ACT Data Science and Informatics Core (DSI)

Pamela Shaw and Rod Walker

Kaiser Permanente Washington Health Research Institute
Pamela.A.Shaw@kp.org, Rod.L.Walker@kp.org

ACT Symposium 2026

KAISER PERMANENTE®

Data Science and Informatics (DSI) Core

Kaiser Permanente Washington Health Research Institute (KPWHRI)

David Cronkite
Chay Davis
Kristin Delaney
Hongyuan Gao
Luesa Healy
Abisola Idu
Chloe Krakauer
Jeremiah Litondo
Jennifer Nelson
Daniel Riness
Sundary Sankaran
Pamela Shaw – Core director
Rod Walker – Core director
Onchee Yu

University of Washington (UW)

Seo-Eun Choi
Laura Gibbons
Michael Lee
Duryah Mohamath
Shubhabrata ‘Joey’ Mukherjee
Karen Velderrain-Lopez

University of Wisconsin

Qiongshi Lu
Zijie Zhao

Brown University

Rebecca Hubbard

Alumni (KPWHRI, UW, UPENN)

Nolan Cole
Peter Frawley
Dustin Key
Serah Kimachia
Brandon Klinedinst
Bonnie Lind
Kristin Linn
Jingxin Liu
Elizabeth Sanders
Jenny Shen
Kyle Tracy
Yinxiang Wu
Jing Zhou
Weiwei Zhu

DSI Core Specific Aims

Aim 1. Manage a high quality and secure ACT Repository of clinical, biological, and EHR data

Aim 2. Provide comprehensive statistical analysis support to translate these data into research

Aim 3. Promote broad use of the data via efficient data sharing with external investigators

Aim 1: Database infrastructure

- Designed and developed a modernized participant tracking database, transitioning away from a legacy Access system
- New web-based user interface (UI) using .NET development platform
- New SQL server relational database back-end
    - Created stored procedures and views to support UI and reports, decoupling front-end / back-end objects for easier maintenance
    - New ability to track and audit data changes across users and time
- Event system assists tracking as a participant moves through the study
- Automated processes for transferring data into the system from external sources, catching data anomalies, and making internal data updates to move participants through the steps of the study
    - Framework to build other ancillary processes linked to the database
- New reporting using SSRS
- Aligned with new developer standards at KPWHRI

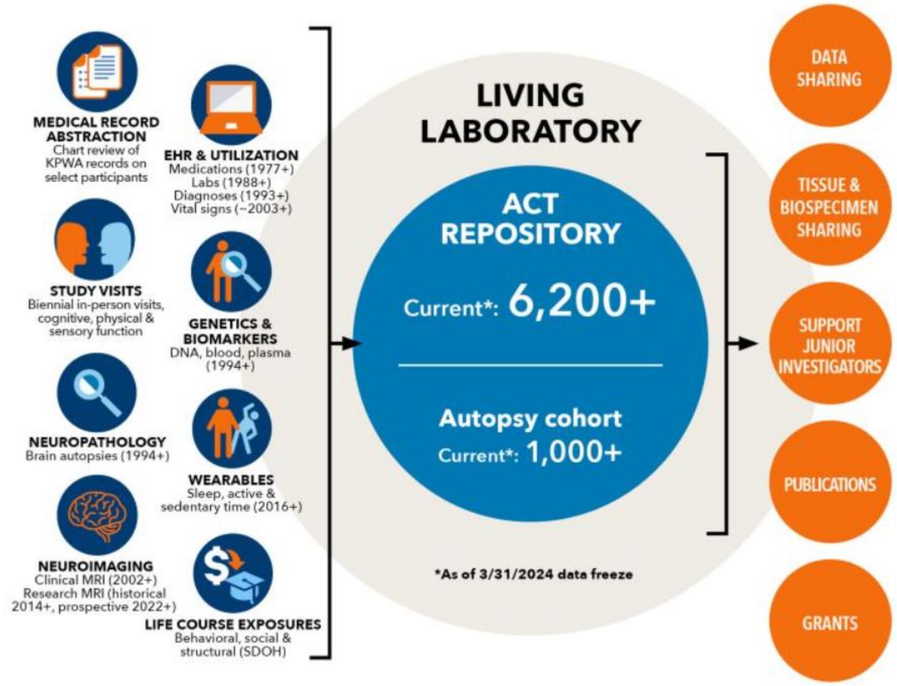
Aim 1: Digital data capture

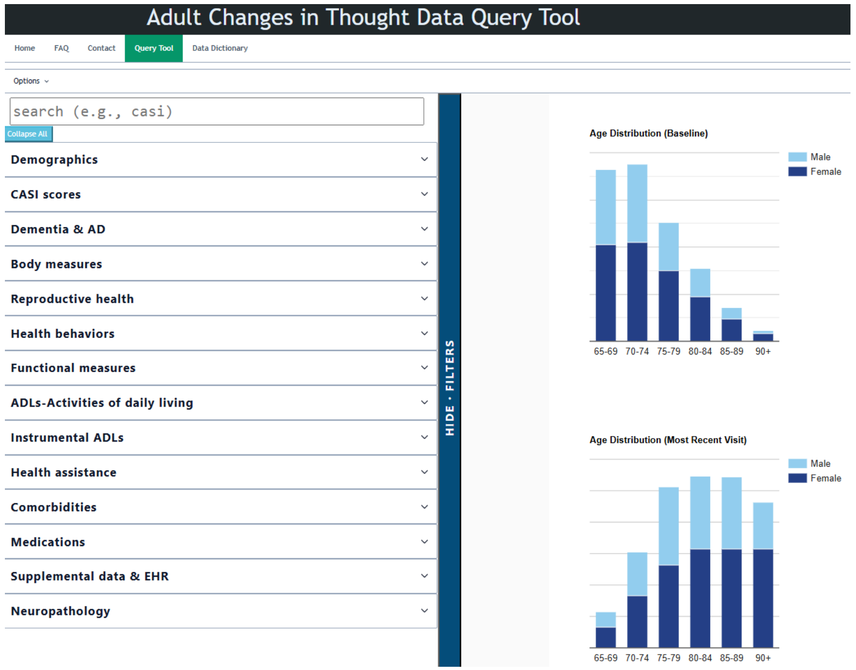
- Designed REDCap instruments to replace legacy data entry processes for paper ACT study visit forms
  - Supports more built-in data quality control
  - Allows simultaneous data entry from multiple users across different locations
  - Additions or modifications due to recommended changes from the ACT Forms Workgroup easier to implement in the new system
  - Automated programs developed to upload new participant data from ACT Central and download data from REDCap to integrate with our historical forms data

- Ongoing work to adapt these instruments for use in Clinical Core workflows and integration with a new digitally collected cognitive battery

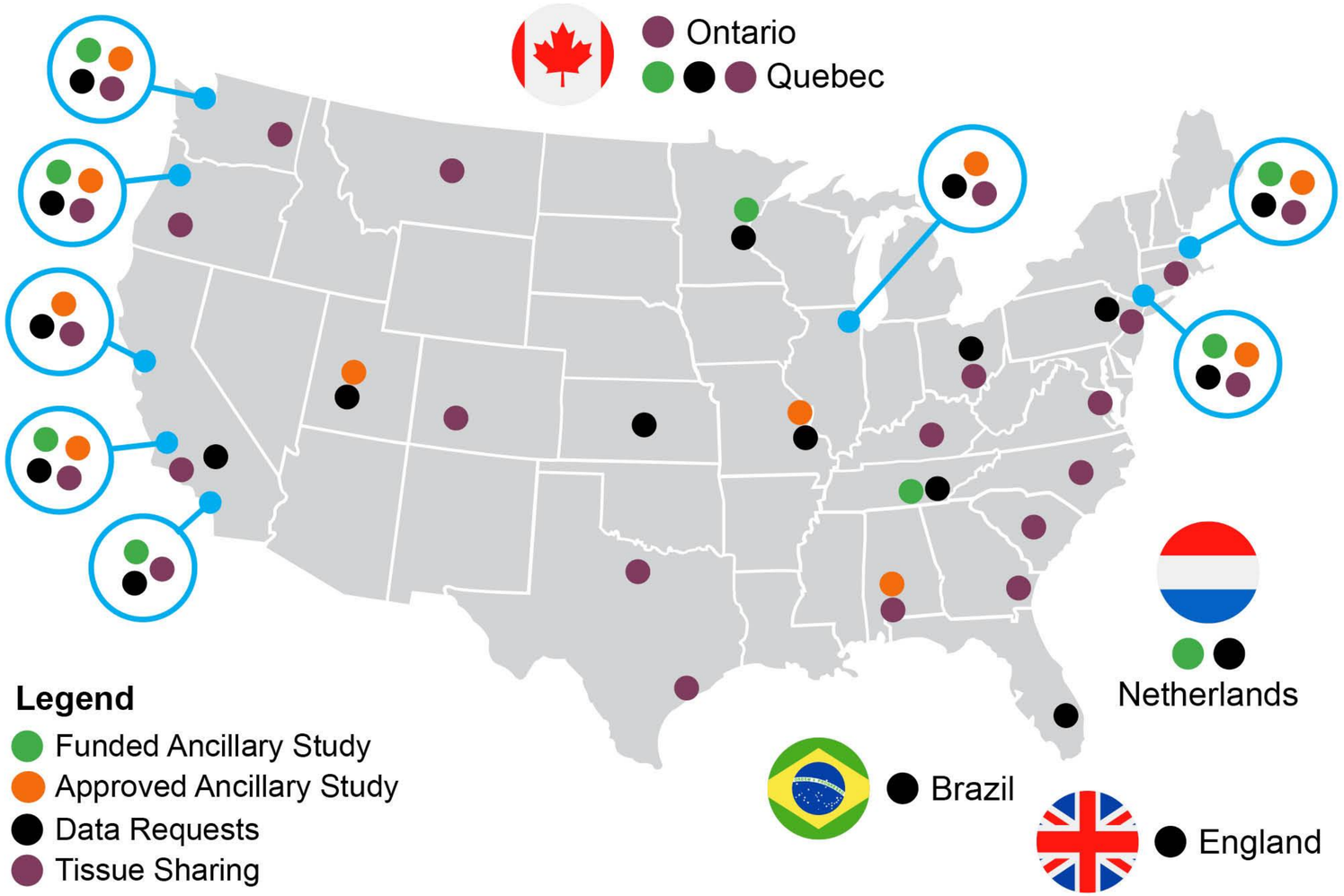
Aims 1,3: Analytic data curation and sharing

- Continued the key DSI Core process of preparing ACT Analytic Data ‘Freeze’
  - 3 major releases during the cycle (working on 4th)
- Significant overhaul of process and documentation
  - Major expansions to included data elements
  - More user-friendly and detailed data dictionary
  - ~100 different content links with descriptions spanning 2000+ variables
- Enhanced the Data Query Tool on ACT website https://www.actagingresearch.org/index.php/resources/data-query-tool
- These activities are critical for **Data Sharing (Aim 3)**





Aim 3: Data sharing



ACT has fulfilled:

- >150 data requests from >30 institutions around the world during the current cycle
- >450 tissue requests
- >30 ancillary studies approved with ~half of those funded
- >160 publications using ACT data or resources since 2021

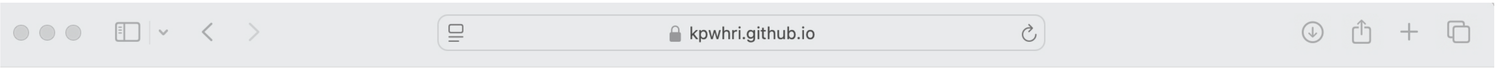
https://kpwhri.github.io/actstats/

Aim 2: ACT Statisticians Interest Group

- ACT Statisticians Interest Group (SIG) Founded in 2021, with launch of U19
- Purpose of the ACT SIG
  - Develop statistical guidance for handling the analytical complexities present in the ACT Study, both for internal and external researchers
    - Guidance is meant to improve rigor and help broadly disseminate recommended conventions or analysis approaches across ACT analyses, such as outcome/exposure definitions or handling of missing data
  - Provide a twice-monthly brown bag discussion for statisticians broadly involved in ACT Science and data analyses for works in progress or manuscripts, grant proposals, etc.
  - Develop a working group website as a way to disseminate resources

  https://kpwhri.github.io/actstats/

Aim 2: New statistical guidance and resources available



ACT Statistical Guidance

Home · Our Work · Resources · Contact

Introduction

Welcome to the ACT Statisticians Working Group website. This site is meant to be a source of helpful information for anyone analyzing or thinking about analyzing data from the Adult Changes in Thought (ACT) study. We also provide a way for you to contact us with your questions.

ACT Statistical Guidance documents

The ACT data have a few complexities for which an analyst may appreciate some guidance, which is informed by statisticians who have a more intimate knowledge of the data generation and/or regularly have addressed these issues in analyses of ACT data.

1. Weighting methods to handle selection bias due to missing data
2. Generation of item response theory (IRT) based CASI scores
3. A review paper for methods to analyze the 24-hour activity cycle data, which includes a supplemental appendix

For more information about the ACT study, including data query tools, data documentation, and other study documents visit https://www.ACTagingresearch.org.

ACT
ADULT CHANGES IN THOUGHT STUDY

Kaiser Permanente Washington
Health Research Institute

Aim 2: Contributions to Project 1: Integrating new statistical techniques

Compositional data analysis

- We explored 3 different methods to analyze 24-hour activity HAC compositional data and review paper: isometric substitution modeling, compositional data analysis (CoDA), latent class analysis (Wu et al 2023)

Growth mixture modeling (GMM)

- We applied GMM to group participants by their trajectory profiles (i.e., of cognitive function and Charleson comorbidity index) and to examine how different profiles were associated with other outcomes/exposures (e.g. Physical activity)

Joint modeling

We implemented joint modeling as a way to handle informative missing data patterns in cognitive and physical function trajectories due to dementia or death

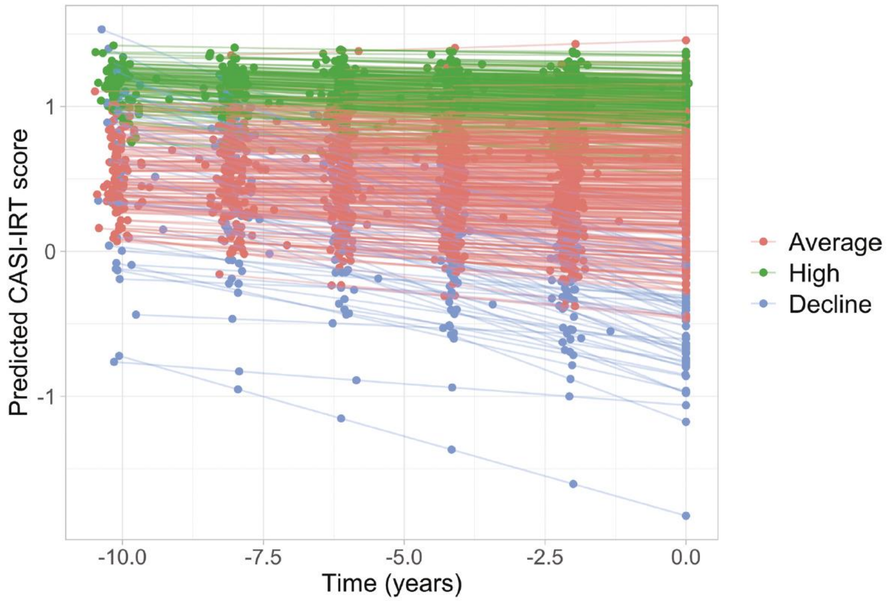
Improved rigor for two-stage regression

- Two stage regression, in which individual trajectories are fit and then estimated features are analyzed in association analyses, is frequently done incorrectly
- We applied state of the art statistical techniques to perform these analyses correctly by incorporating the measurement error of the estimated features

Aim 2 P1 Highlight: How does prior 10-year trajectory of cognition and physical function relate to current activity?

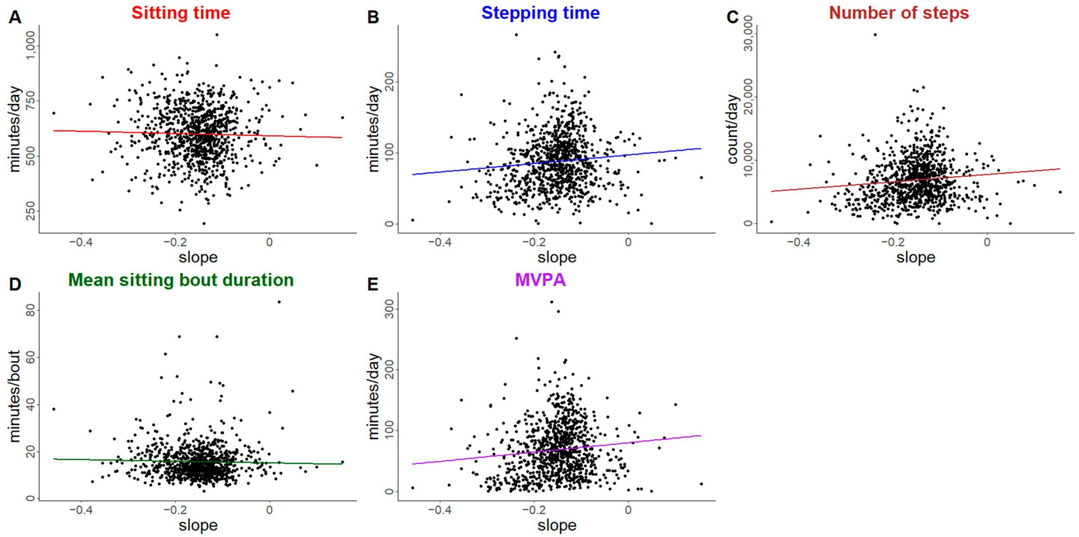
Analysis of cognitive trajectories (Rosenberg et al 2024)

- Growth mixture modeling found 3 trajectory profiles
- Applied 3-step LCA approach (Vermunt 2010) that corrects for class uncertainty
- Declining cognition predicted lower PA providing some evidence of a reverse relationship between PA and cognition in older adults



Analysis of physical function trajectories (Greenwood-Hickman et al 2025)

- We fit physical function (sPPF score) trajectories using mixed effects modeling and extracted individual slopes and intercepts (BLUPs)
- We modeled associations between these slopes/intercepts with 24HAC, using bootstrap to address uncertainty and IPW to address selection bias
- Worsening physical function is associated with lower PA and higher SB, but not sleep or light activity



Aim 2: Contributions to Projects 2 and 3- the highlights

Project 2 – Cognitively defined AD-subgroups

- Updated data pipeline for identifying and categorizing incident AD-dementia cases into cognitively defined subgroups
- Provided statistical support to address selection issues in multiple contexts
  - Clinically assessed antemortem MRI measures
  - Neuropathology from brain donation

Project 3 – Translational pharmacoepidemiology and dementia

- Created exposure and outcome variables by integrating extensive data on medication, blood pressure and comorbidity data from chart review, ACT study visits and KPWA electronic health records, from participants' earliest KPWA enrollment to recent years on ACT
- Onchee Yu served as Site-PI and sole statistician, guided study design and all statistical analysis for three manuscripts addressing Project's scientific aims

Other DSI scientific contributions: a few highlights

- **Life Course Core support**
  - Consulting/mentoring on statistical methodology, analysis code, and manuscripts from mostly student-led projects on associations between life course components—living arrangement and location, internet use, childhood experiences—and cognition/neuropathology/neuroimaging
  - One project has motivated current statistical methods work in DSI to address coarsened data on time spent in living arrangements
- **UW statisticians broad support of ACT 19 science**
  - Dr. Laura Gibbons has made several statistical contributions, including helping researchers link ACT cognitive measures to those in other studies and use principled methods to extrapolate results on the autopsy subset to the broader ACT cohort and ACT study results to the broader Seattle metro population >65
- **UPenn/Brown statistical methods contributions**
  - Dr Rebecca Hubbard's team has led methods work to address complexity in ACT data, including PhD student Jenny Shen developing novel estimators of the average treatment effect (ATE) to combine surrogate EHR data on the entire cohort with gold-standard outcomes from a subset subject to selection bias

Future work

Infrastructure + Data sharing

- Transition to real-time digital data capture (at time of study visit)
- Continue enhancements to study tracking and reporting tools
- Increase frequency and content of data freezes

Statistical guidance and contributions to science

- Construction of innovative exposure constructs for the new science in ACT renewal,
  - Multimodal exposure assessment detailing different aspects of hypertension, diabetes and atrial fibrillation
- Continued improvements in statistical rigor, including development of new guidance
  - Appropriately handling measurement error in exposures and/or outcomes, such as medication exposures or the uncertainty in the predicted CASI IRT score
  - Principled decision making when performing analysis of time-varying exposures