Development of Complex Data Platform for the Stanford Technology Analytics and Genomics in Sleep (STAGES) Study

Eileen B. Leary¹, Rebekka K. Seeger-Zybok¹, Marcin Mazurek², Oleksii Voronoi², Benjamin Lawlor², Cheryl Liane Stephenson², Clete Kushida¹, Emmanuel Mignot¹

1. Center for Sleep Sciences & Medicine, Stanford University, Palo Alto, CA  2. Prometheus Research, LLC, New Haven, CT.

INTRODUCTION
Sleep is critical to both physical and mental health. Sleep deprivation impairs performance, judgment, mood, and is a preventable contributor to accidents. We all experience sleep; yet why we sleep and how the brain generates sleep remain biological mysteries because we lack the tools and data needed to gain a comprehensive picture of sleep.

To address this need, the Stanford Center for Sleep Sciences and Medicine, with assistance from a multi-disciplinary team, has set up the infrastructure for a large-scale project to develop and disseminate essential tools and data to the scientific community to advance the field of sleep medicine. These data and tools will be crucial for our understanding of the genetic architecture of sleep and will improve detection, treatment, and prevention of sleep disorders. Our goal is to be a catalyst for change in the sleep field.

METHODS
The Stanford Technology Analytics and Genomics in Sleep (STAGES) study is a prospective cross-sectional, multi-site study which is outlined in Figure 1.

3,000 sleep clinic patients will be enrolled at 17 data collection sites from five centers including: Stanford University (Stanford, CA), Mayo Clinic (Rochester, MN), MedSleep (~6 locations in Ontario, Canada), St. Luke’s Hospital, (Chesterfield, MO); and Geisinger (~7 locations across Pennsylvania and New Jersey). New data collection sites are being added.

A comprehensive dataset will be collected, catalogued and stored using a robust data platform (see Figure 2), including:
- Alliance Sleep Questionnaire (ASQ) an on-line sleep/medical history questionnaire
- In-lab nocturnal polysomnography data (one night PSG)
- Computerized Neurocognitive Battery (CNB) developed by UPenn
- Actigraphy using the Huami Actiwatch device over 2 weeks with sleep diary
- Genetics - Genome wide association data
- Stored biological samples (DNA, plasma, serum) for future biomarker research
- 3-D facial image
- Medical record data

DATA PLATFORM:
The STAGES team partnered with Prometheus Research, LLC to design and build a customized platform using their open source integrated research registry, RexStudy. Requirements included: (1) mechanisms to securely transfer data from multiple disparate data sources and match to correct subject, (2) patient portal to enroll participants via electronic consent and accommodate electronic case report forms, (3) extensive data validation and error reporting, (4) granular access privileges for different user classes, (5) sophisticated querying and analytic tools, and (6) ability to export all data for complete portability.

Each phase of the project, including creation of custom APIs, was heavily tested in a user acceptance testing (UAT) environment before being deployed.

MACHINE LEARNING TOOLS:
Machine learning and signal processing techniques will be used for this project to create software and algorithms that will be employed to streamline PSG analysis, extract meaningful sleep phenotypes, and standardize analysis in large samples. We will validate this software in smaller, existing cohorts and use the software in the large 30,000 sample in conjunction with GWAS. All resources developed will be shared.

RESOURCE SHARING:
The project’s philosophy is to make all resources generated by the study available to the widest potential audience with the least number of restrictions. Therefore, all data, biological samples, and electronic products collected or developed under the STAGES project will be available for any interested researcher.

RESULTS
The STAGES data platform currently holds data from more than 1,700 enrolled research participants. Fifty members of the STAGES team have accounts for entering, analyzing, or monitoring the 75 plus data tables, and so far, there are more than 50 queries for data monitoring and quality control.

We have also built a multi-dimensional e-consent system that allows coordinators to consent subjects using 3 different pathways (1) Before coming to their appointment, participant review and complete the consent electronically through the Prometheus Patient Portal, (2) During their visit, a staff member has the participant complete the consent electronically through the Prometheus site, (3) During the visit, the consent is completed on paper and then uploaded to Prometheus.

CONCLUSION
The STAGES data platform was successfully deployed and is being used to collect, monitor, and analyze data from multiple external sources. We are continuing to add functionality to enhance the system’s utility. All study data will be shared with the scientific community. In parallel, we are developing analytical tools such as machine learning and new statistical methods that will assist in the interpretation of these data. Access to these data and tools will spark new research opportunities and genetic analysis, which will result in new diagnostic biomarkers for sleep disorders and a better molecular understanding of sleep regulation.

SUPPORT/CONTACT INFORMATION
This project is funded by the Klarman Family Foundation.
To learn more, contact:
Principal Investigator, Emmanuel Mignot, MD, PhD at mignot@stanford.edu
Project Director: Eileen Leary, MS, RPSGT at leary@stanford.edu