

Breakout Session 5:

National Sleep Research Resource

Dr. Shaun Purcell

*Associate Professor, Brigham & Women's Hospital, Harvard Medical
School*

The National Sleep Research Resource (NSRR)

NIH STRIDES
January 2024

Screenshot of the National Sleep Research Resource (NSRR) website. The page features a dark blue header with the NSRR logo and navigation links. Below the header, there are two main buttons: "EXPLORE DATASETS" and "SHARE YOUR DATA". A blog post is featured with an illustration of a child sleeping in a bedroom at night. The blog post title is "Exploration of sleep disturbances in children and adolescents with and without autism in a paediatric sample referred for polysomnography". The footer includes statistics: 47,841 individuals represented, 7.97 TB stored on the resource, and 1 PB shared with researchers.

Shaun Purcell
smpurcell@bwh.harvard.edu
<http://zzz.bwh.harvard.edu>





NATIONAL
SLEEP RESEARCH
RESOURCE

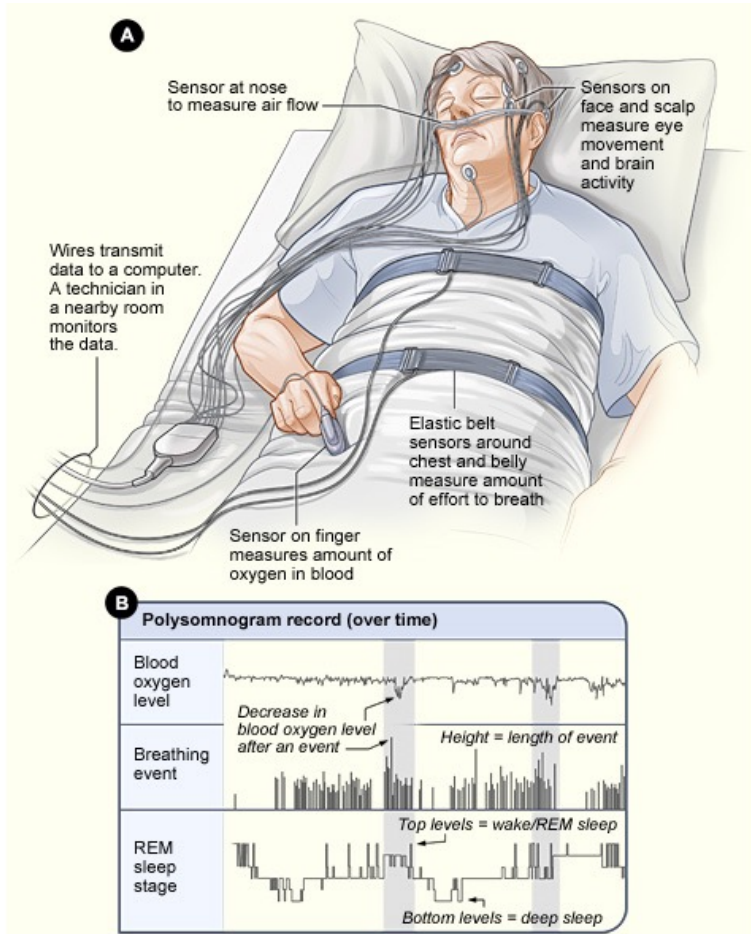
ADVANCING SCIENCE GLOBALLY
THROUGH DATA SHARING

“The National Sleep Research Resource (NSRR) offers free web access to large collections of de-identified physiological signals and clinical data elements collected in well-characterized research cohorts and clinical trials.”

Founded in 2014 by **Dr. S Redline**

Funded by the **National Heart, Blood and Lung Institute** (Resource Grant → NIH contract)

Polysomnography signal data



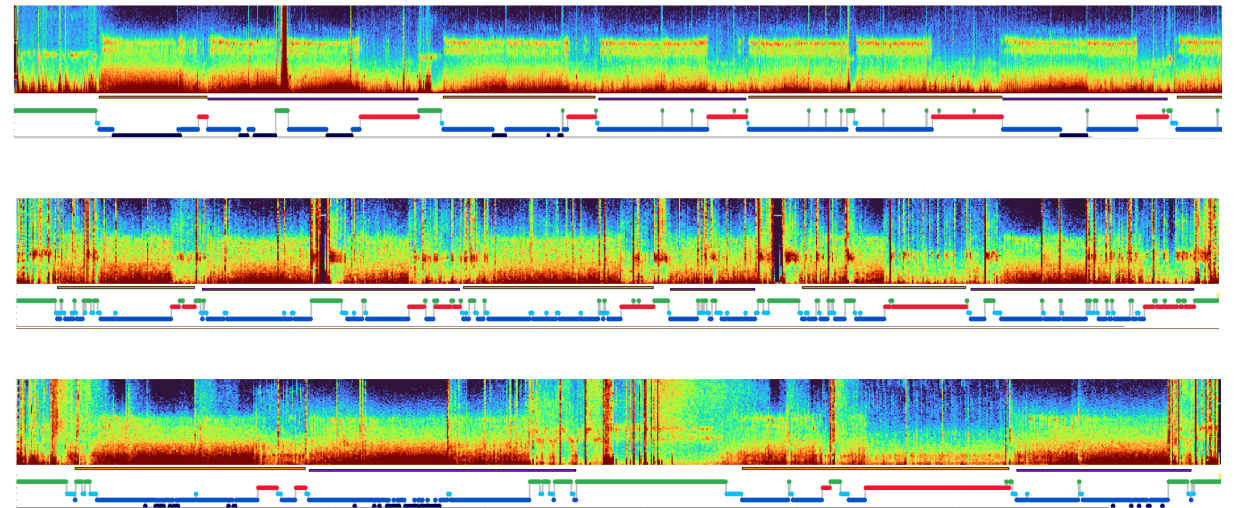
Classically, PSG data often reduced to only ~dozen common metrics

- sleep stage duration, onset latency, apnea hypopnea index, etc

Rich, multi-modal & dynamic raw signal data

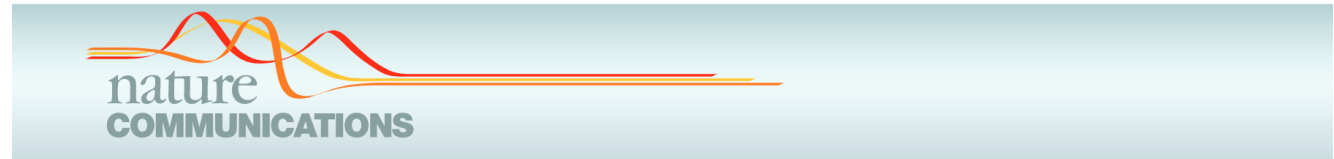
- often requires specialized tools/analytic approaches (historically, often via in-house *ad hoc* Matlab scripts)

Whole night EEG spectrograms & hypnograms for 3 individuals



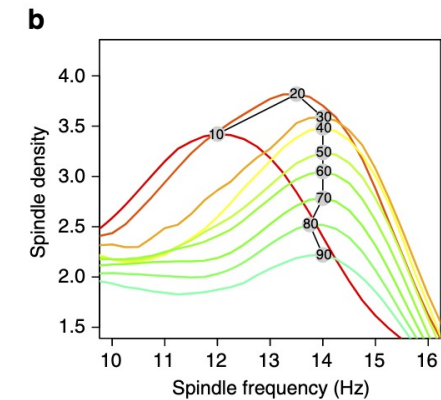
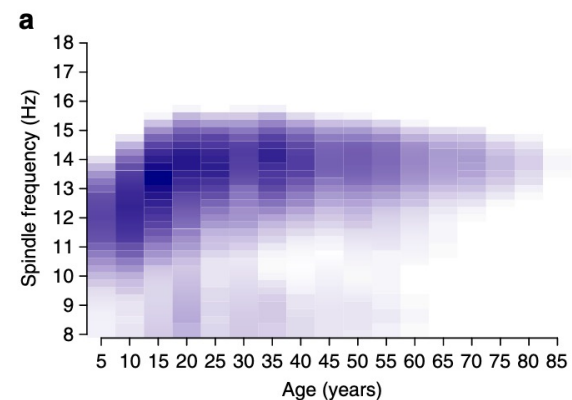
Scaling up sleep science

- NREM EEG microarchitecture
- $N > 10,000$ aged 2-90
- Captures person-to-person variability in markers of sleep & brain function/development
- Opportunities to link to genetic risk

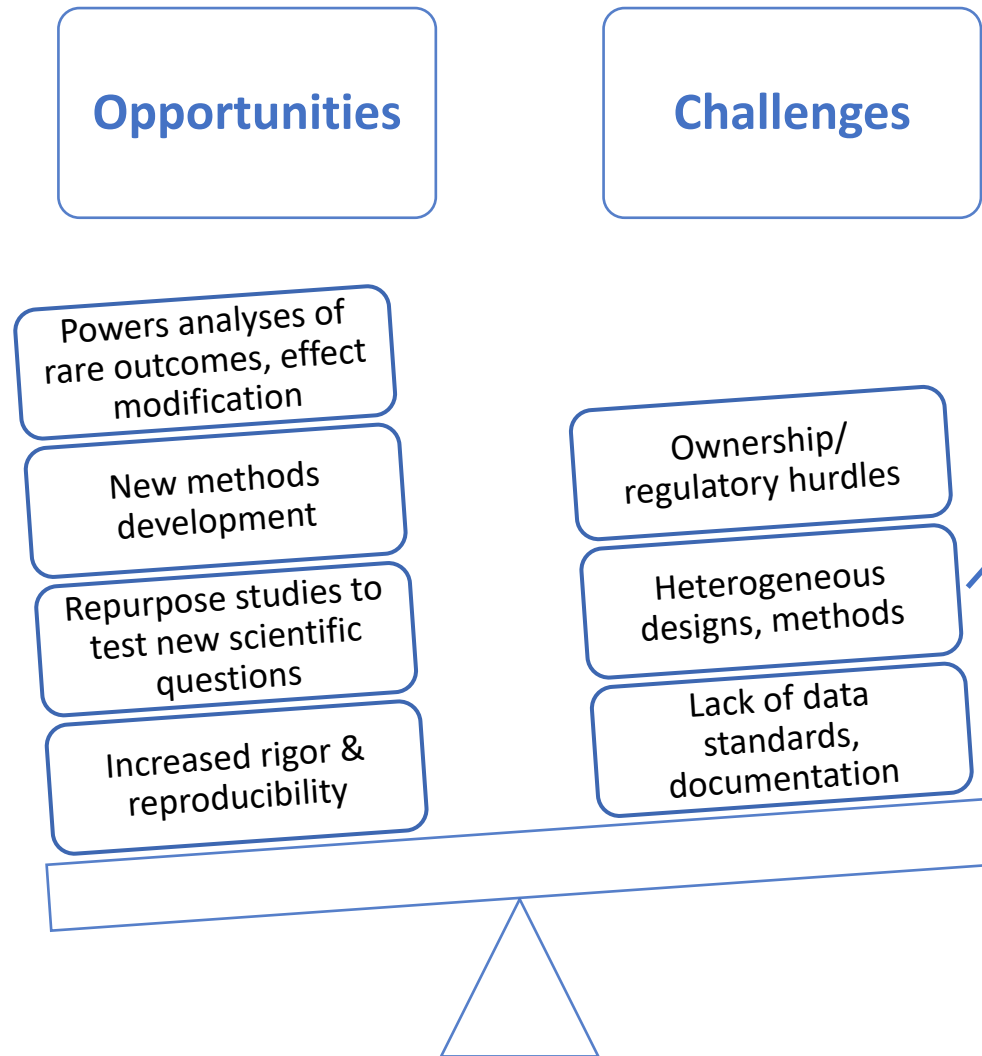


Characterizing sleep spindles in 11,630 individuals from the National Sleep Research Resource

S.M. Purcell^{1,2,3}, D.S. Manoach^{2,4,5}, C. Demanuele^{2,4,5}, B.E. Cade^{6,7}, S. Mariani^{6,7}, R. Cox^{2,8}, G. Panagiotaropoulou^{2,4,5}, R. Saxena^{9,10,11}, J.Q. Pan¹², J.W. Smoller^{2,4,13}, S. Redline^{2,6,7,*} & R. Stickgold^{2,8,*}



Data archiving & aggregation: important but often challenging



End user's perspective:

Combining different datasets can be logistically/technically challenging

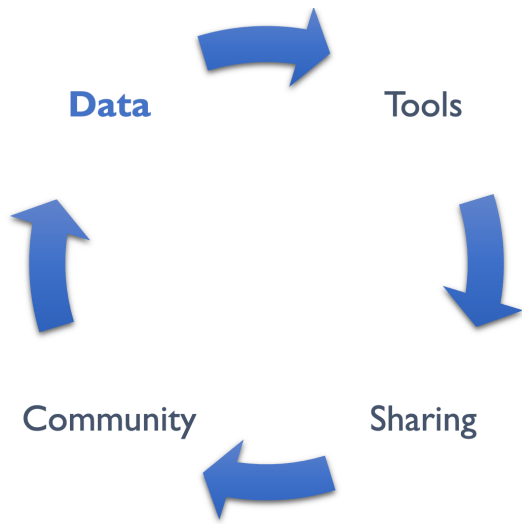
Subtle biases can be amplified in large datasets

The original studies were likely not specifically designed to answer your current research question



NSRR's attempts to alleviate some of these issues:

- harmonized & well-documented datasets
- tools to work with large datasets
- technical support



~ 50,000 individuals

> 30,000 full PSGs

~7000 actigraphy studies

2 TB of data shared weekly

4,643 Data Access Use Agreements

Browser window: Datasets - Sleep Data - National X | <https://sleepdata.org/datasets>

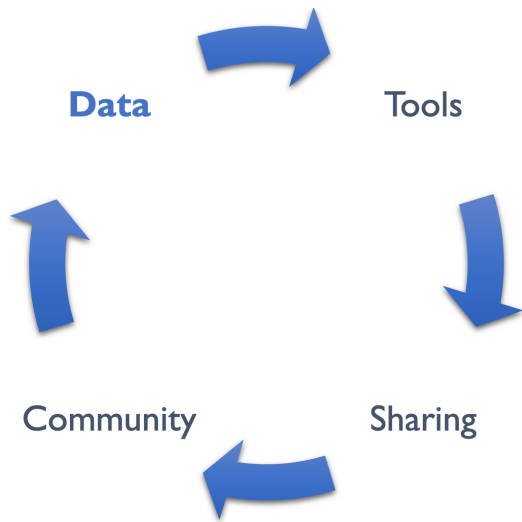
Navigation: NSRR About **Datasets** Tools Forum Blog Share data Search Sign in Sign up

Datasets

datasets << 1 to 25 of 26 >>

Filters: All Ages | All Data | Featured | [Browse the Cohort Matrix](#)

	SUBJECTS	AGES	TIMEFRAME	DATA
Apnea Positive Pressure Long-term Efficacy Study	1,516	18 - 84	2003 - 2008	Polysomnography
5,548 Files · 144 GB 353 Variables Request Data Access				
Efficacy Assessment of NOP Agonists in Non-Human Primates	5	14 - 19	2019	Polysomnography
40 Files · 2.85 GB Request Data Access				
Maternal Sleep in Pregnancy and the Fetus	106	18 - 42	2015 - 2019	Polysomnography
444 Files · 92.8 GB 37 Variables Request Data Access				
Apnea, Bariatric surgery, and CPAP study	49	26 - 64	2011 - 2014	Polysomnography
441 Files · 45.4 GB 108 Variables Request Data Access				
Assessing Nocturnal Sleep/Wake Effects on Risk of Suicide	971	18 - 52	2020 - 2021	Questionnaires Only
19 Files · 3.43 MB 301 Variables Request Data Access				



> 5,000 defined variables

Ongoing harmonization of variables across studies, mapping to CDEs

Extensive documentation on study design

NSRR_BMI - APPLES Variables - X

https://sleepdata.org/datasets/apples/variables/nsrr_bmi

NSRR About Datasets Tools Forum Blog Share data

Search Sign in Sign up

Apnea Positive Pressure Long-term Efficacy Study

Docs Files Variables

apples > variables > Harmonized > Anthropometry > nsrr_bmi v0.1.0

NSRR_BMI ★

Overview

GRAPHS

[Histogram](#)

MORE

[Related](#)

[History](#)

Name: nsrr_bmi

Label: Body mass index (BMI)

Description: Harmonized by the NSRR team to align with [TOPMed](#) and [BioDataCatalyst standards](#). Source: [bmbiqlquan](#)

Units: kilograms per square meter

Type: numeric

Tags: [bmi](#)

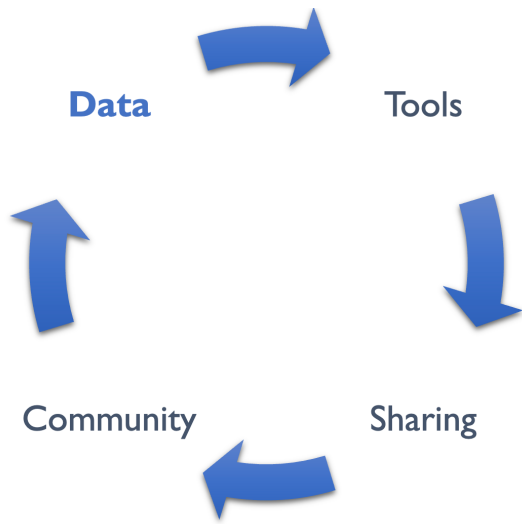
Body mass index (BMI) By Visit

BMI Range (kg/m²)	Subjects (Baseline)
17 to 21	24
21 to 26	287
26 to 31	492
31 to 35	279
35 to 40	217
40 to 45	119
45 to 49	36
49 to 54	23
54 to 58	23
58 to 63	6
63 to 68	2
68 to 72	4

Visit (Number) vs Body mass index (BMI)

Visit	N	Mean	StdDev	Median	Min	Max	Unknown	Total
Baseline (BL)	1,512	32.1	± 7.8	30.4	16.8	72.4	4	1,516
Clinical Evaluation (CE)	-	-	-	-	-	-	1,516	1,516
Diagnostic Visit (DX)	-	-	-	-	-	-	1,516	1,516
CPAP Titration Visit (CPAP)	-	-	-	-	-	-	1,516	1,516
Two Month Post-CPAP Neurocognitive Visit (2M)	-	-	-	-	-	-	1,516	1,516
Four Month Post-CPAP Neurocognitive Visit (4M)	-	-	-	-	-	-	1,516	1,516
Six Month Post-CPAP Neurocognitive Visit (6M)	-	-	-	-	-	-	1,516	1,516

© 2023 National Sleep Research Resource Privacy Contact Us | Funded By NHLBI



Raw physiologic signals (EDF) & annotations on 10,000s of individuals

In total, ~30 years' worth of multi-modal sleep signal data

APPLES Files - Sleep Data - Nat X

https://sleepdata.org/datasets/apples/files/polysomnography

80% ☆

NSRR About Datasets Tools Forum Blog Share data

Search Sign in Sign up

Apnea Positive Pressure Long-term Efficacy Study

Docs Files Variables

apples files polysomnography

1 to 100 of 2,208

apples-130001.annot	2023-02-27 13:49:03	201 KB
apples-130001.edf	2023-02-22 16:10:58	59.9 MB
apples-140003.annot	2023-02-27 13:49:03	228 KB
apples-140003.edf	2023-02-22 16:10:59	58 MB
apples-140004.annot	2023-02-27 13:49:03	195 KB
apples-140004.edf	2023-02-22 16:10:59	57.4 MB
apples-140007.annot	2023-02-27 13:49:03	202 KB
apples-140007.edf	2023-02-22 16:10:59	58.7 MB
apples-140008.annot	2023-02-27 13:49:03	217 KB
apples-140008.edf	2023-02-22 16:11:00	60.4 MB
apples-140009.annot	2023-02-27 13:49:03	187 KB
apples-140009.edf	2023-02-22 16:11:00	55.3 MB
apples-140011.annot	2023-02-27 13:49:03	226 KB
apples-140011.edf	2023-02-22 16:11:01	62.5 MB
apples-140012.annot	2023-02-27 13:49:03	194 KB
apples-140012.edf	2023-02-22 16:11:01	52.1 MB
apples-140013.annot	2023-02-27 13:49:03	183 KB

nsrr download apples/polysomnography

Don't have the NSRR gem?

Install using our [NSRR Gem Installation Instructions](#).

© 2023 National Sleep Research Resource Privacy Contact Us | Funded By NHLBI

Controlled access to human subjects data

- All data de-identified, but need to adhere to data sharing language in original participant informed consent
- Stream-lined process for DAUA, with proposals reviewed by NSRR staff
- Provide in-house IRB review for users without local IRB

The screenshot displays the NSRR Data Request portal. The top navigation bar includes links for NSRR, About, Datasets, Tools, Forum, Blog, Share data, and Webinars. The main content area is titled "DATA ACCESS AND USE AGREEMENT" and contains the following sections:

- Data Request:** Standard (Individual) v1.1.2.5625e02, with a link to "switch to organization".
- DAUA Progress:** A list of items with checkboxes: Page 1, Page 2, Page 3, and Signature, all of which are checked.
- Uploads:** A section for uploading documents.
- Datasets:** A list of datasets, currently showing "NOP-NHP".
- Proof:** A section for providing proof of IRB review, currently showing "Brigham and Women's Hospital".
- Intended Use of Data:** A section for describing the intended use of the data, with a text area and a note: "Please provide enough detail about your research and how you intend to use the data. All requests are reviewed for appropriateness based on the description provided. Incomplete descriptions will result in delays in granting access."
- Security:** A section for attesting to data security, with a checked checkbox: "I attest that data will be stored on a secure password protected device."
- HIPAA Training:** A section for attesting to HIPAA training, with a checked checkbox: "I attest that I have completed a Human Subjects Protections Training Course."

At the bottom, there is a link for more information about privacy and security trainings: <https://www.hhs.gov/hipaa/for-professionals/training/index.html>. A final note states: "For avoidance of doubt, permissible uses may include use of the Data/Datasets for research evaluation and testing of a product or technology but will not extend to proposals that include or incorporate the Data/Datasets into such product. BWH will provide the Data/Datasets requested by the Data User upon".



NATIONAL SLEEP RESEARCH RESOURCE

ADVANCING SCIENCE GLOBALLY
THROUGH DATA SHARING

- Ongoing **NHLBI contract** to sustain NSRR and facilitate integration with the NHLBI BioData Catalyst (BDC) ecosystem
 - to link sleep data with large amounts of genetics, omics, imaging, bioassays, cognitive, medical and other data available on many of the NSRR cohorts
 - to transition away from our locally-hosted data-download model
 - to foster modern cloud-based, reproducible workflows and analysis capabilities
- **STRIDES** supplement: pending BDC-readiness for data ingestion, to support interim transfer of NSRR data to the cloud, enabling cloud-savvy users direct access

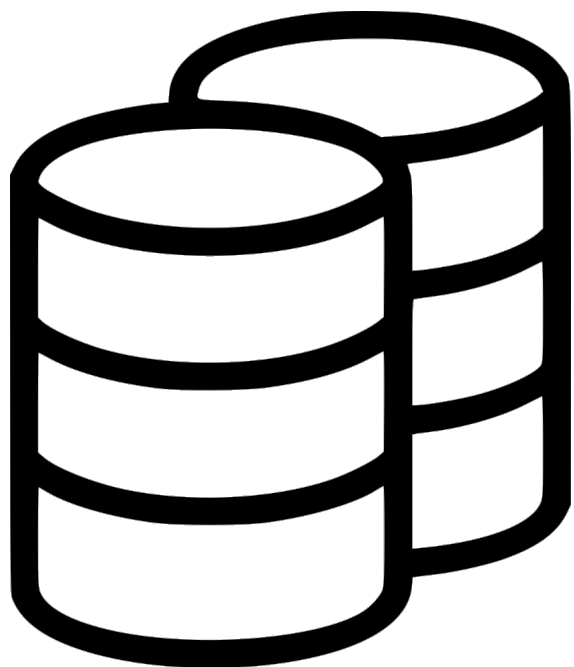


NATIONAL SLEEP RESEARCH RESOURCE

ADVANCING SCIENCE GLOBALLY
THROUGH DATA SHARING

- Configured a secure AWS-STRIDES environment
- Custom-built authentication/authorization system for controlled access
- Python client library for users to pull data from S3 buckets to EC2 instances
- Migrated all Cleveland Family Study data

- Bottlenecks encountered:
 - **regulatory approvals** required to transfer key cohorts from on-premise servers to NIH-administered STRIDES AWS environment
 - non-trivial work with NIH, depositors and parent cohorts to amend original data hosting agreements
 - unable to use STRIDES credits on AWS accounts administered by our local institution
 - **lessons learned:** moving forward broader, host-agnostic data use agreement templates for new studies



DATA

Data need tools

What can we learn from these data?

“Reference-driven analysis”

Using old data to inform the analysis of new data

Making tools accessible

Sharing tools as well as data



TOOLS



Many published results cannot be reproduced

- open tools and data important for reproducibility
- but often not transparently reported or shared

Literature review of all sleep methods/algorithms papers published in 2019:

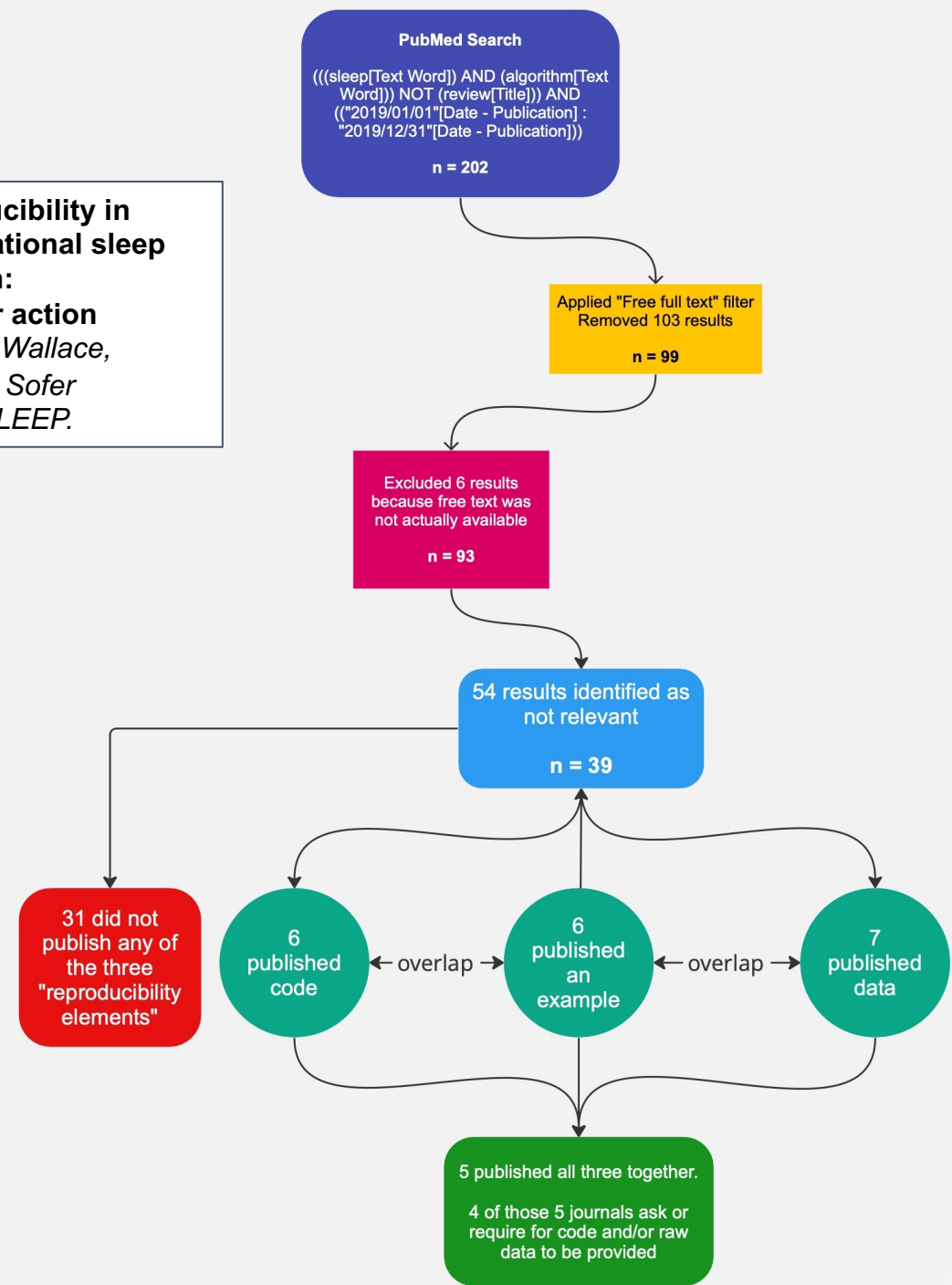
80% of relevant papers did not publish either:

1) code, 2) an example or 3) any data

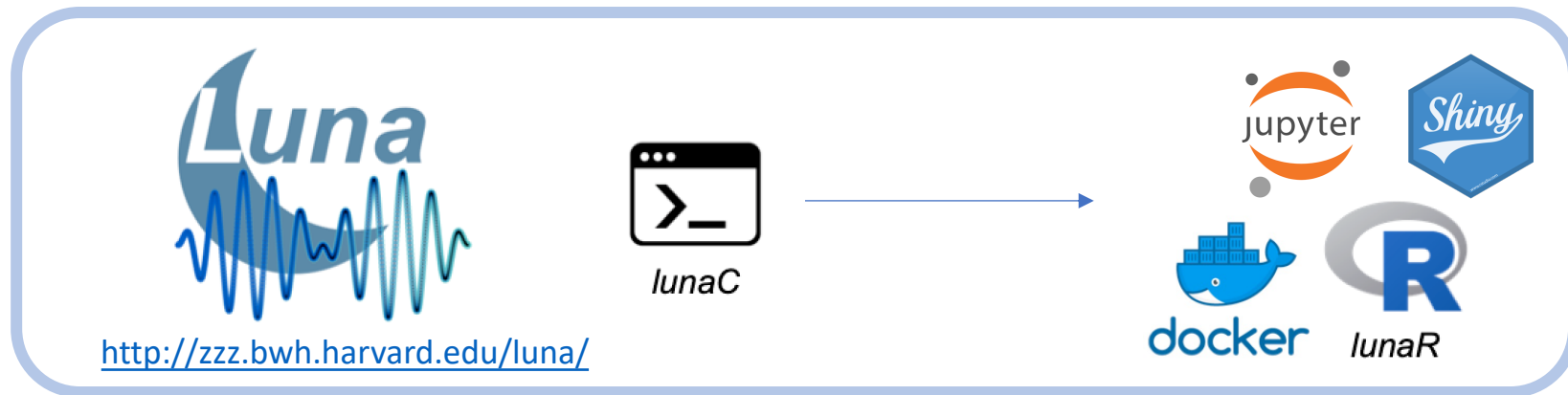
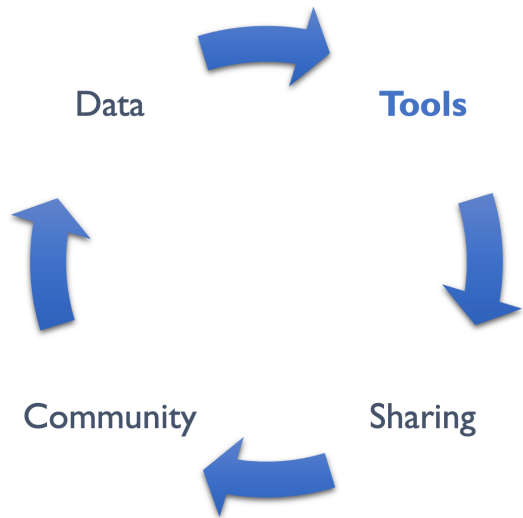
Of 5 articles that supplied all of these, 4 were published in journals that explicitly ask for code and/or raw data

Reproducibility in computational sleep research:

a call for action
Cassidy, Wallace, Purcell & Sofer (2024) SLEEP.



1) Making in-house tools more accessible

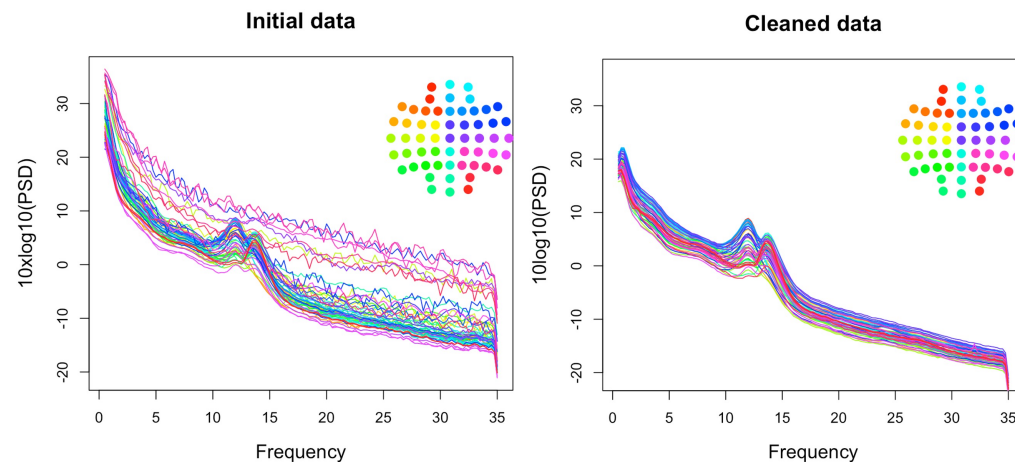


2) Sharing pipelines as well as data

NAP: NSRR Automated Pipeline

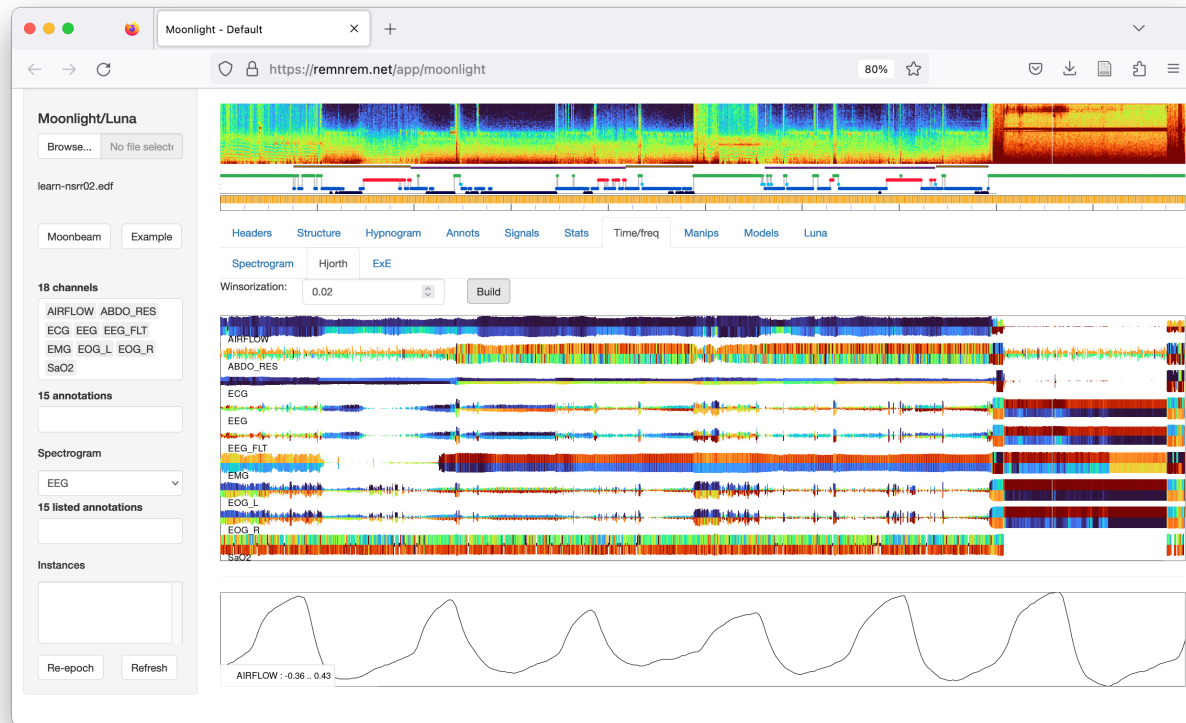
Harmonize/flag issues w/ labels, referencing, units, sample rates, filtering, polarities, corrupt signals, artifact, non-standard channels, (automated) staging alignment, etc

Uniform annotation format

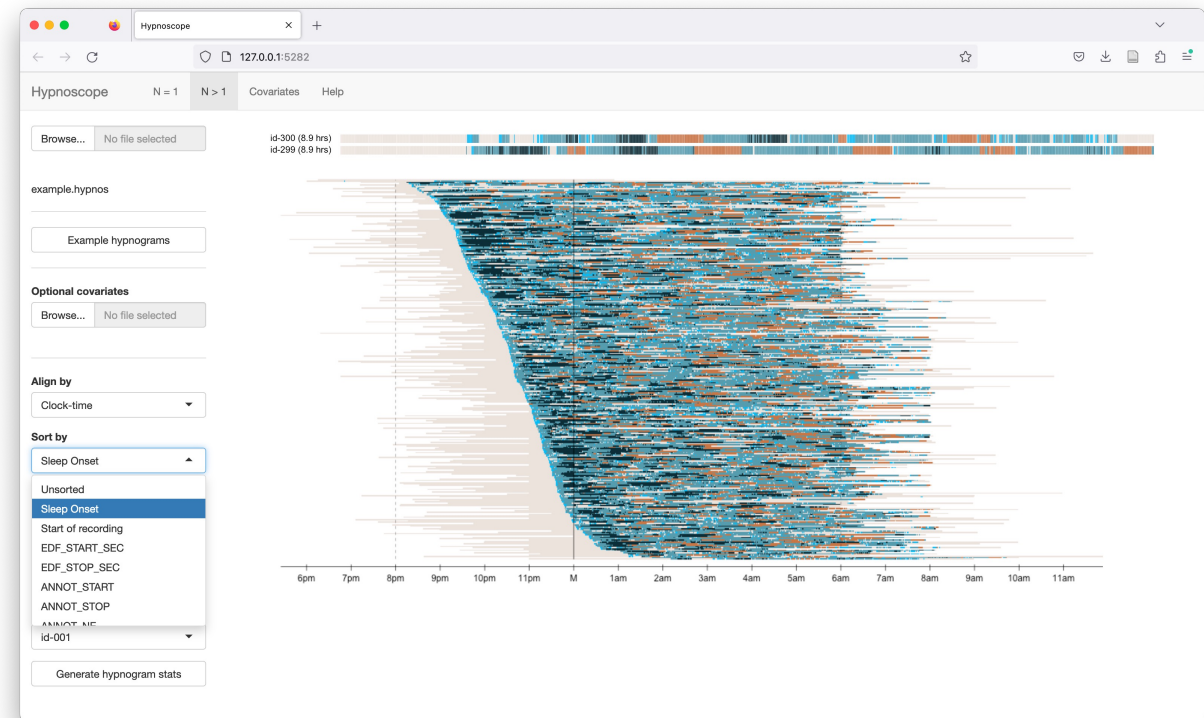


Cloud-based apps for sleep signal data

Moonlight



Hypnoscope



Alternative ways to access NSRR data

Moonbeam

The screenshot shows the Moonbeam web interface. At the top, there are tabs for 'NSRR' and 'User URLs'. Below that, there is a field for 'NSRR token (http://sleepdata.org/token)' with a masked input and an 'Authenticate' button. A dropdown menu is open, showing '10 cohorts' with 'Apnea Positive Pressure Study' selected. Other cohorts listed include Cleveland Children Sleep Health Study, Cleveland Family Study, Childhood Adenotonsillectomy Trial, Hispanic Community Health Study, Multi-Ethnic Study of Atherosclerosis, and Mignot Nature Communications. To the right, there is a field for '1104 records' and an 'Import' button. A 'Cancel' button is at the bottom right.

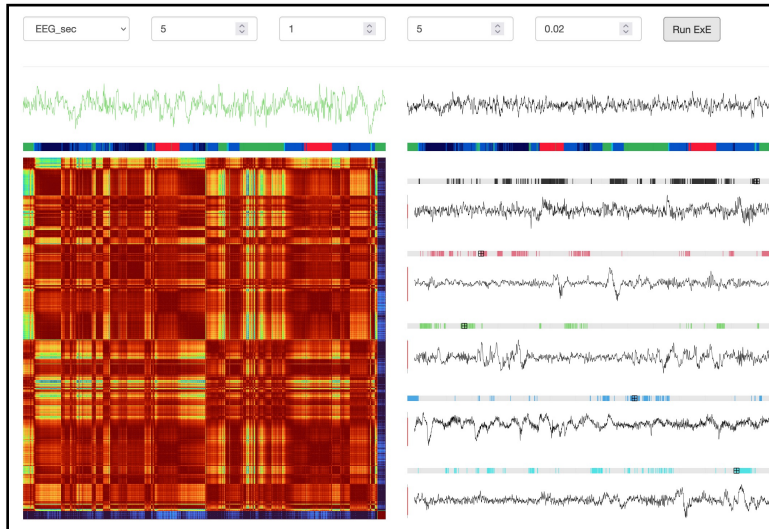
Clinical, demographic & other phenotypic data

The screenshot shows a data table in the Moonbeam interface. The table has columns for 'Variable', 'Value', 'Units', and 'Description'. The data is for cohort 'shhs1-200226'. The table shows variables like 'nsrr_age', 'nsrr_sex', 'nsrr_bmi', and 'nsrr_flag_spsw'. The 'Description' column provides detailed information about each variable, including harmonization standards and source information.

Variable	Value	Units	Description
nsrr_age	47	years	Subject age - Harmonized by the NSRR team to align with TOPMed and BioDataCatalyst standards . Source: :age_s1: at Visit 1 and :age_s2: at Visit 2
nsrr_sex	female	.	Subject sex - Harmonized by the NSRR team to align with TOPMed and BioDataCatalyst standards . Source: :gender:
nsrr_bmi	41.3123583	kilograms per square meter	Body mass index (BMI) - Harmonized by the NSRR team to align with TOPMed and BioDataCatalyst standards . Source: :bmi_s1: at Visit 1, and :bmi_s2: at Visit 2
nsrr_flag_spsw	full scoring	.	Study scored sleep/wake rather than with full sleep staging if there is not sufficient quality on EEG lead. Studies scored as sleep wake have all sleep stages scored using a default of Stage 2 and no arousals are scored. - Harmonized by the NSRR team. Source: :staging5:

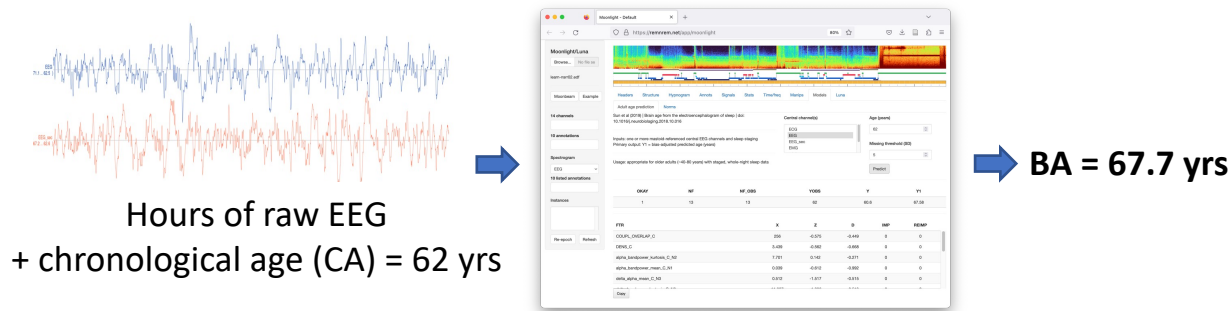
Showing 1 to 986 of 986 entries

Signal data & tools



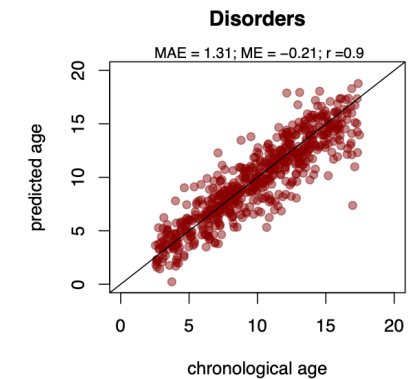
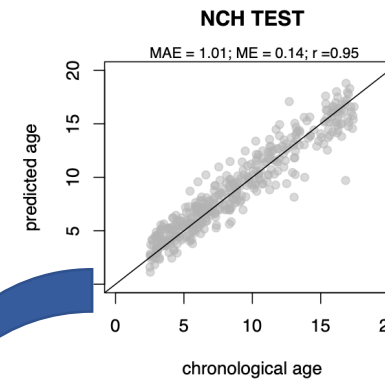
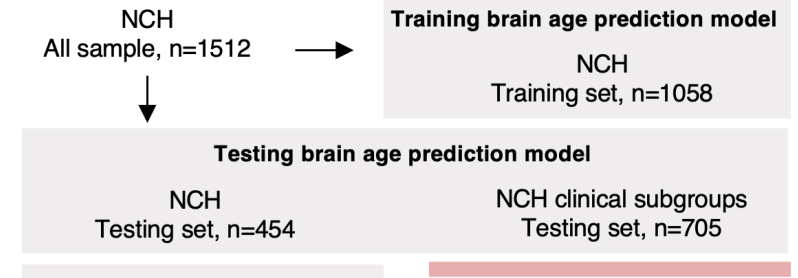
Sharing models as well as data & tools

- Automated sleep staging
- Biological age (BA) prediction
- Disease risk/status
- etc.



"Soup-to-nuts" cloud-based analyses

i.e. staging, QC, feature extraction, missing data imputation & prediction (but with intermediate steps still visible/modifiable)



Kozhemiako et al (2023) Mapping Typical and Altered Neurodevelopment with Sleep Macro- and Micro-Architecture. *NeuroImage: Clinical*

$$r(CA, BA) = 0.95$$

$\Delta = BA - CA$ tracks with diagnostic status for several neurodevelopmental delays

Summary

- Primary goal: to make NSRR data and tools more accessible
- For controlled-access data, cloud environments can bring additional layers of regulatory and administrative challenges, especially when working with legacy datasets and a still-evolving broader ecosystem
- NSRR ingestion into BDC now on a smoother path
 - uploading new cohorts
 - sleep-specific CDEs
 - integration of NSRR tools, e.g. via Seven Bridges workflows

Acknowledgements

Brigham and Women's Hospital, Boston

Susan Redline, Shaun Purcell, Paige Sparks
Daniel Mobley, Michael Rueschman, Ying Zhang
Meg Tully, Susan Surovec, Matthew Kim
Michael Prerau, Senthil Palanivelu,
Rebecca Robbins, Stephany Zhivotovsky
Samar Jawde, Remo Mueller, Dennis Dean,
Sara Mariani

University of Kentucky

GQ Zhang, Satya Sahoo, Licong Cui

Beth Israel Deaconess Medical Center

Ary Goldberger, Madalena Costa

NHLBI

Marishka Brown, Weiniu Gan

All NSRR data depositors and users

<http://sleepdata.org>

