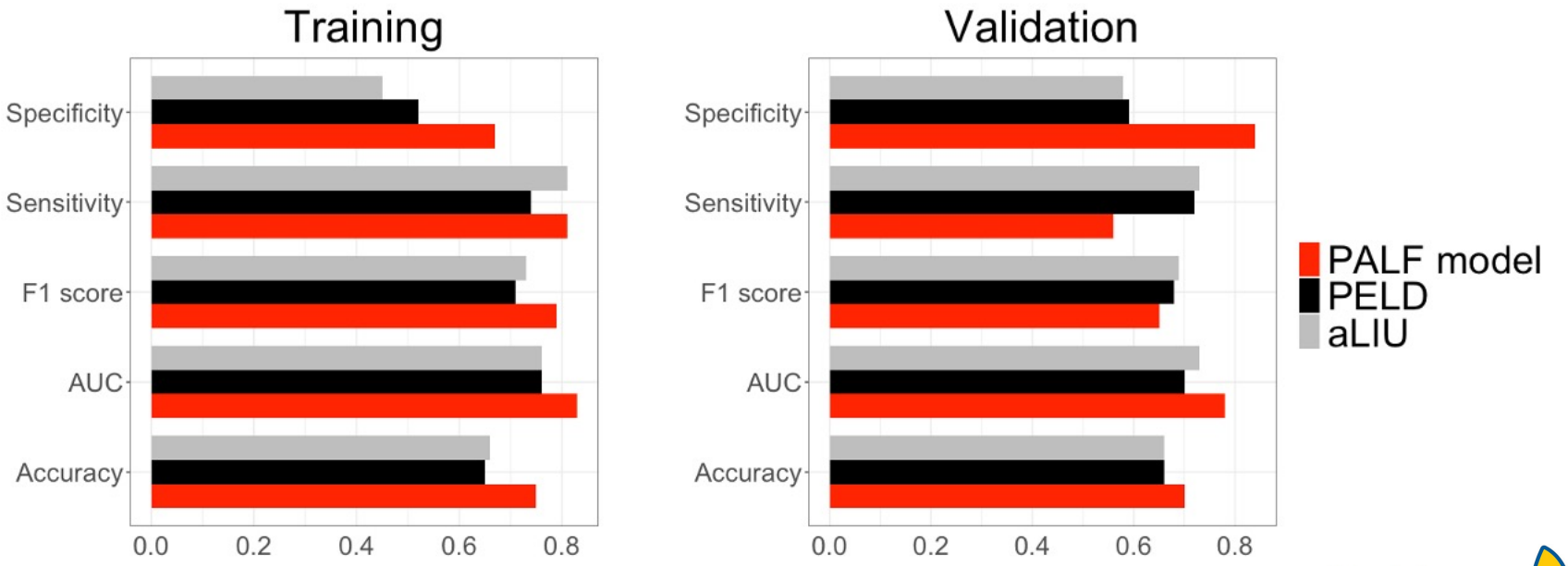
Training ROC Curve



Validation ROC Curve



The PALF model outperformed both PELD (C-statistic 0.76) and Liver Injury Unit (C-statistic 0.76) score in predicting SNL



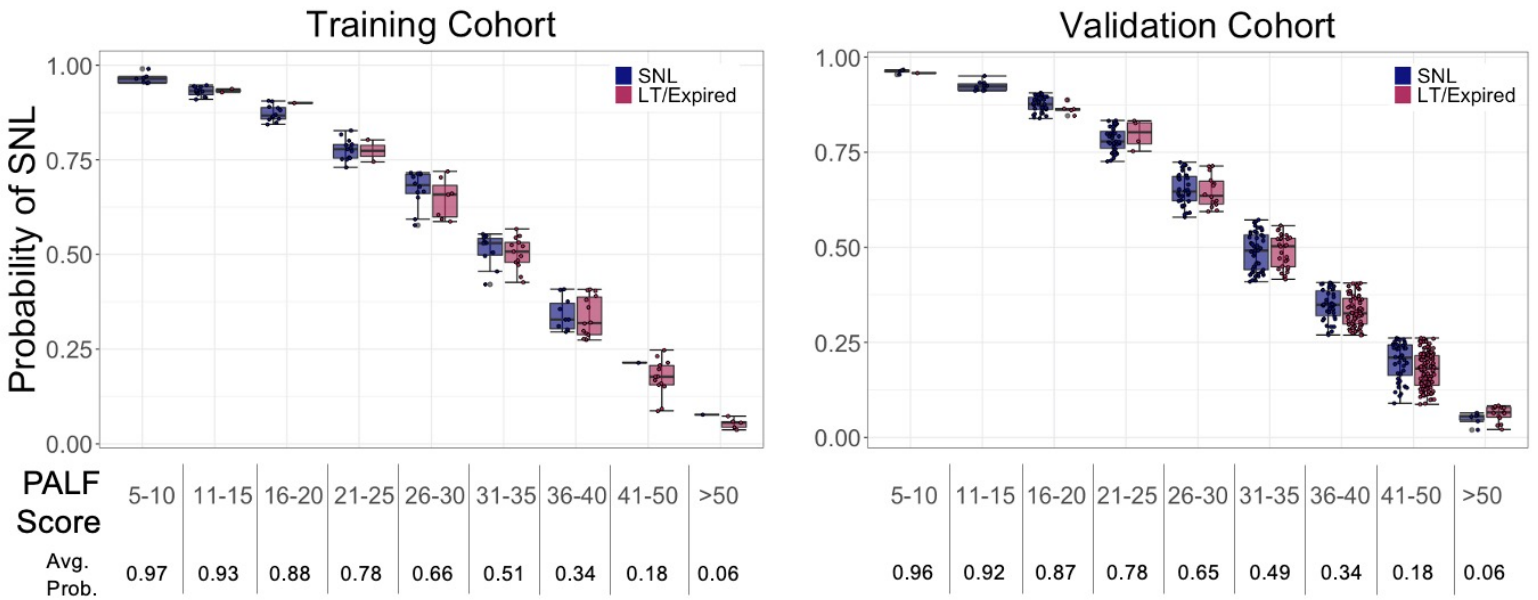
EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

Children's Hospital LOS ANGELES

Creation of the CHALF Score

$$\text{PALF Score} = \frac{\left[1 - \left(1.8846 + 2.773\ \ln(\text{albumin}) - 0.8103\ \ln(\text{total bilirubin}) - 0.8615\ \ln(\text{ammonia})\right)\right] * 15 + 50}{2}$$

- **Range 5-60**
- **Scores >30 predict lower chance of survival without LT and higher chance of mortality**



EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

Children's Hospital LOS ANGELES

Strengths and Limitations

- **Strengths:**
  - Large sample size in both training and validation cohorts
  - Ethnically and clinically diverse patient population in CHLA cohort
    - Real world experience with not all patients being evaluated for LT and/or being on a study
  - Model relies in easily and frequently ordered admission lab tests

- **Weaknesses:**
  - Class imbalance within clinical outcomes can impact statistical modeling
  - Some patients who died may have survived with LT and some patients who could have survived without transplant may have undergone LT

EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

Children's Hospital LOS ANGELES

Conclusion

- The CHALF Score accurately predicts SNL using common admission labs (albumin, ammonia, and total bilirubin)

- This novel, externally validated score, offers a reliable measure of SNL and potential guidance for early referral for transplant evaluation

EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

Children's Hospital LOS ANGELES

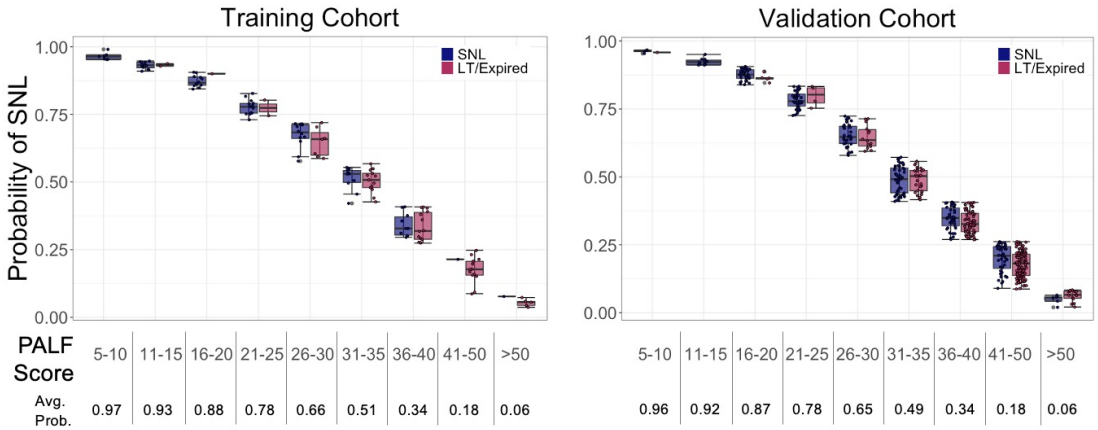
Original Clinical Science—Liver

CHALF Score: A Novel Tool to Rapidly Risk Stratify Children in Need of Liver Transplant Evaluation During Acute Liver Failure

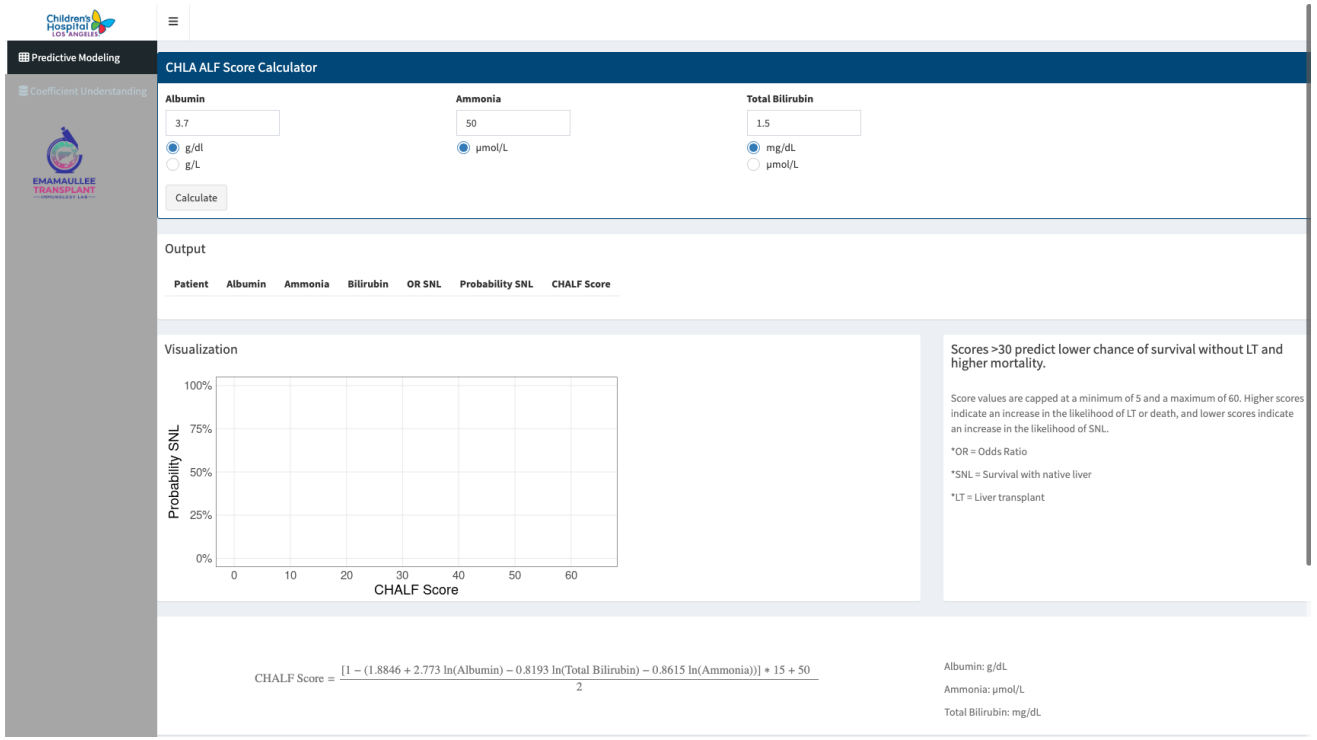
Johanna M. Ascher-Bartlett, MD,[1,2] Sarah Bangerth, MS,[2,3] Shannon Jordan, BS,[2,3] Carly Weaver, BS,[1] Sarah Barhouma, MHA,[2,3] Kambiz Etesami, MD,[2,3] Rohit Kohli, MBBS, MS,[1,2] and Juliet Emamaullee, MD, PhD, FRCSC[2,3]

- **Uses admission bilirubin, ammonia, and albumin**
- **Range 5-60**
- **Scores >30 predict lower chance of survival without LT and higher chance of mortality**

$$\text{PALF Score} = \frac{[1 - (1.8846 + 2.773 \ln(\text{albumin}) - 0.8103 \ln(\text{total bilirubin}) - 0.8615 \ln(\text{ammonia}))] * 15 + 50}{2}$$



CHALF Score App



EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

Children's Hospital LOS ANGELES

Future Directions

- Multi-center real world validation with time series data
  - Kings College, Columbia, Kansas City-Mercy Children's Hospital
  - ICU transfer study in partnership with UK PICU Collaborative

EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

Children's Hospital LOS ANGELES

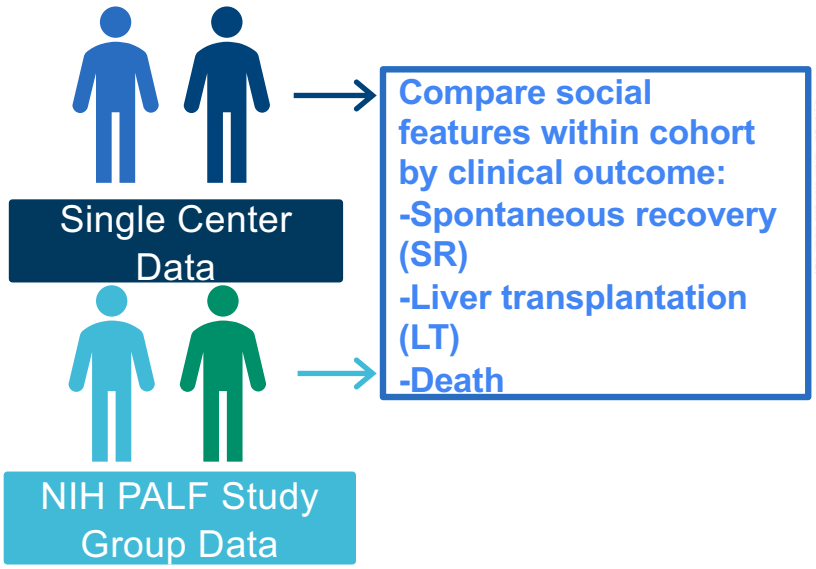
Additional applications of NIDDK PALF Study Group Data and Repository

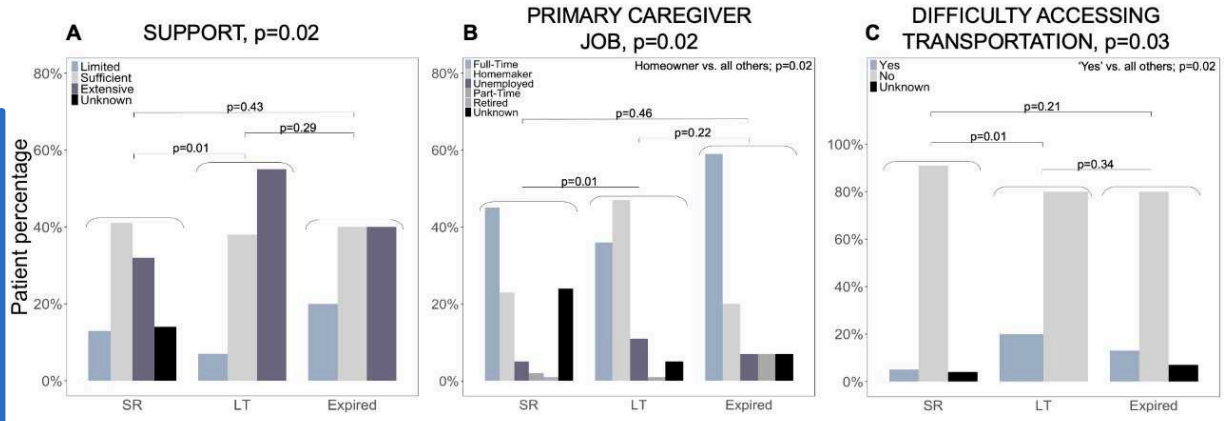
Finance, race, ethnicity and spoken language impact clinical outcomes for children with acute liver failure

PEDIATRIC TRANSPLANTATION

HYPOTHESIS: Social determinants of health may impact clinical outcomes following pediatric acute liver failure (PALF) as well as present barriers to obtaining timely care in a transplant

DESIGN & OUTCOMES:





Social factors including family support systems, caretaker employment, and patient spoken language may impact clinical outcome following acute liver failure in children.

CONCLUSION: The observed differences in clinical outcomes in PALF related to social determinants of health may highlight unconscious biases held by transplant teams. Proactive strategies to support and engage families with risk factors could improve patient outcomes.

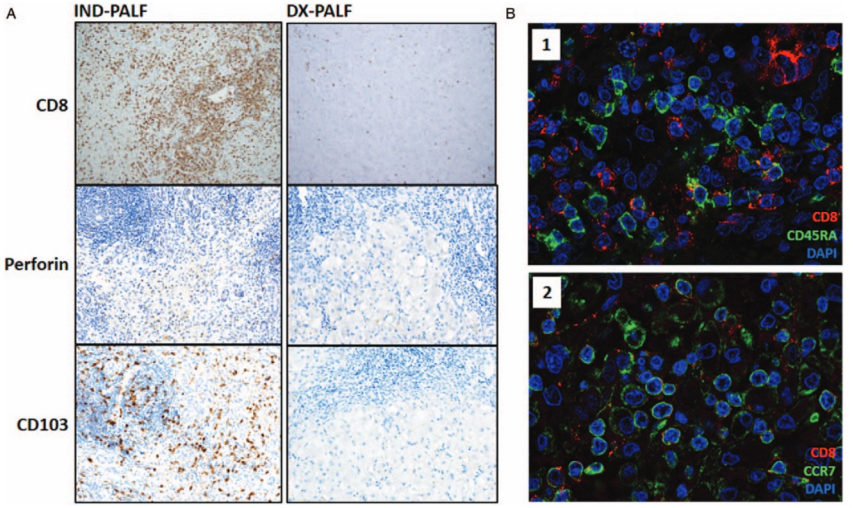
Ascher Bartlett et al. 2023

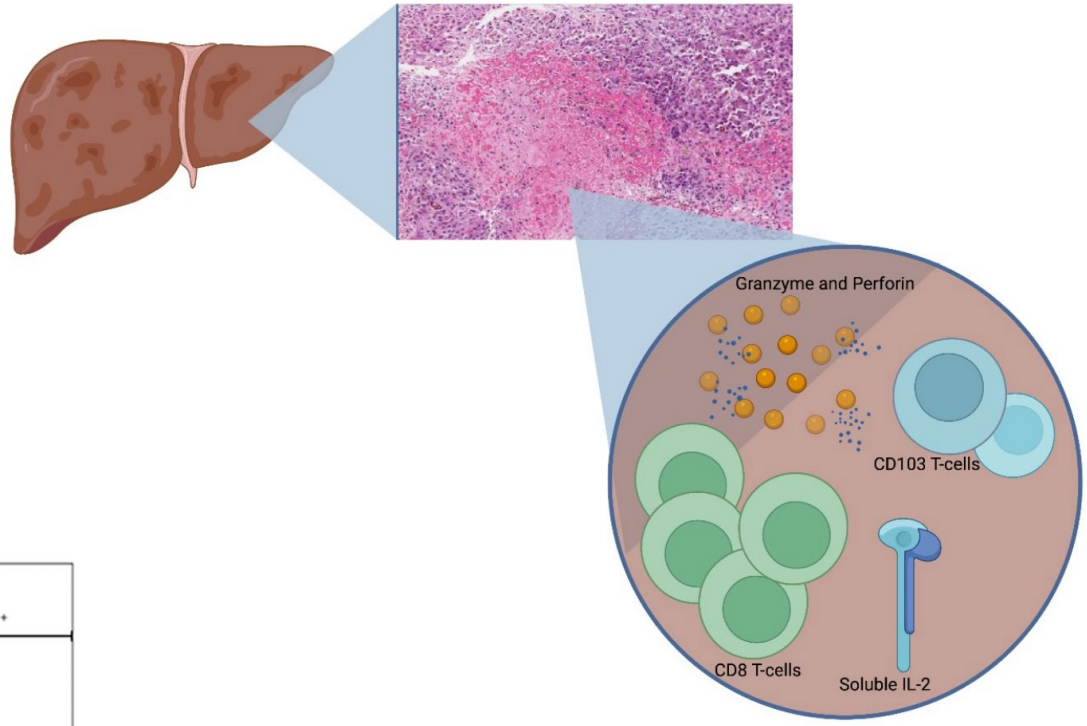
WILEY

International Pediatric Transplant Association

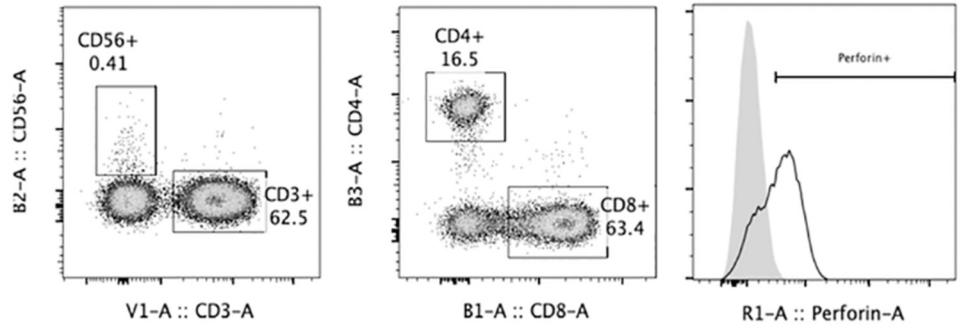
PALFSG Identified CD8 T-cell Signatures in Liver and Blood

Liver Tissue





Blood



Chapin et al (2020). J Pediatr Gastroenterol Nutr.
Chapin et al (2023). PLoS One.
Ascher-Bartlett et al (2022). Liver Transplantation

EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

SCIENCE ADVANCES | RESEARCH ARTICLE

IMMUNOLOGY

Spatially resolved immune exhaustion within the alloreactive microenvironment predicts liver transplant rejection

Arianna Barbetta[1]†, Brittany Rocque[1]†, Sarah Bangerth[1], Kelly Street[2], Carly Weaver[3], Shefali Chopra[4], Janet Kim[1], Linda Sher[1], Brice Gaudilliere[5], Omid Akbari[6,7], Rohit Kohli[3,8], Juliet Emamaullee[1,6,8]*

Allograft rejection is common following clinical organ transplantation, but defining specific immune subsets mediating alloimmunity has been elusive. Calcineurin inhibitor dose escalation, corticosteroids, and/or lymphocyte depleting antibodies have remained the primary options for treatment of clinical rejection episodes. Here, we developed a highly multiplexed imaging mass cytometry panel to study the immune response in archival biopsies from 79 liver transplant (LT) recipients with either no rejection (NR), acute T cell–mediated rejection (TCMR), or chronic rejection (CR). This approach generated a spatially resolved proteomic atlas of 461,816 cells (42 phenotypes) derived from 96 pathologist-selected regions of interest. Our analysis revealed that regulatory (HLADR$^+$ T_{reg}) and PD1$^+$ T cell phenotypes (CD4$^+$ and CD8$^+$ subsets), combined with variations in M2 macrophage polarization, were a unique signature of active TCMR. These data provide insights into the alloimmune microenvironment in clinical LT, including identification of potential targets for focused immunotherapy during rejection episodes and suggestion of a substantial role for immune exhaustion in TCMR.

Online April 12, 2024

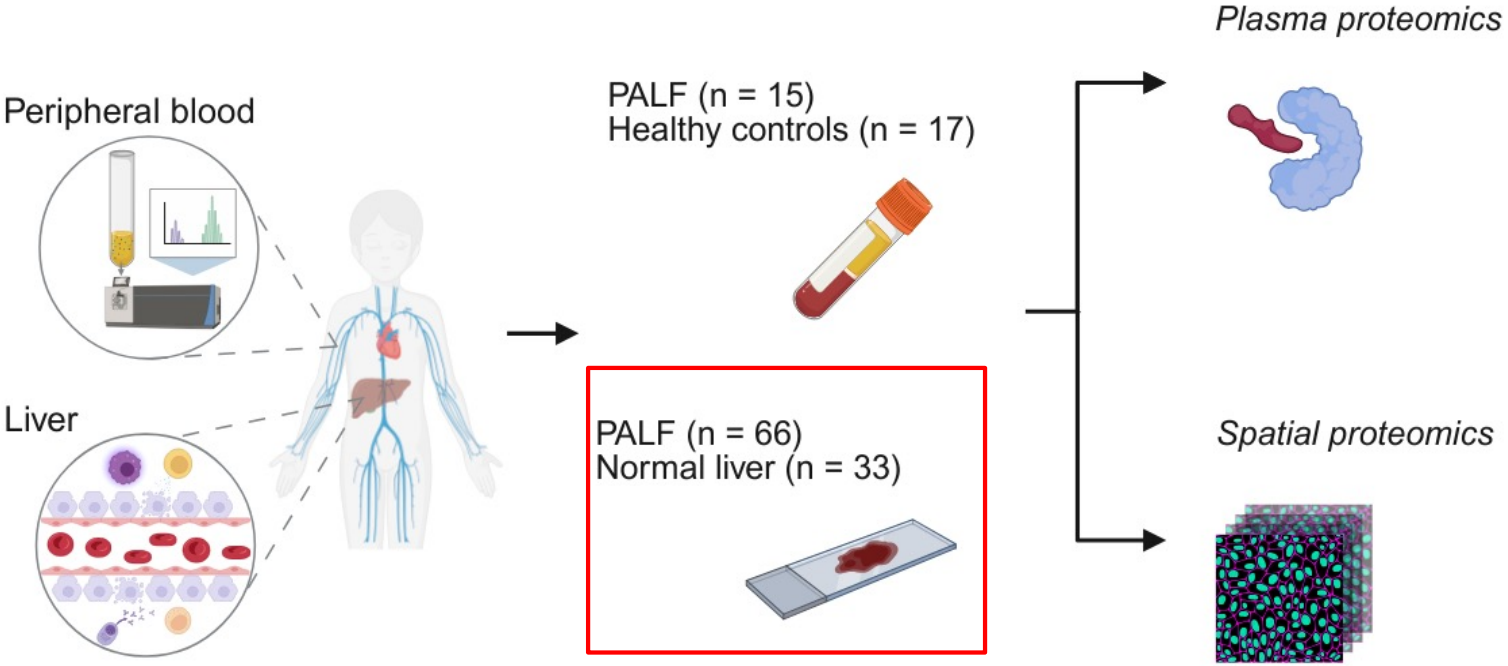
scientific reports

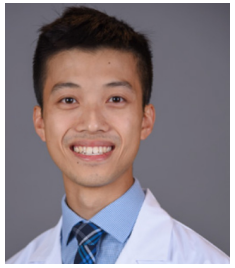
OPEN

Integrated workflow for analysis of immune enriched spatial proteomic data with IMmuneCite

Arianna Barbetta[1], Sarah Bangerth[1], Jason T. C. Lee[1], Brittany Rocque[1,2], Evanthia T. Roussos Torres[3,4], Rohit Kohli[5,7], Omid Akbari[3,6] & Juliet Emamaullee[1,3,7]✉

Experimental Design:
Single center analysis of liver tissue and blood in patients with PALF





Jason Lee MD
(Surgical Resident/PhD Student

MAMAULLEE TRANSPLANT IMMUNOLOGY LAB

Children's Hospital LOS ANGELES | Keck School of Medicine of USC

Request ID: 23430
Requestor Name: Juliet Emamaullee
Study: PALF
Date: July 21, 2022 /version 1

NIDDK Repository Sample Availability Estimation Report

Disclaimer: All available sample volumes and/or masses are ESTIMATED, not actual. This report is based on current sample inventory and sample availability may be impacted by other requests that are fulfilled between issue of this report and disbursement of your materials.

This form is NOT a sample request.

If based on the information in this report, the NIDDK Central Repository holds sufficient samples for your Research Project, attach this form to your grant application (e.g., X01, R01) or follow the steps outlined at the end of this report to request samples via the Administrative Review process (if applicable).

1. Description of the proposed sample request

This request is for 0.05 mL of serum and 1 unit of tissue slides from participants in the PALF study, stratified by clinical outcome (recovery, liver transplant, and death). Specimens are requested within 7 days of study enrollment.

2. Current PALF sample inventory

The NIDDK Central Repository verifies that the following samples are resident in the NIDDK

EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

Children's Hospital LOS ANGELES | Keck School of Medicine of USC

Available Volume of 1 Tissue Slide within 7 Days of Enrollment

Volume (unit)	Total Samples	% Reduction	Impact
X = 1	0	100%	Depleted
1 < X <= 4	4	25 - 99%	Significant
4 < X <= 10	42	10 - 24%	Moderate
X > 10	9	<10%	Modest

Available Volume of 1 Tissue Slide within 7 Days of Enrollment, by Clinical Outcome

Volume (unit)	Samples for Participants who Recovered	Samples for Participants who Received a Transplant	Samples for Participants who Died	% Reduction	Impact
X = 1	0	0	0	100%	Depleted
1 < X <= 4	2	2	0	25 - 99%	Significant
4 < X <= 10	8	31	3	10 - 24%	Moderate
X > 10	1	8	0	<10%	Modest

Conclusions

- The NIDDK PALF Study Group Data and Tissue Repository have been valuable resources to augment and validate our single center studies and experience at Children's Hospital-Los Angeles

- NIDDK PALF Study Group data were used to validate
  - CHALF Score to predict need for LT in children with acute liver failure
  - Impact of social determinants of health on clinical outcomes in PALF

EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

Children's Hospital LOS ANGELES | Keck School of Medicine of USC

Acknowledgements

- Emamaullee Research Group
  - Arianna Barbetta MD (doctoral student)
  - **Jason Lee MD, MS (resident postdoc)**
  - **Johanna Ascher-Bartlett MD (resident postdoc)**
  - **Sarah Bangerth, MS (Data Scientist)**
  - Brittany Rocque MD MS (resident postdoc)
  - **Carly Weaver BS (Clinical Research Coordinator)**
  - Medical Student RSP:
    - Kate Guion (MS-3)
    - Shresthra Vijayendra (MS-3)
    - Shannon Jordon (MS-3)
    - Deepika Sarode (MS-4)
  - Lauren Pickard (undergrad)
  - Sarah Barhouma (undergrad)
  - Tricia Saputera (undergrad)
  - Pranay Singh (undergrad)
  - Asish Misra MD PhD (transplant fellow)
- Collaborators
  - Ellison Institute (Dr. Jerry Lee, Dr. Nolan Ung)
  - CHLA : Dr. Yanni, Dr. Kohli
  - Pathologists: Dr. Shefali Chopra/Dr. Shengmei Zhou
- Mentoring committee
  - Dr. Omid Akbari
  - Dr. Shahab Asgharzadeh (CHLA SC2 Core Team)
  - Dr. Rohit Kohli
- Department of Surgery
  - Dr. Linda Sher
  - Dr. Vaughn Starnes

https://github.com/julietusc/

We are grateful for the following support:

- NCI K08-CA245220-01 (2020-2025)
- ACS Jacobson Promising Investigator Award (2023)
- Wright Foundation Pilot Grant (2022-2023)
- AASLD CTORA Grant (2020-2022)
- ASTS/Natera Faculty Development Grant (2020-2022)
- Gilead Research Foundation Liver Scholar (2020-2022)
- USC Computational Genomics Pilot Grant (2021-2022)
- USC Pilot Core Award x 2 (2021)
- USC Research Center for Liver Diseases Pilot Grant (2019-2020)
- USC Dean's Pilot Grant (2018-2019)
- **CHLA Liver Transplant Program Startup Funding**
- **One Legacy Foundation Fellowship Grant (Rocque/Misra/Ascher-Barlett)**
- **Broad Institute Clinical Research Fellowship Grant (Rocque/Ascher-Bartlett)**
- **CIRM Clinical Fellowship (Lee)**
- **Saban Research Fellowship 2022-2024 (Ascher-Bartlett)**
- ASTS Fellowship Grant 2021-2023 (Misra)
- ASTS Presidential Student Mentor Award 2021-2023 (Sarode/Guion/Kim)
- USC Undergraduate Research Award Program Grant (Pickard/Eguizabal)
- USC Provost Award/SURF Award (Barhouma)

EMAMAULLEE TRANSPLANT IMMUNOLOGY LAB

JOHNS HOPKINS
SCHOOL of MEDICINE

AI and Body Composition : Novel Methods to improve accuracy

Prasanna Santhanam, MBBS, MD
Associate Professor of Clinical Medicine and Oncology

ELSEVIER

Diabetes & Metabolic Syndrome: Clinical Research & Reviews

Volume 18, Issue 8, August 2024, 103113

Obesity prediction: Novel machine learning insights into waist circumference accuracy

Carl Harris [a], Daniel Olshvang [b], Rama Chellappa [a b], Prasanna Santhanam [c]

Background

- Obesity is a major public health challenge globally, linked to diabetes, CVD, fatty liver, cancer.
- Rising prevalence due to socioeconomic disadvantages, lifestyle changes, reduced productivity.
- Traditional metrics: Weight and BMI – limited sensitivity, high variability across ethnicities.
- Asians have higher body fat % for the same BMI, making BMI inadequate for CVD risk stratification.

Background

- Waist Circumference (WC)Stronger predictor of cardiometabolic risk than BMI.
- WC ≥ 94 cm in men identifies high risk for CVD/T2DM (sensitivity >84%, specificity >78%).
- In Look AHEAD, WC increase during intervention predicted higher CVD risk.
- Better predictor even in children and older adults; trajectories correlate with T2DM onset.

Scope

- Despite evidence, WC is rarely measured in routine practice.• Measurement variability and site differences affect prevalence estimates.
- Guidelines recommend WC as a vital sign, but standardization is lacking.
- AI offers alternative: predict WC from readily available variables.

Scope

- Despite evidence, WC is rarely measured in routine practice.• Measurement variability and site differences affect prevalence estimates.
- Guidelines recommend WC as a vital sign, but standardization is lacking.
- AI offers alternative: predict WC from readily available variables.

Conformal Predictions

- Conformal Prediction with Uncertainty Quantification
- Conformal prediction (CP) is a distribution-free framework for quantifying prediction uncertainty in machine learning, offering valid confidence sets or prediction intervals under the assumption of data exchangeability.
- Unlike conventional point estimates, CP provides guarantees on the error rate, making it attractive for high-stakes applications such as medicine and finance.

Conformal Predictions

- Distribution-free guarantees – CP does not require strong distributional assumptions and ensures coverage probability (e.g., 95%) holds on average across future predictions (Shafer & Vovk, 2008).
- Uncertainty quantification – By constructing prediction sets/intervals, CP explicitly communicates model uncertainty instead of a single deterministic prediction (Angelopoulos & Bates, 2021).

Conformal Predictions

- Model-agnostic applicability – CP can be applied to any machine learning algorithm (regression, classification, survival analysis, etc.), serving as a wrapper around the base model (Vovk et al., 2005).
- Calibration via nonconformity scores – Predictions are assessed using a "nonconformity measure," comparing new instances to calibration data to adjust prediction intervals dynamically.
- Practical utility – Widely explored in healthcare AI for risk stratification, diagnostic support, and survival analysis, where calibrated uncertainty estimates are crucial (Johansson et al., 2021).

Novelty

- First application of conformal prediction to anthropometry/obesity research.
- Provides prediction intervals instead of single point estimates.
- Addresses uncertainty – critical for clinical adoption.
- Extends prior conformal prediction work (oncology, pathology, racial disparities) to metabolic health.

Research Aim

- Develop ML models to predict WC using demographic and anthropometric data.
- Employ uncertainty quantification using conformal prediction.• Validate across diabetic (Look AHEAD) and general (NHANES) populations.
- Hypothesis: Conformal prediction provides robust WC estimates with clinical utility.

Datasets and Variables

NHANES:• Nationally representative cross-sectional dataset (2003–2016).• Anthropometry, labs, demographics • n=38,493 after exclusions.

Look AHEAD:• RCT of lifestyle intervention in T2DM patients (n=4,899 at baseline).• NIH-funded, 8-year follow-up.• Baseline data used for external validation.

Predictors:

- Age, gender, weight, height, race/ethnicity, education level.
- Race/ethnicity collapsed into Hispanic, White/Black, Other/Mixed.
- Education collapsed into five categories.
- Gender-specific models due to differences in fat distribution.

Methods

Prediction Task

- Goal: Predict WC (cm) with confidence intervals.
- Conventional ML regression: single estimates, no uncertainty.
- Our approach: Conformal prediction → intervals reflecting uncertainty.
- Adaptive intervals: Narrow for typical cases, wide for outliers/extreme cases.
- Empirical coverage: % of true outcomes within prediction intervals.
- Prediction set size: width of interval (narrower preferred).
- Conditional coverage: subgroup performance by gender, race, education, WC levels.
- External validation: Train on NHANES, calibrate/test on Look AHEAD.

Methods-Preprocessing and Model Training

Preprocessing – Concatenated NHANES (2003–2016), tabulated Look AHEAD.

- One-hot encoding for categorical variables.
- Standardized continuous variables.
- Missing data exclusions: NHANES reduced from 45,377 to 38,493; Look AHEAD complete at 4,899.

Training- NHANES: 50% train, 25% calibration, 25% test

.• Look AHEAD: 50% calibration, 50% test.

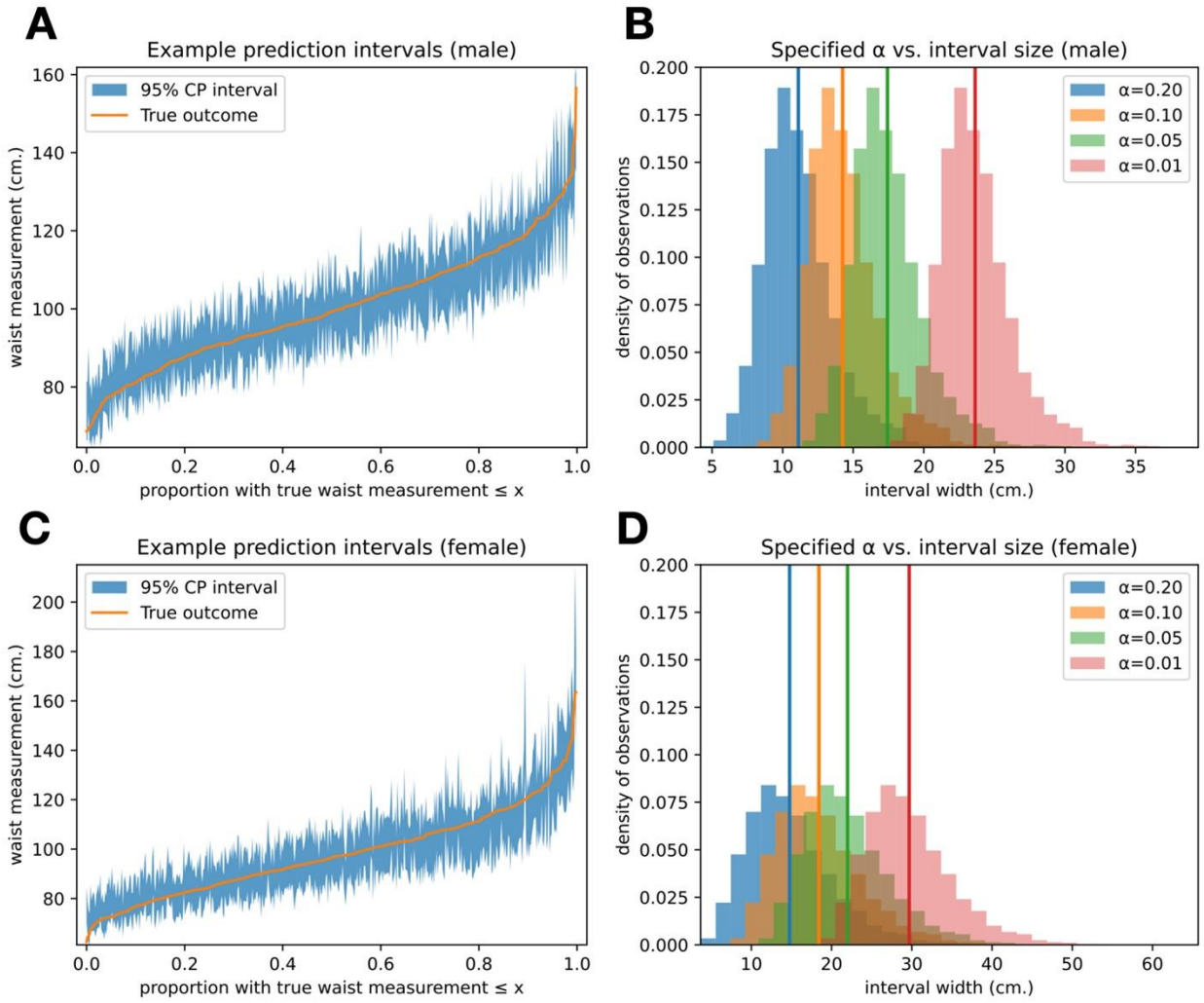
- Grid search for hyperparameters via 10-fold CV.
- Separate upper/lower models for conformal prediction.
- Point predictions with XGBoost; Shapley values for feature importance.

Results

NHANES

- Coverage: 95.5% (men), 95.4% (women).
- Adaptive interval widths: larger for extreme WC.
- Narrower average intervals with higher α values.
- Consistent performance across subgroups; only Hispanic men showed marginal under-coverage.

- **Overall Results from NHANES**
  - Achieved high coverage rates, demonstrating superior performance.

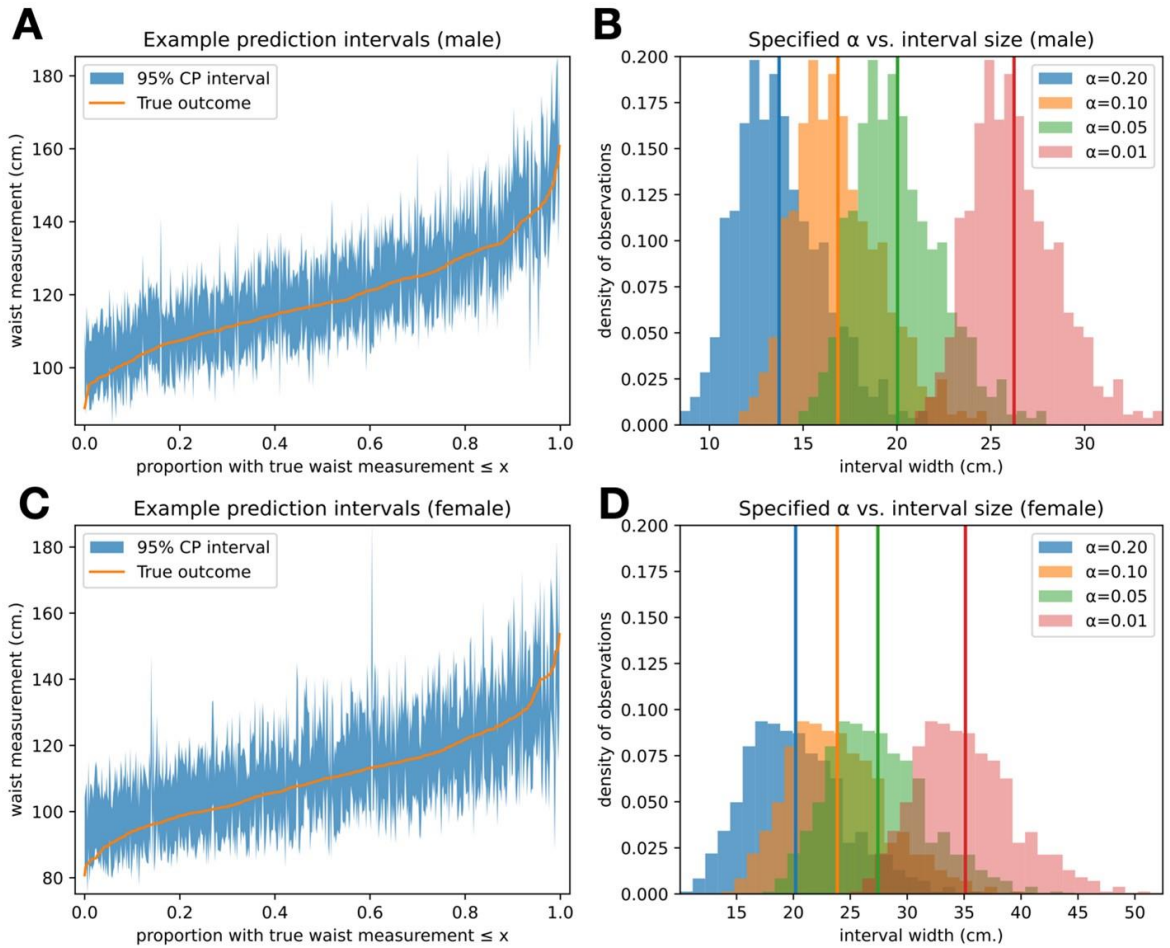


Results

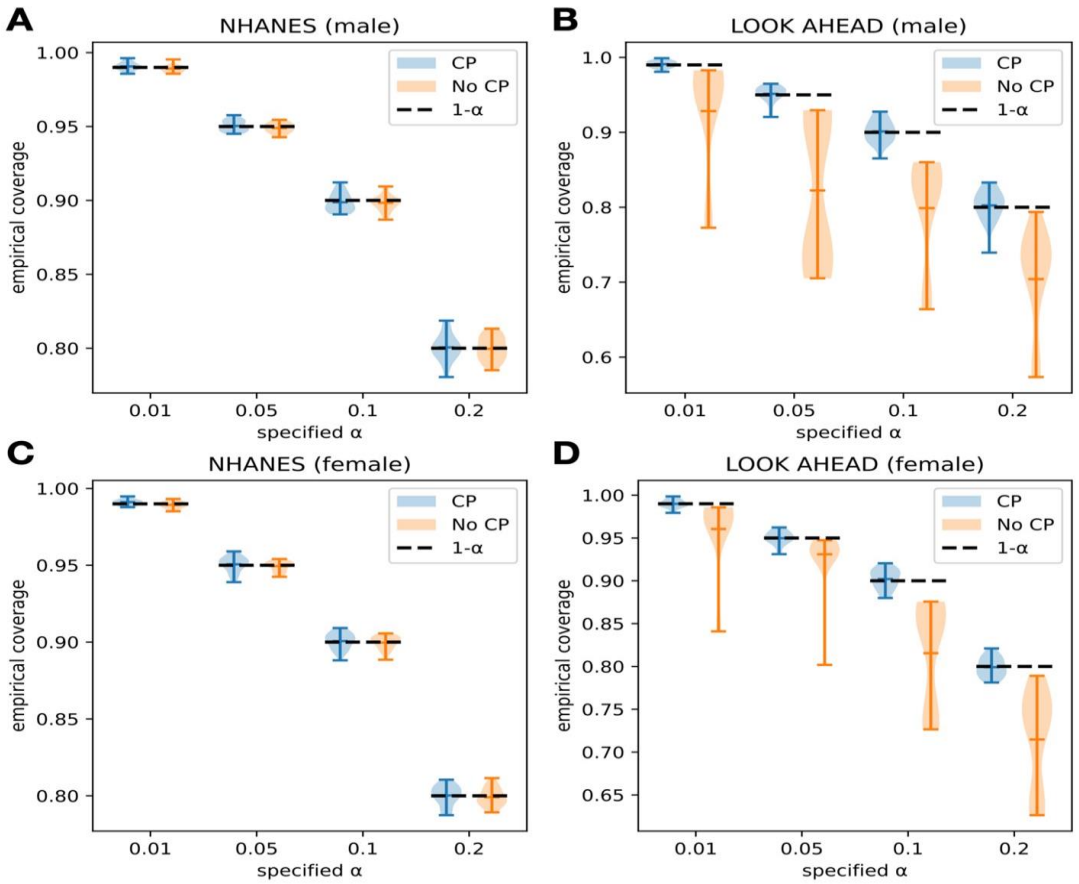
LOOK AHEAD

- Coverage: 95.1% (men), 95.2% (women).
- Interval widths larger than NHANES → reflects population shift (general vs. diabetic).
- Subgroup coverage ~95%; under-coverage only at WC extremes.
- Demonstrates robustness across domain shifts.

- **Performance Metrics**
  - Coverage rates and prediction interval widths demonstrated robustness of the model.

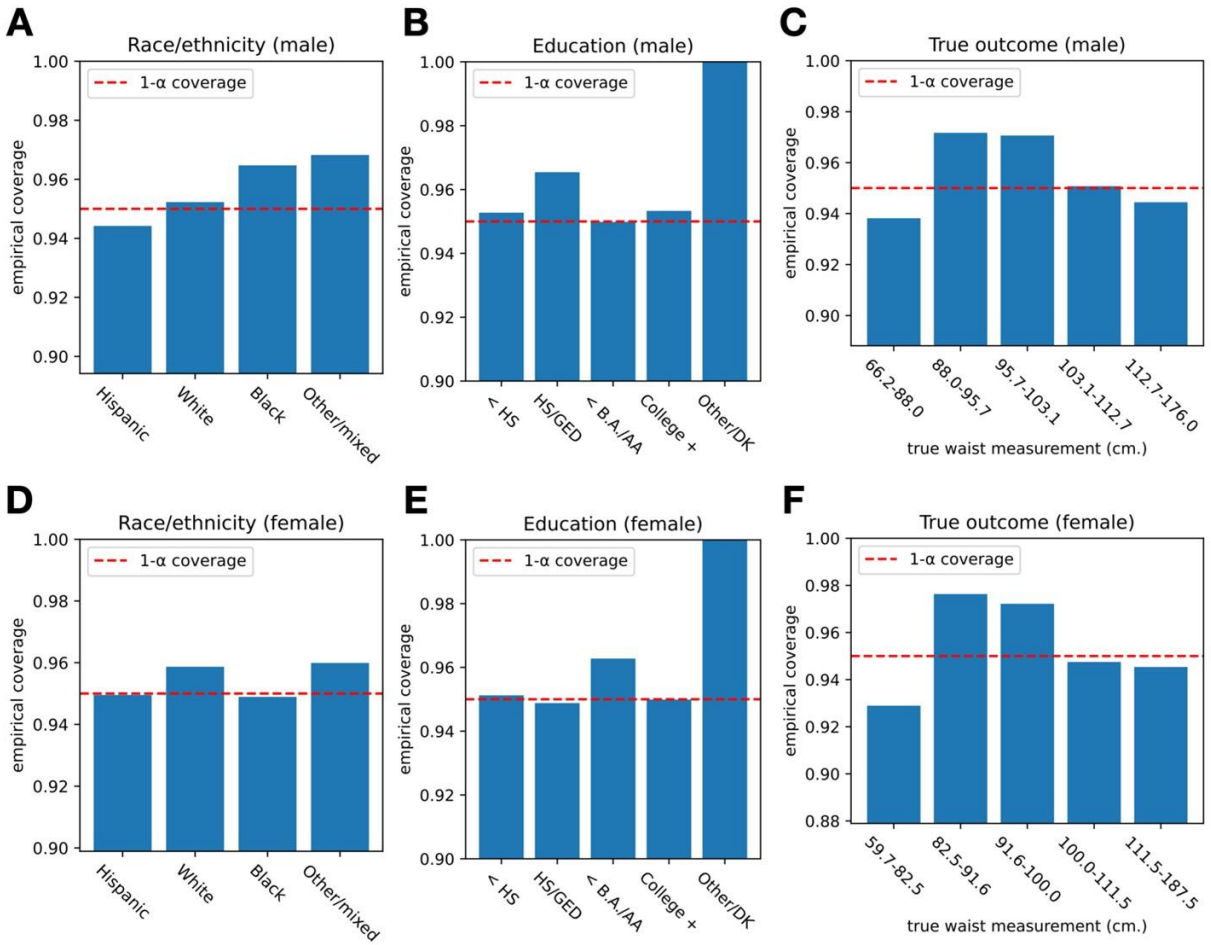


- **Validation Results from Look AHEAD**
  - Model maintained high coverage rates across different demographics, validating its consistency.



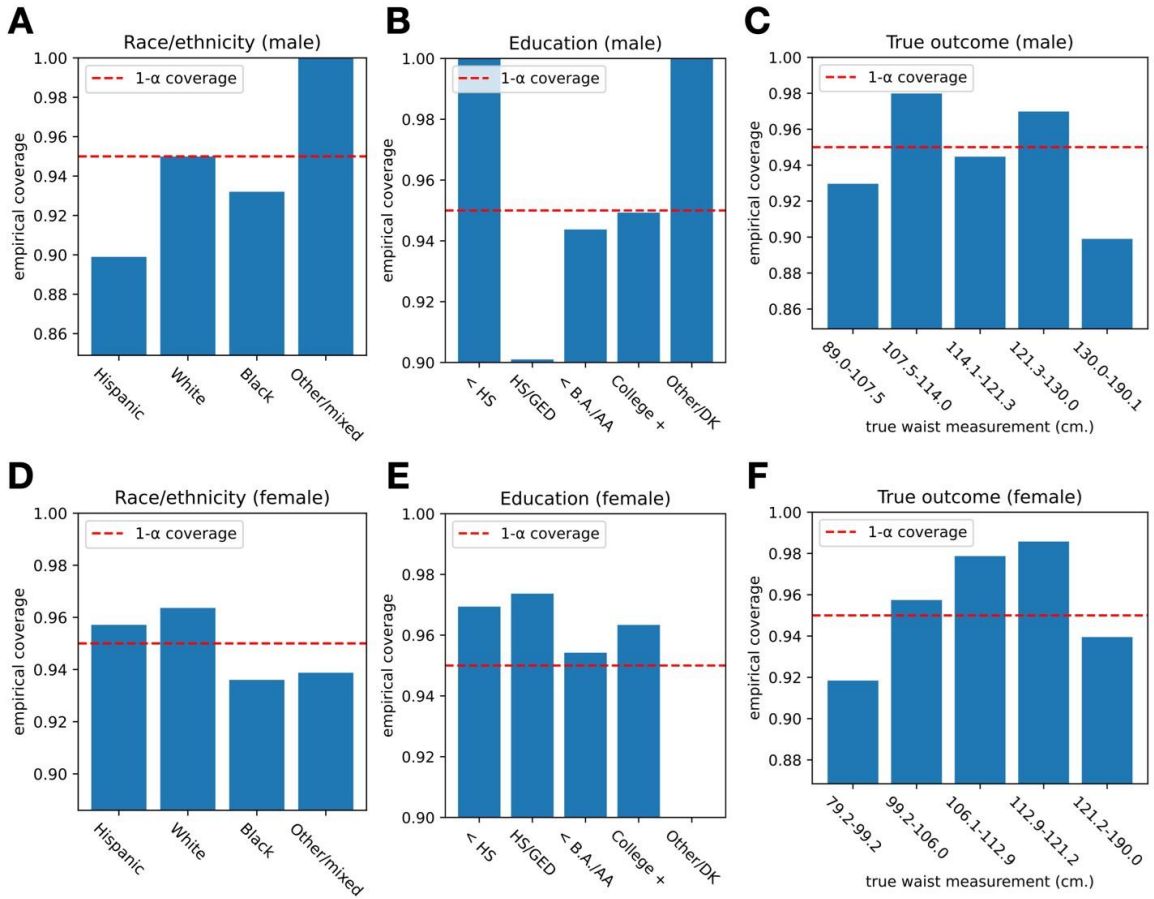
- **Overall Results from NHANES**

  Achieved high coverage rates, demonstrating superior performance.



External Validation

Confirmed the model's performance on external datasets, ensuring reliability



Results

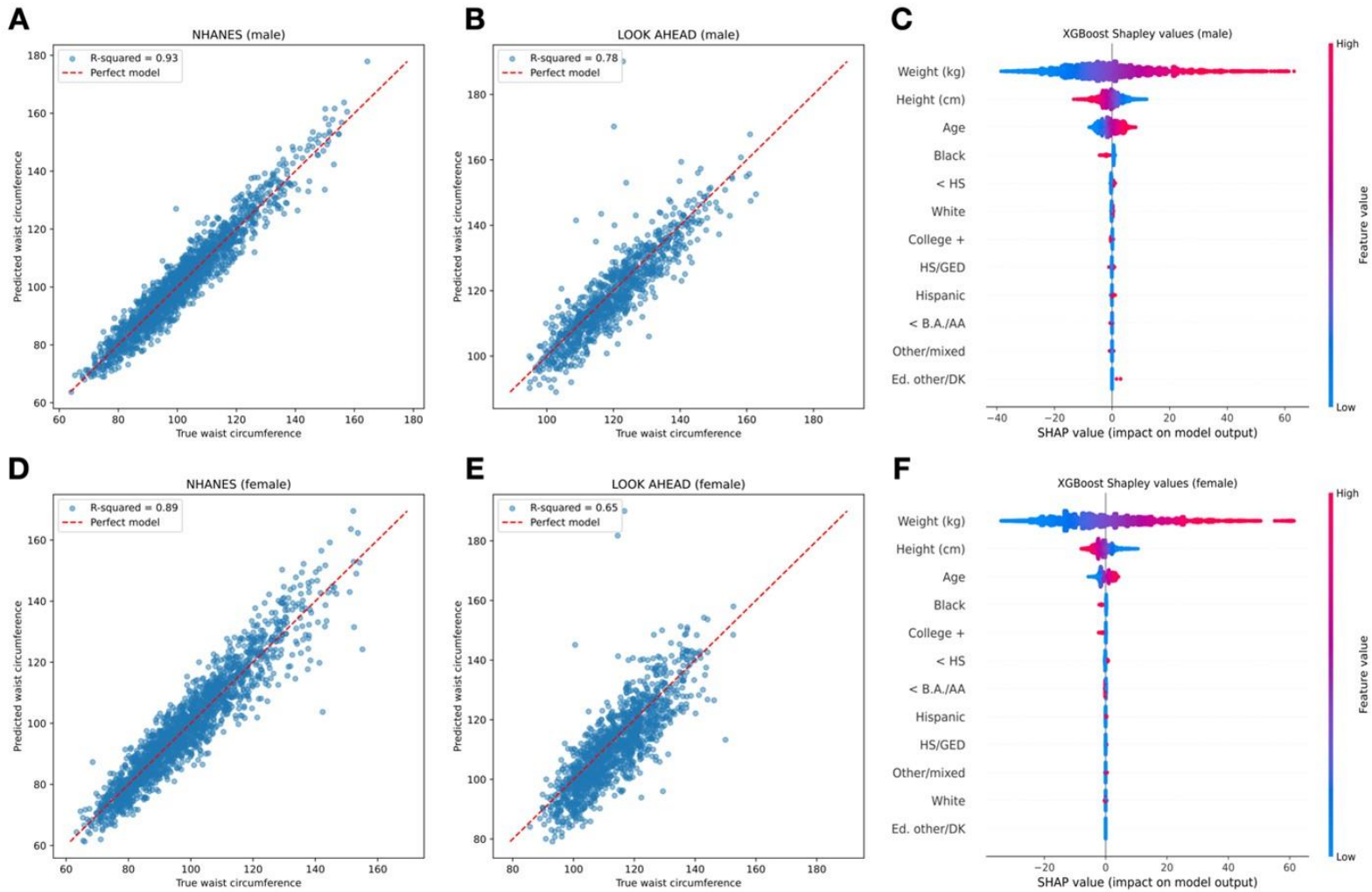
Resampling

- NHANES: Conformalized and non-conformalized coverage ≈ 95%.
- Look AHEAD: Non-conformalized underperformed (coverage <95%).
- Conformalized models recovered expected coverage after calibration.
- Confirms advantage of conformal prediction under data shift.

Point Prediction Models

- XG Boost regression: high accuracy in NHANES (R^2 strong).
- Look AHEAD: lower accuracy, especially in females (distributional shift).
- Feature importance (Shapley values):

– Weight: strongest positive predictor.
– Height: negative predictor (controlling for weight).
– Age: positive association with WC.


A
NHANES (male)
R-squared = 0.93
Perfect model
Predicted waist circumference
True waist circumference
B
LOOK AHEAD (male)
R-squared = 0.78
Perfect model
Predicted waist circumference
True waist circumference
C
XGBoost Shapley values (male)
Weight (kg)
Height (cm)
Age
Black
< HS
White
College +
HS/GED
Hispanic
< B.A./AA
Other/mixed
Ed. other/DK
SHAP value (impact on model output)
High
Feature value
Low
D
NHANES (female)
R-squared = 0.89
Perfect model
Predicted waist circumference
True waist circumference
E
LOOK AHEAD (female)
R-squared = 0.65
Perfect model
Predicted waist circumference
True waist circumference
F
XGBoost Shapley values (female)
Weight (kg)
Height (cm)
Age
Black
College +
< HS
< B.A./AA
Hispanic
HS/GED
Other/mixed
White
Ed. other/DK
SHAP value (impact on model output)
High
Feature value
Low

Discussion

Strengths:

- First application of conformal prediction to obesity research.
- Robust across datasets despite domain shift.
- Clinically interpretable with uncertainty quantification.

Limitations:

- Cross-sectional data, no causality.
- Excluded diet/physical activity.
- External validation limited to Look AHEAD.
- Measurement variability in WC remains an issue.

Clinical Implications

- Reliable WC predictions without physical measurement.
- Reduces inter-operator variability.
- Facilitates integration of WC into clinical practice.
- Improves obesity risk stratification across diverse populations.

Clinical Implications and Future Directions

Implications

- Reliable WC predictions without physical measurement.
- Reduces inter-operator variability.
- Facilitates integration of WC into clinical practice.
- Improves obesity risk stratification across diverse populations.

Future

- Expand validation across more diverse datasets.
- Incorporate additional predictors (diet, PA, biomarkers).
- Develop EHR-based decision support systems.
- Explore utility in pediatrics, geriatrics, and minority populations.

Conclusions

- WC is a critical but underutilized cardiometabolic biomarker.
- ML with conformal prediction enables accurate, reliable, uncertainty-aware WC predictions.
- Potential to transform obesity risk assessment and clinical adoption.

Acknowledgments:

- Look AHEAD participants & NIH/NIDDK.
- No conflicts of interest.

National Institute of Diabetes and Digestive and Kidney Diseases

Central Repository

Contacts

- **Dr. Adam E Gaweda -** adam.gaweda@louisville.edu
- **Dr. Juliet Emamaullee -** juliet.emamaullee@med.usc.edu
- **Dr. Prasanna Santhanam -** psantha1@jhmi.edu

Upcoming Webinar: Building Real-World Data (RWD) Linkages with a Focus on Quality and Reusability

- **Date:** September 25th from 2-3pm ET
- **Agenda:** This session is geared towards researchers interested in learning about how privacy preserving record linkage (PPRL) and real-world data (RWD) is being used for cross agency data linkage and generating real-world evidence.
  - Quality considerations/barriers to using RWD in research, such as bias, underrepresentation of populations in data, and technical issues of data harmonization and linkage
  - Considerations and best practices when generating, stewarding, and using RWD
  - Examples of data linkage implementations across federal health
  - Governance framework for data linkage
- **Scan the QR code register**

Thank You!