

east africa open science training

2018 nov

working open success stories *(and non-success stories)*

brian m. bot | principal scientist | sage bionetworks

sage bionetworks



aim to redefine how complex biological data
is gathered, shared and used

decentralization *

the process of redistributing or dispersing functions, powers, people, or things away from a central location or authority

* wikipedia

technology enabled biomedical research

recruitment

data generation

data analysis

decentralization

technology is only one piece

principles



people



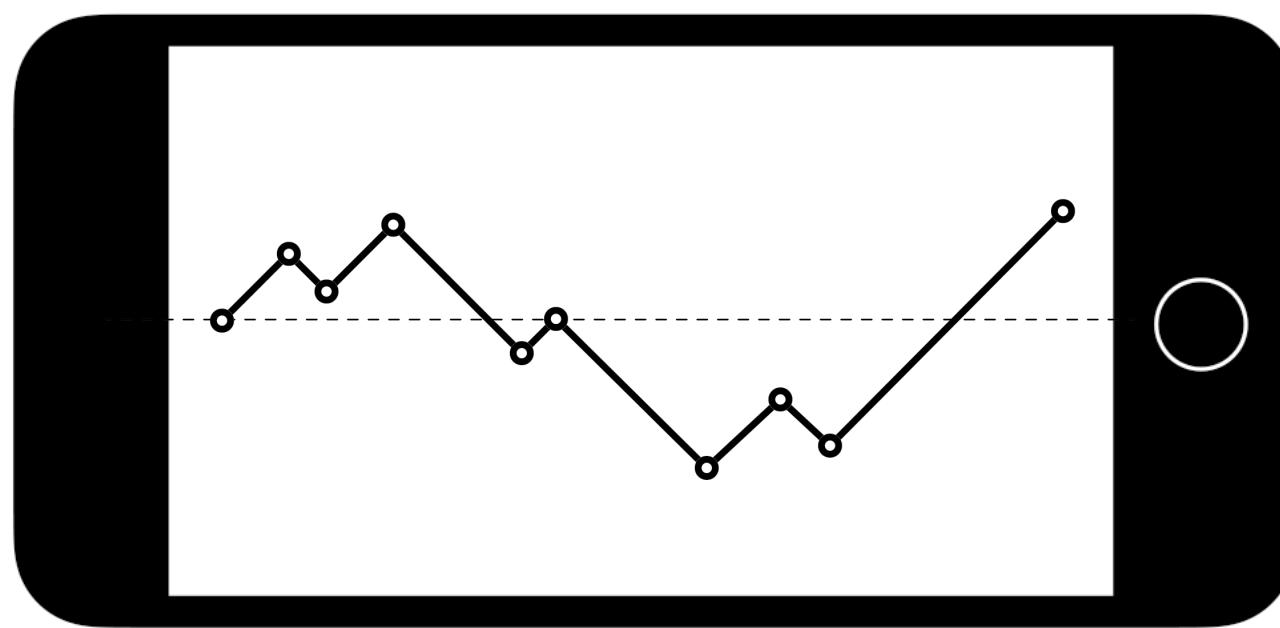
platforms



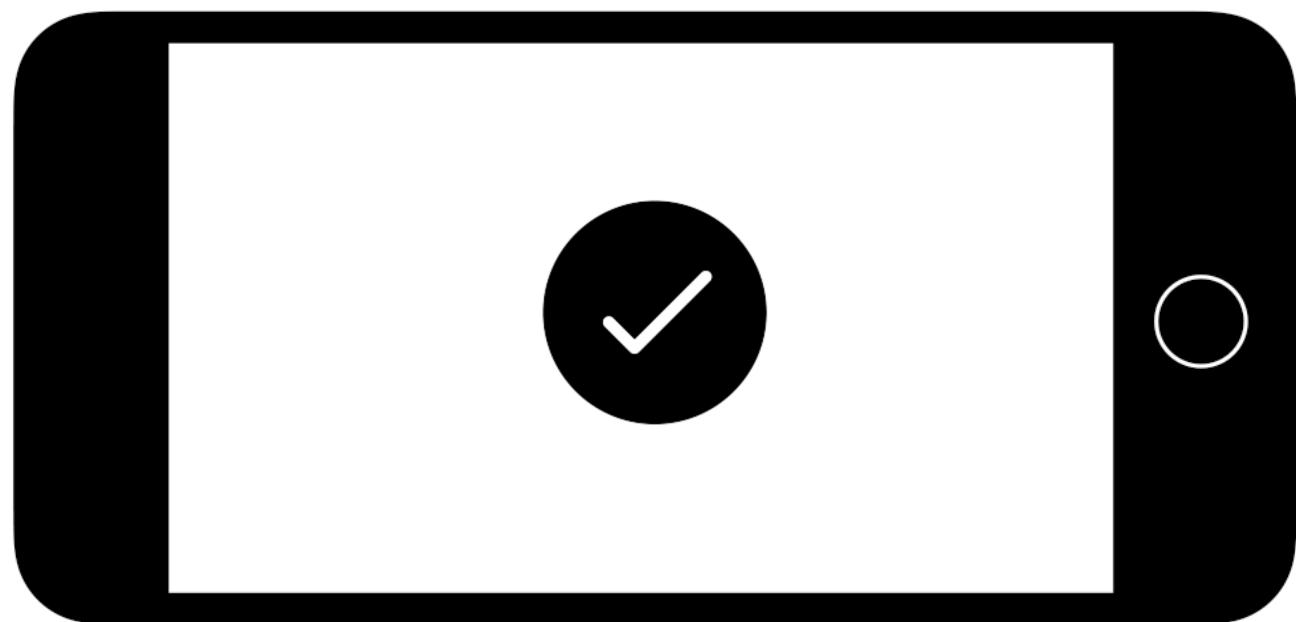
decentralization

use case 1: remote assessments

mobile Health



mHealth



proliferation of digital (health) technologies



potential of digital health technologies

- increased fidelity of health measurements
- reduce inefficiencies / costs
- improve access
- make medicine more personalized
- allow individuals to better track and manage their own health

emerging challenges of digital health technologies

- participant engagement and retention
- management of increasingly sensitive and identifiable data
- accuracy of remote measurements
- an ever evolving landscape
- *(realizing the potential discussed on the last slide)*

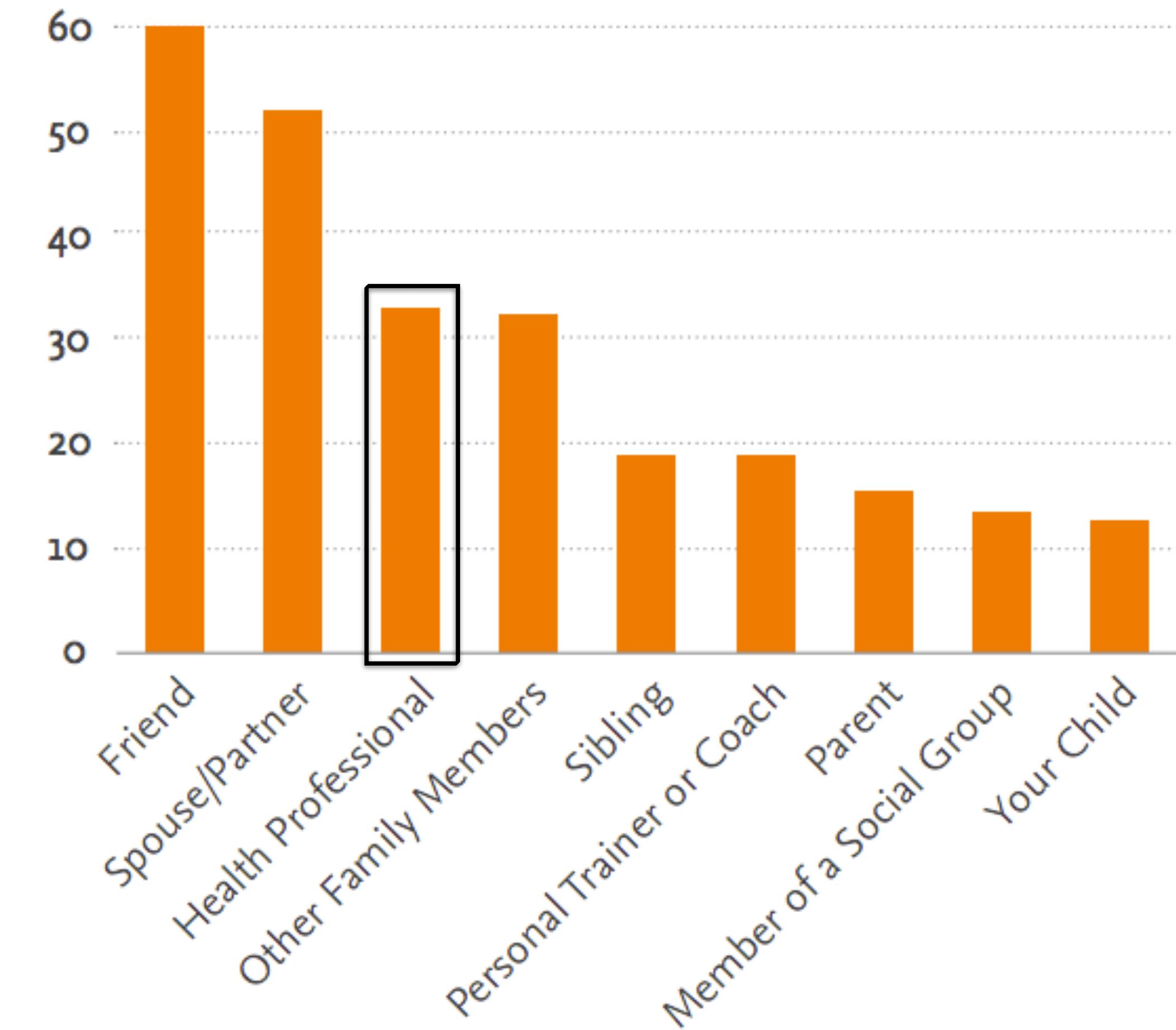


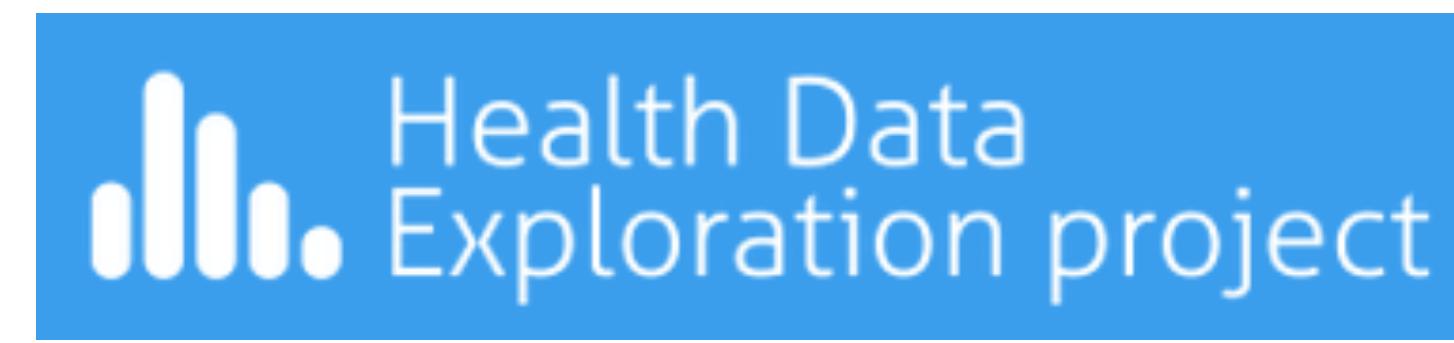
Robert Wood Johnson Foundation

Figure 4. Who do PHD users share with?

Who do you share your PHD with?

Percent of responses



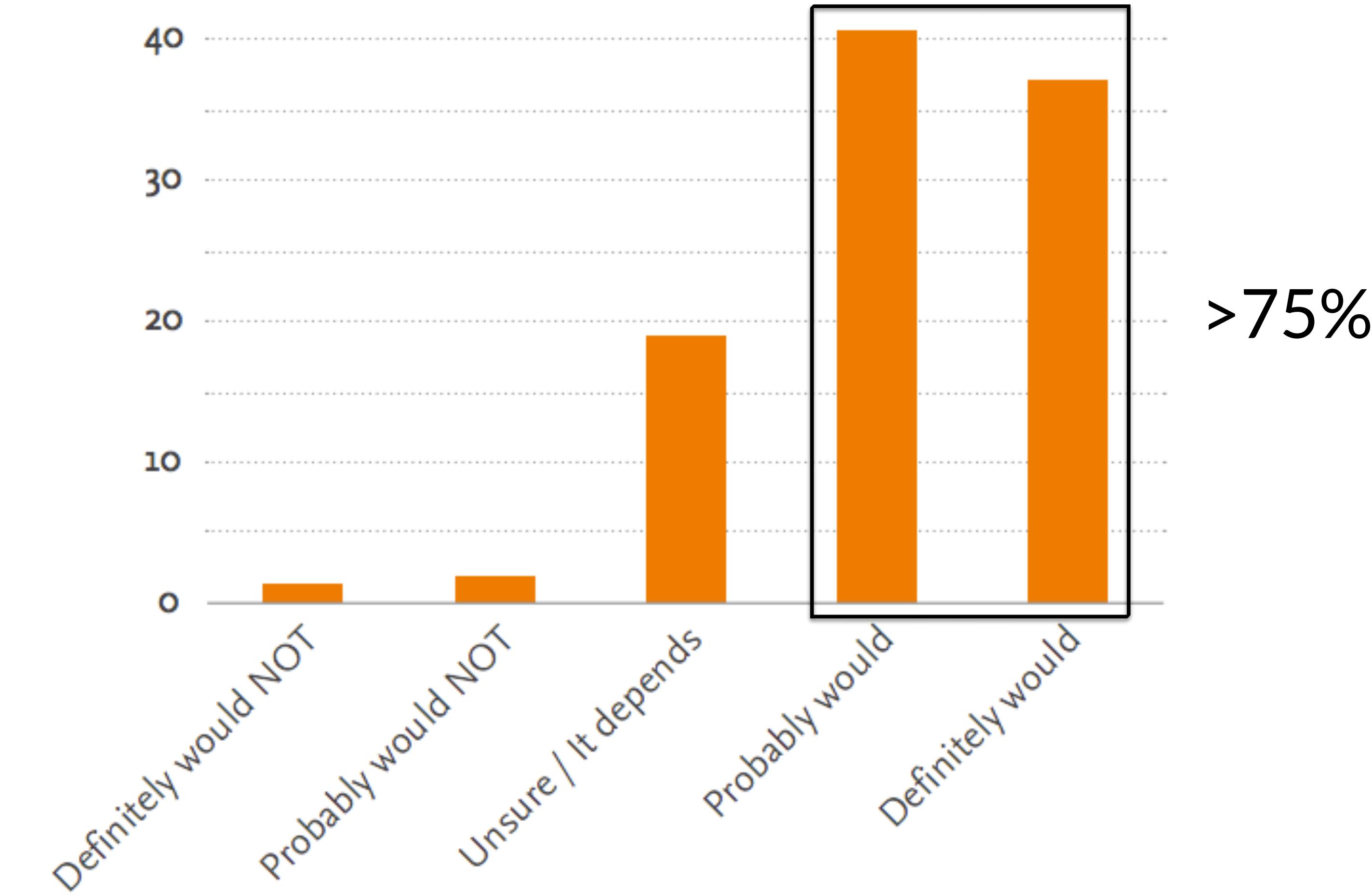


Robert Wood Johnson Foundation

Figure 5. **Sharing with Researchers**

Would you be willing to share your PHD with researchers?

Percent of responses



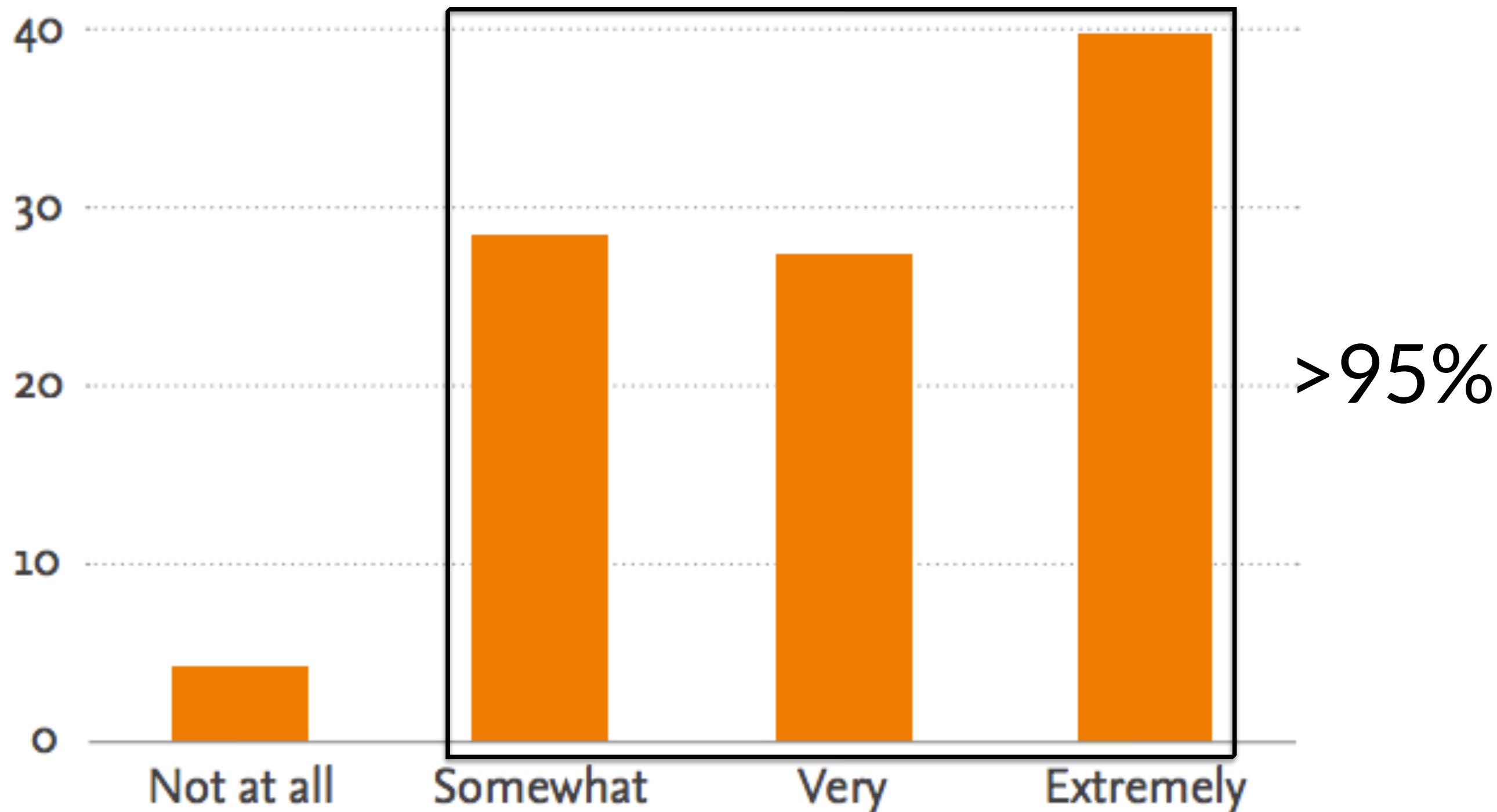


Robert Wood Johnson Foundation

Figure 6. Attitudes towards anonymity of Personal Health Data

How important is it to you that your PHD be kept anonymous?

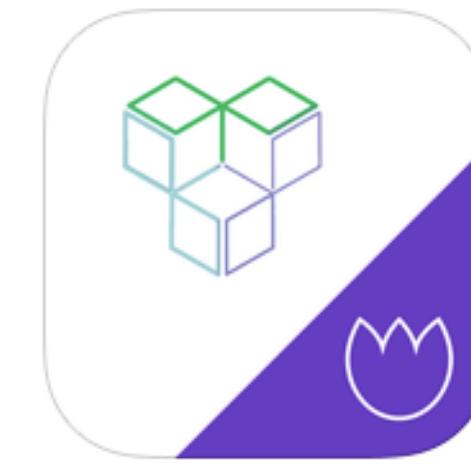
Percent of responses





A software framework
made specifically for
medical research

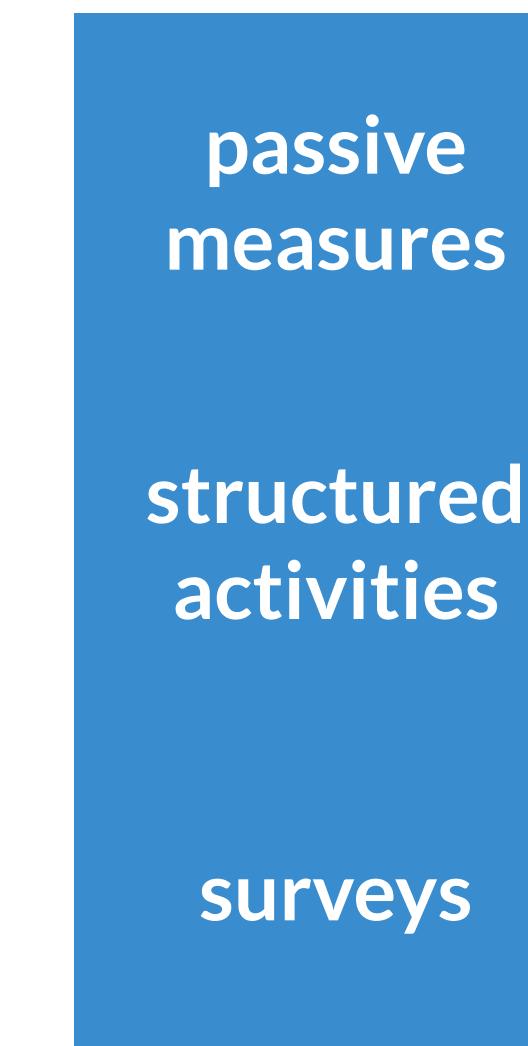




mPower

mPower helps decipher Parkinson's disease.

The variability in Parkinson's disease symptoms has left many questions unanswered. So the University of Rochester and Sage Bionetworks created the mPower app to precisely measure data such as dexterity, balance, memory, and gait. This information could help researchers better understand how various symptoms are connected to Parkinson's disease. In turn, participants could start to recognize their own signs and symptoms.



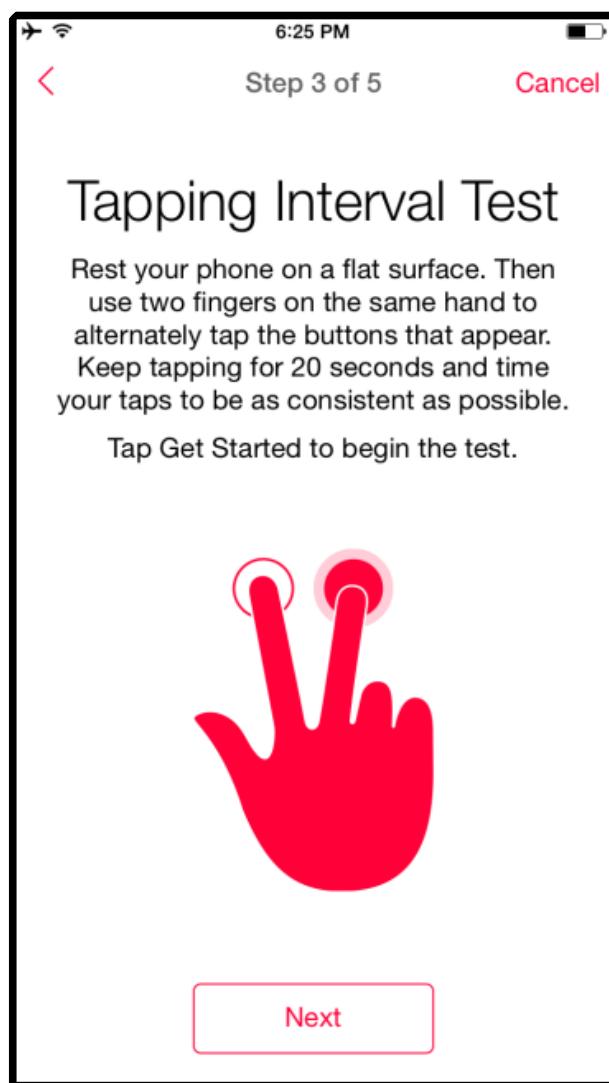
motor initiation	gait/balance	hypophonia	memory
gps - displacement vectors	gps - displacement vectors	-	-
tapping activity	walking/standing activity	voice activity	memory game
MDS-UPDRS PDQ8	MDS-UPDRS PDQ8	MDS-UPDRS PDQ8	MDS-UPDRS PDQ8



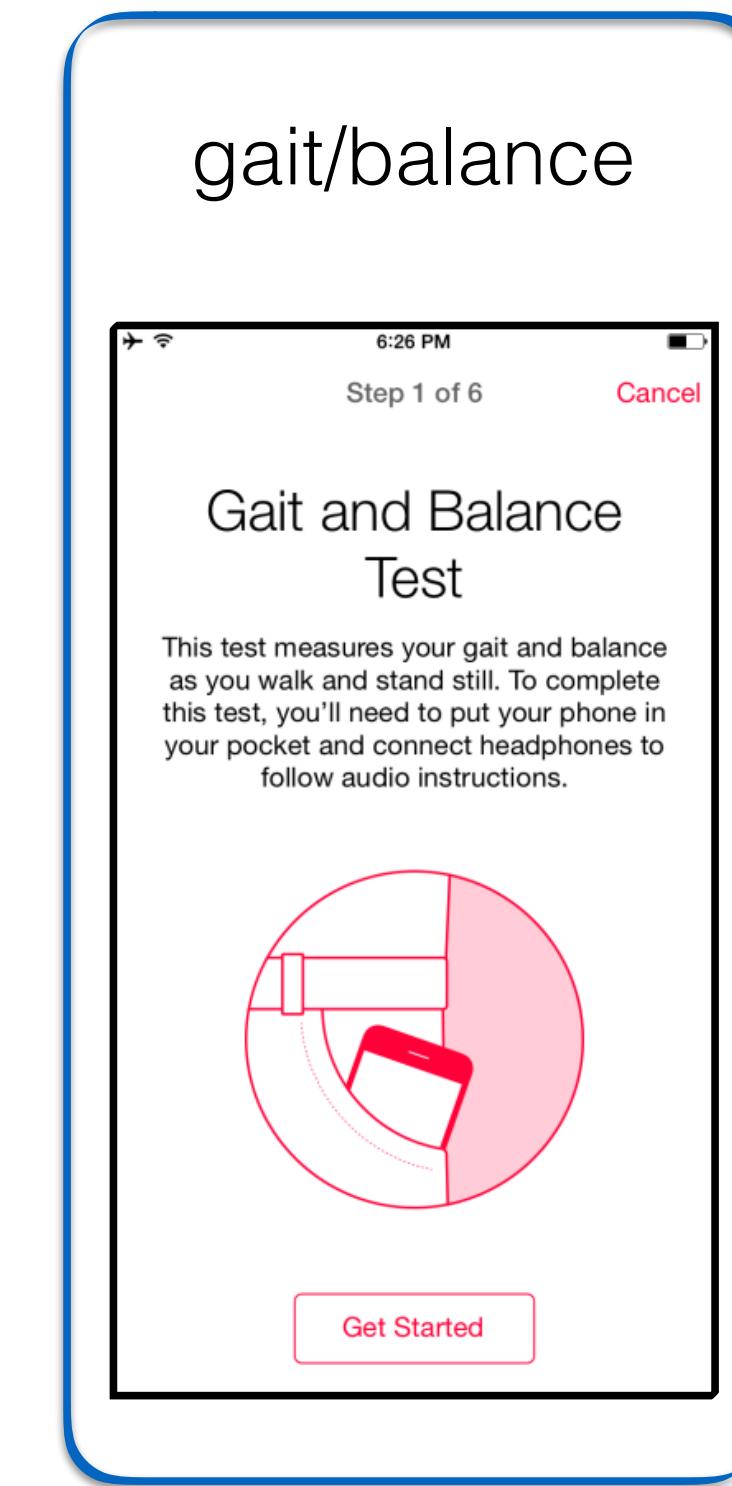


mPower activities

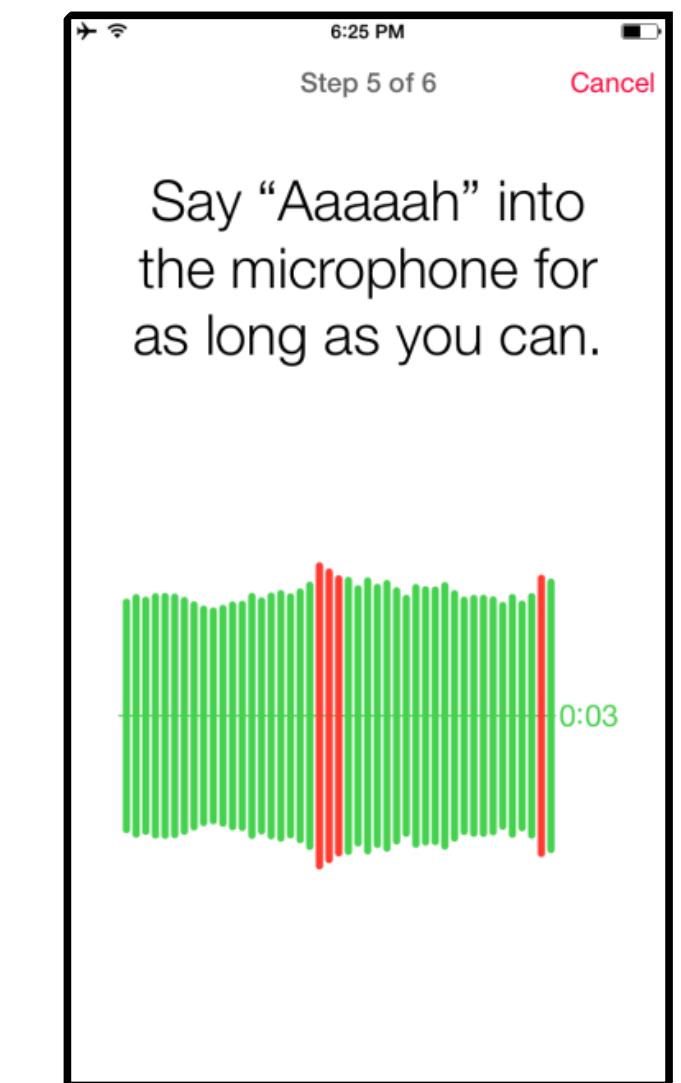
motor initiation



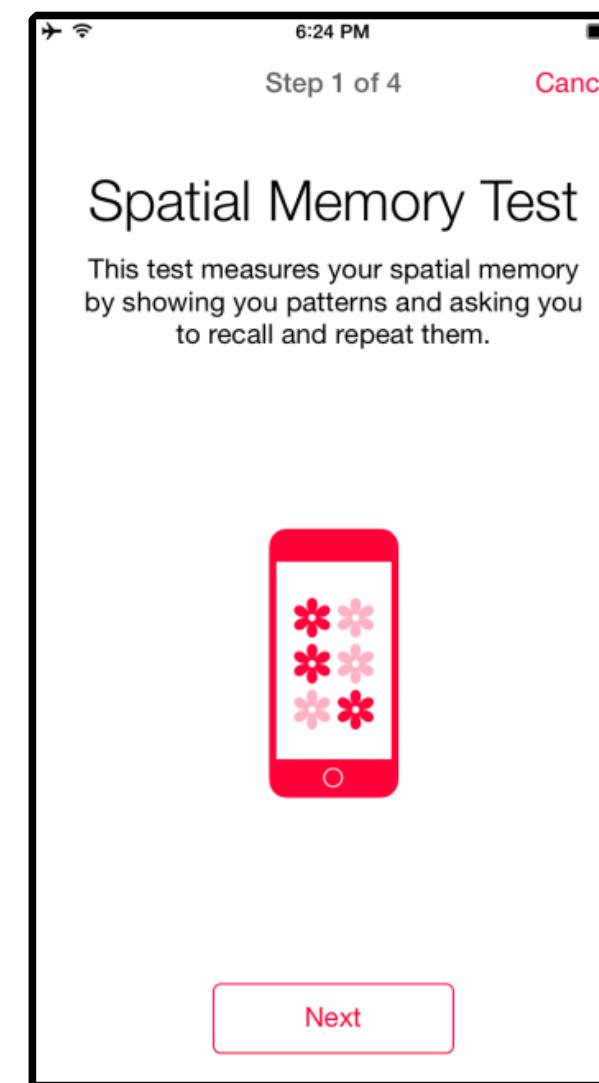
gait/balance



hypophonia

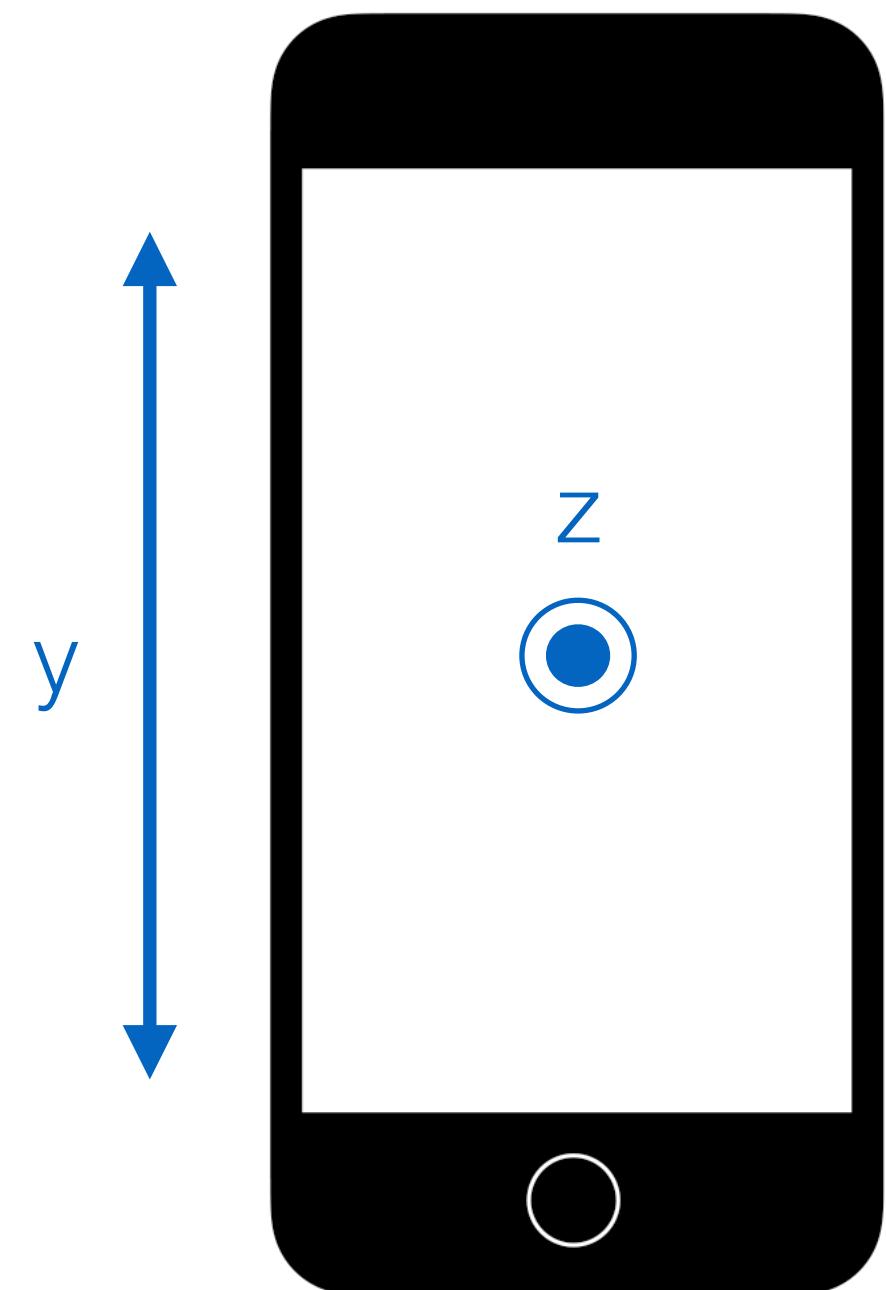


memory





gait / balance



userAcceleration

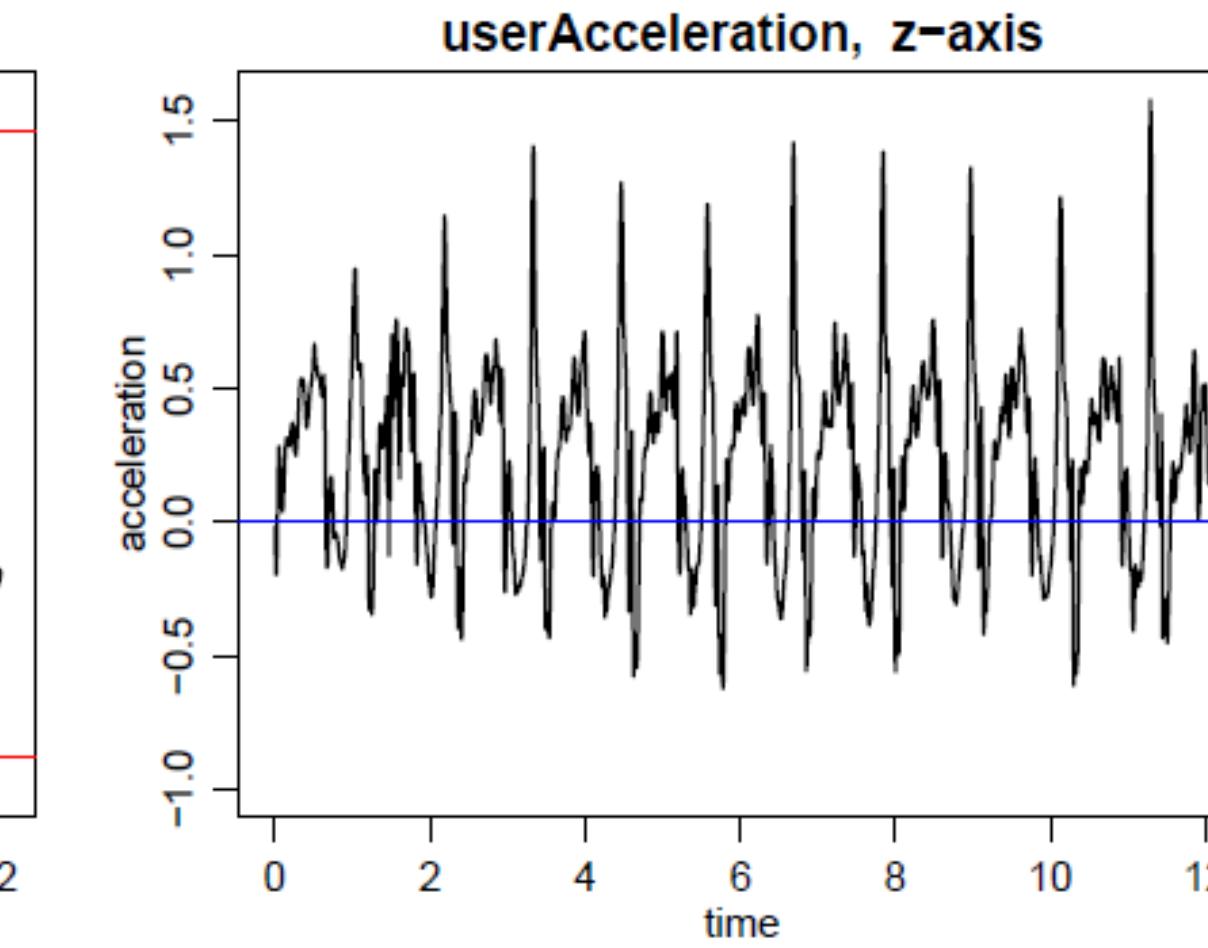
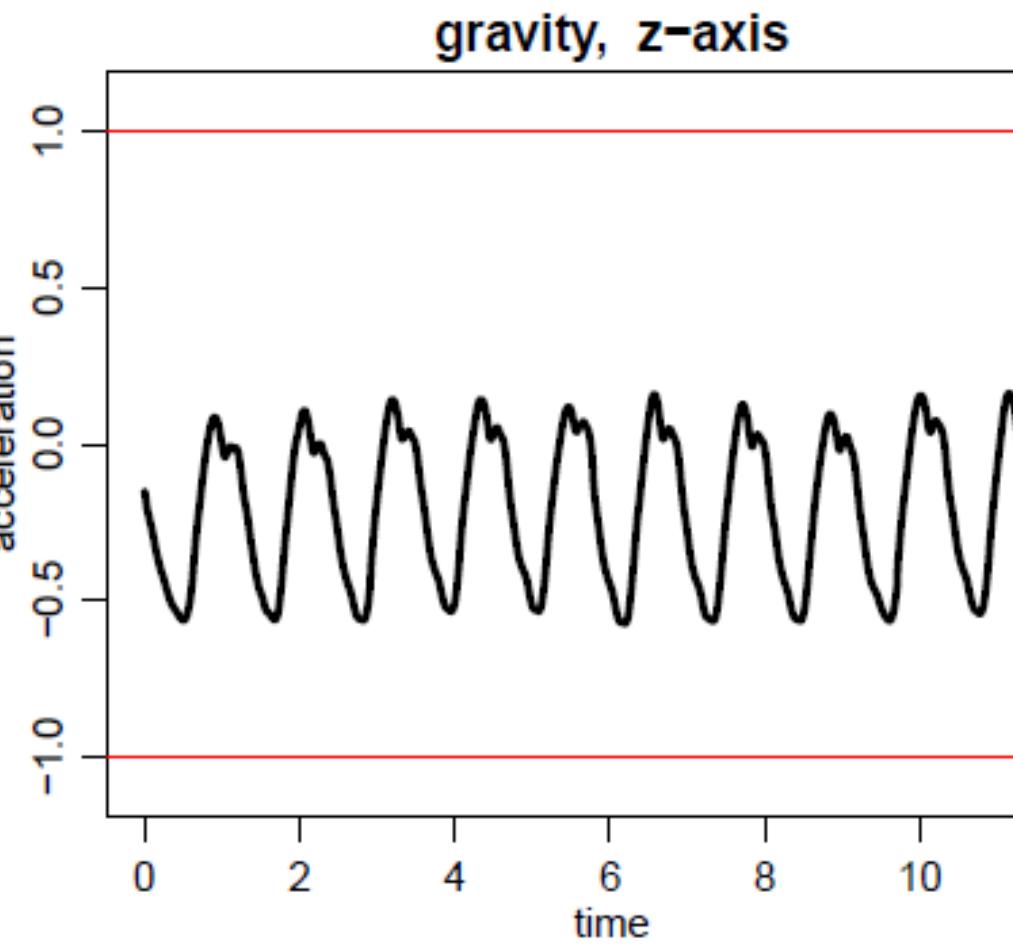
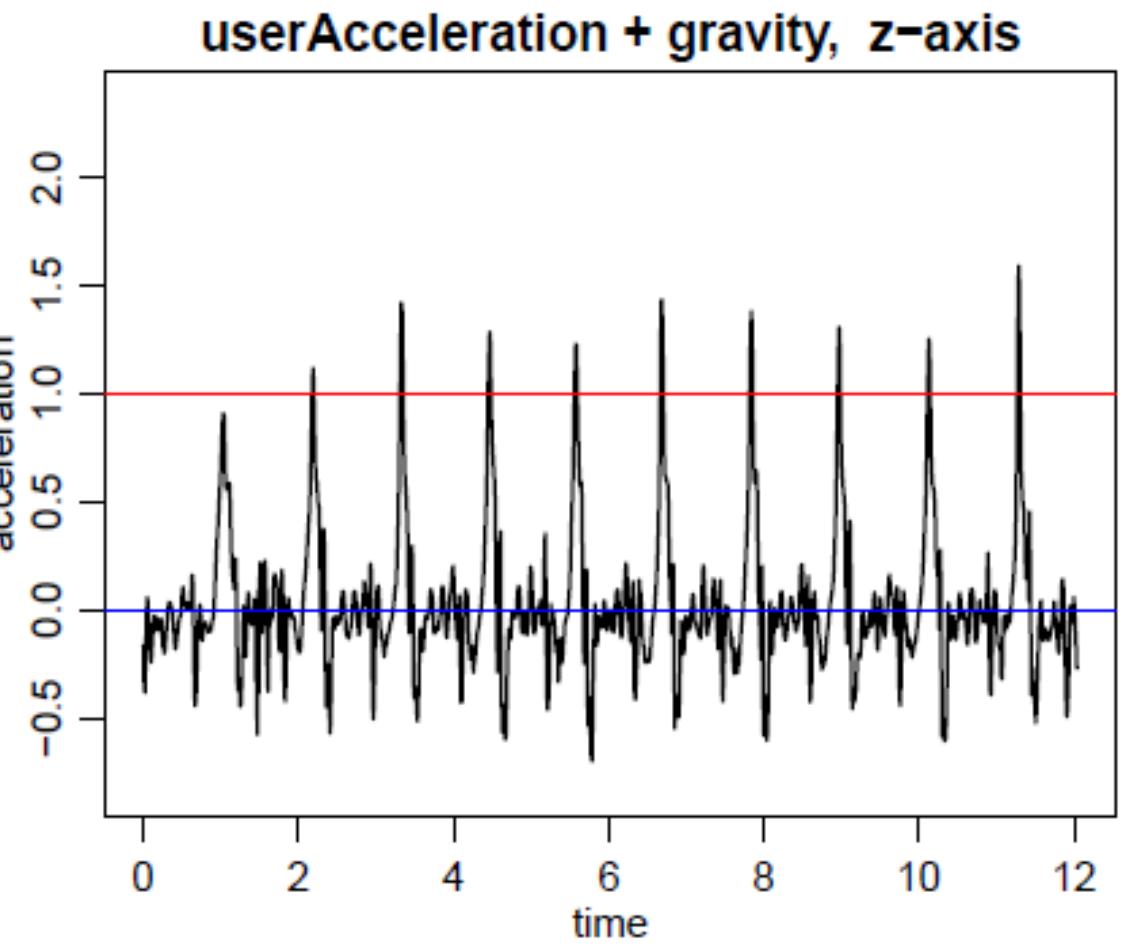
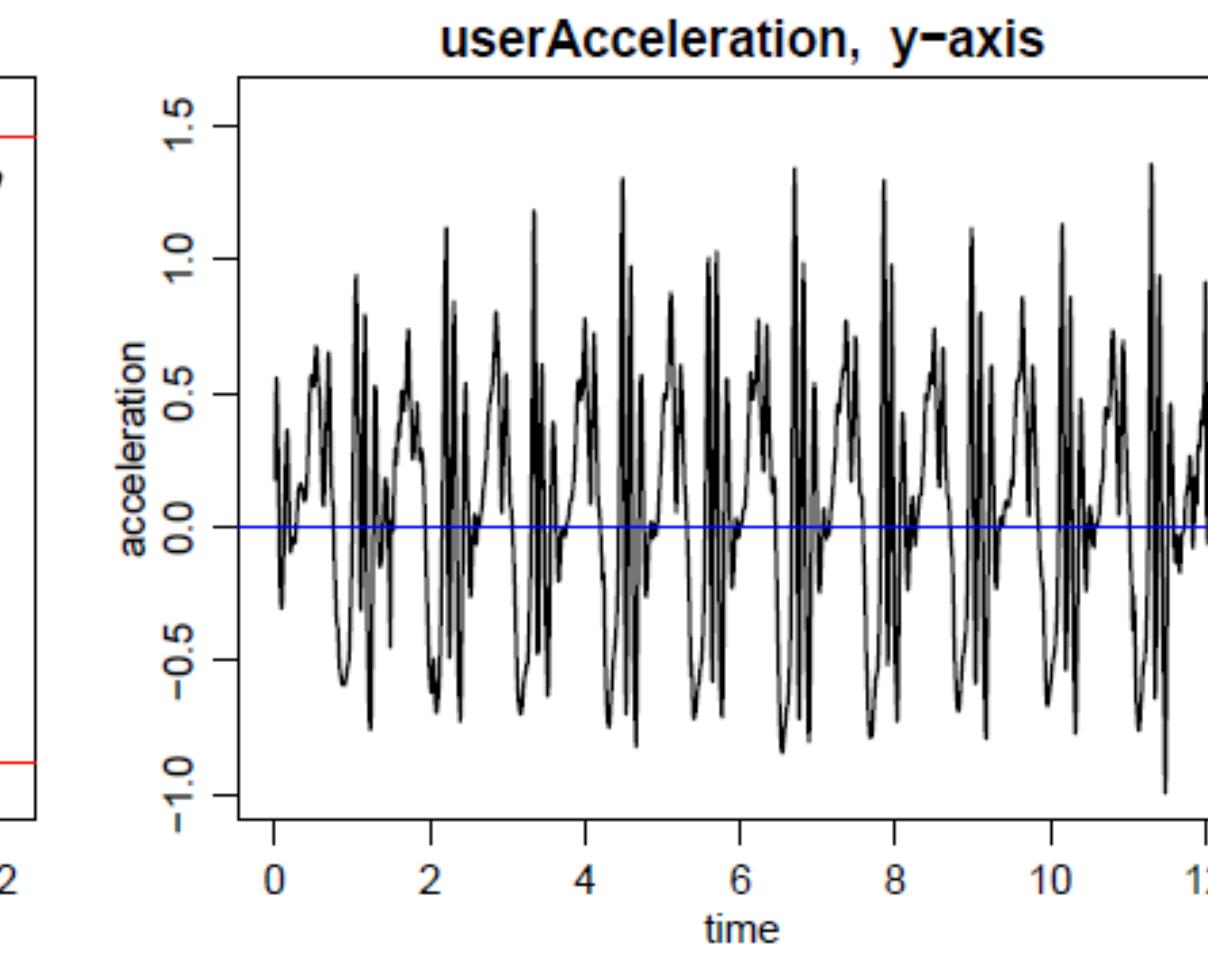
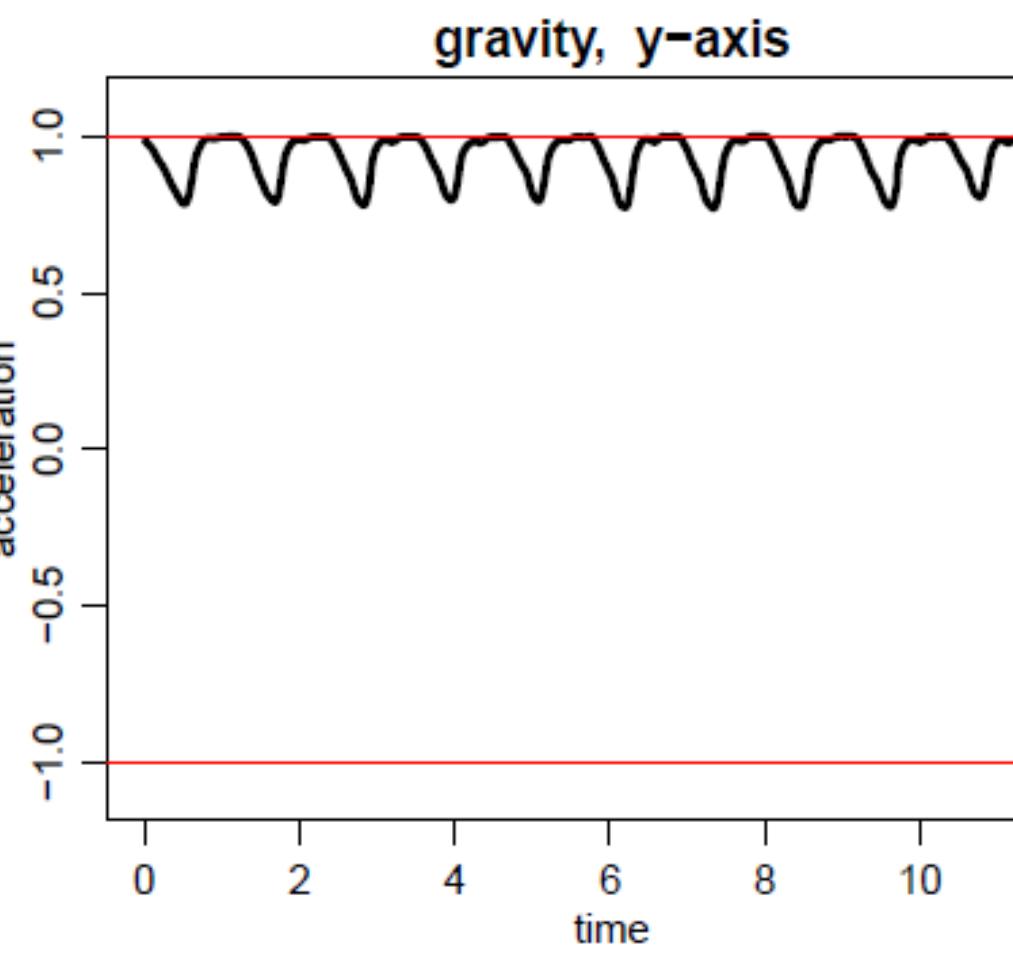
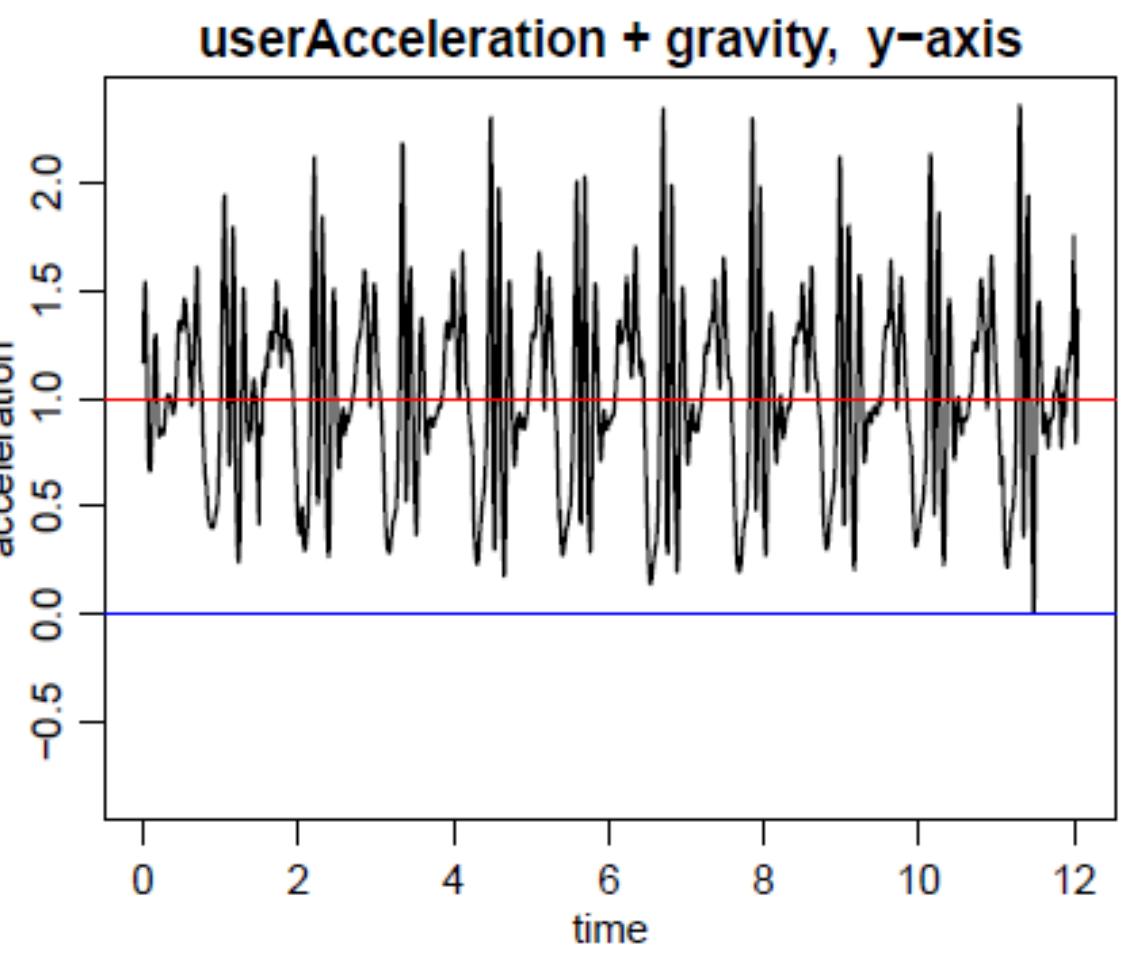
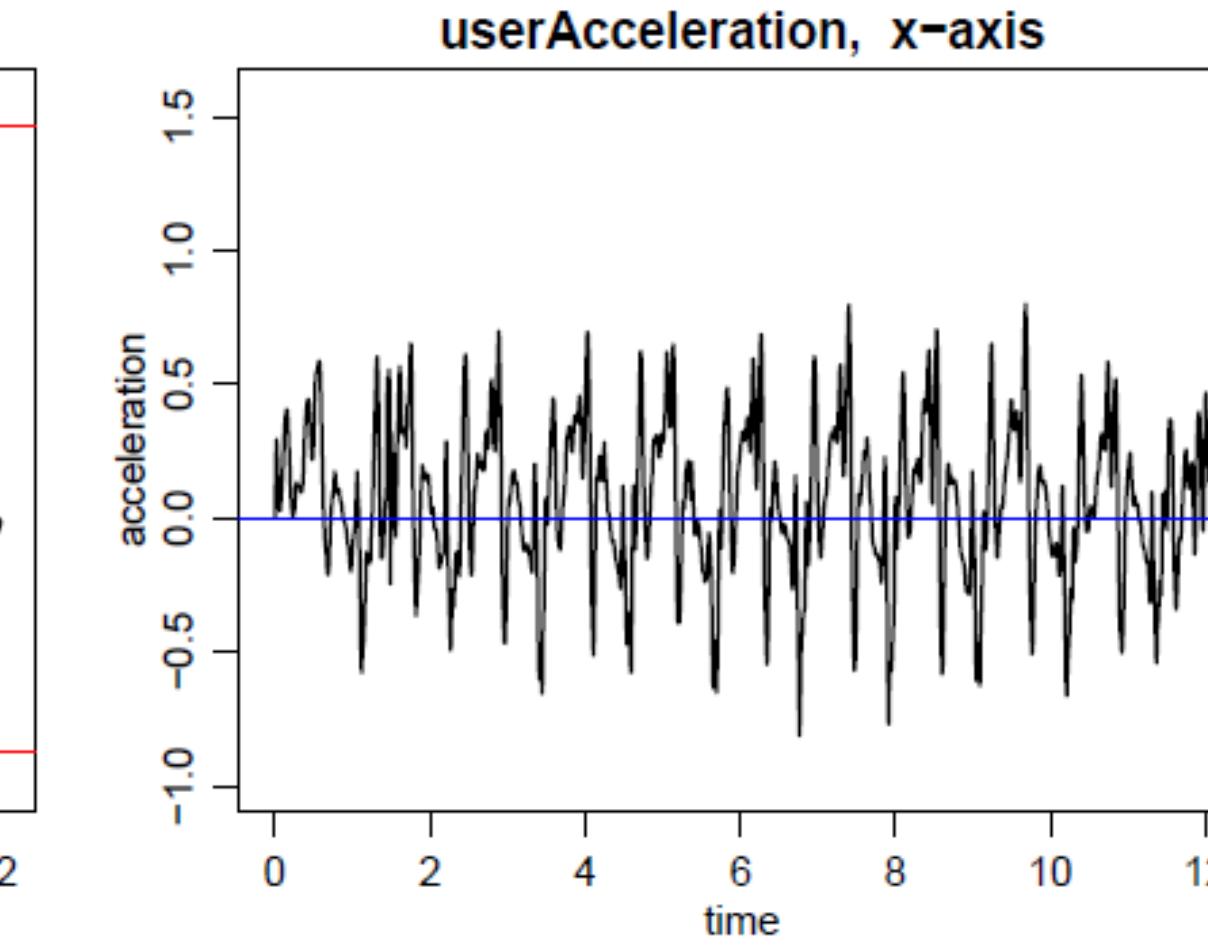
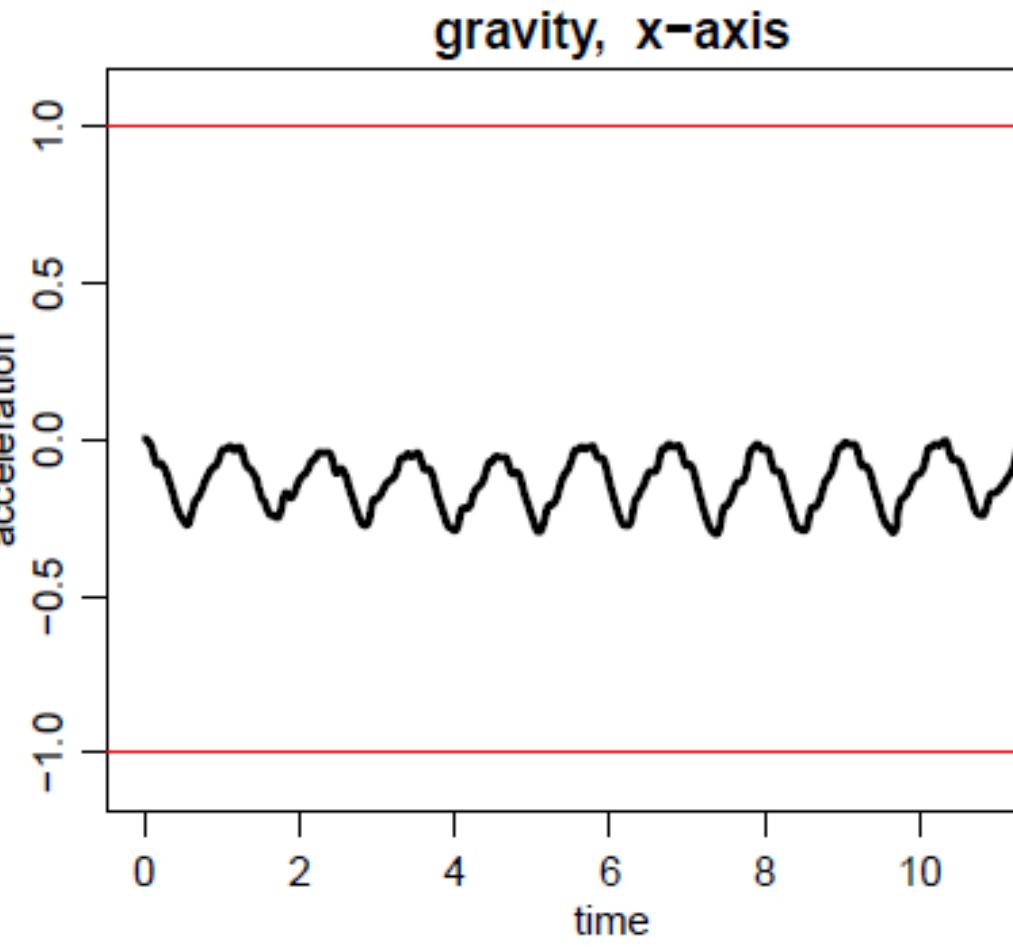
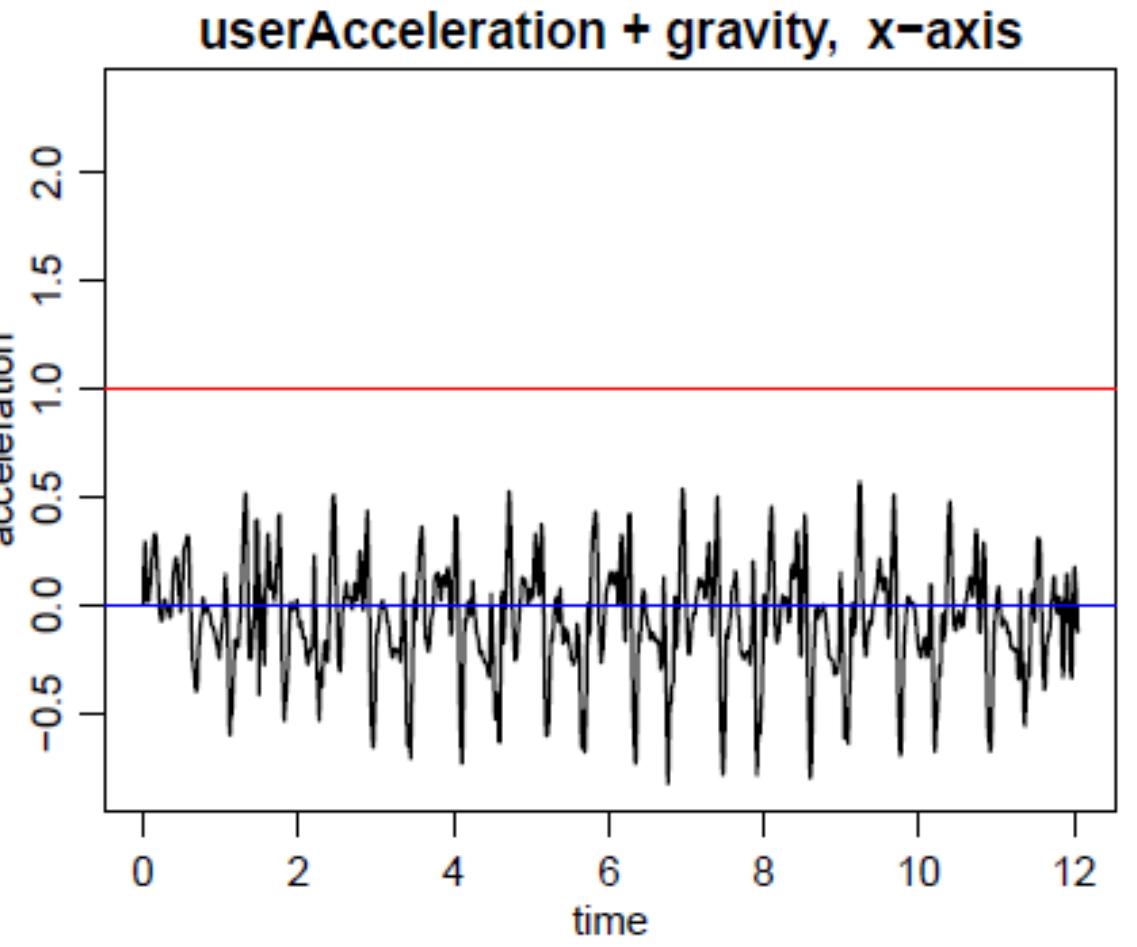
gravity

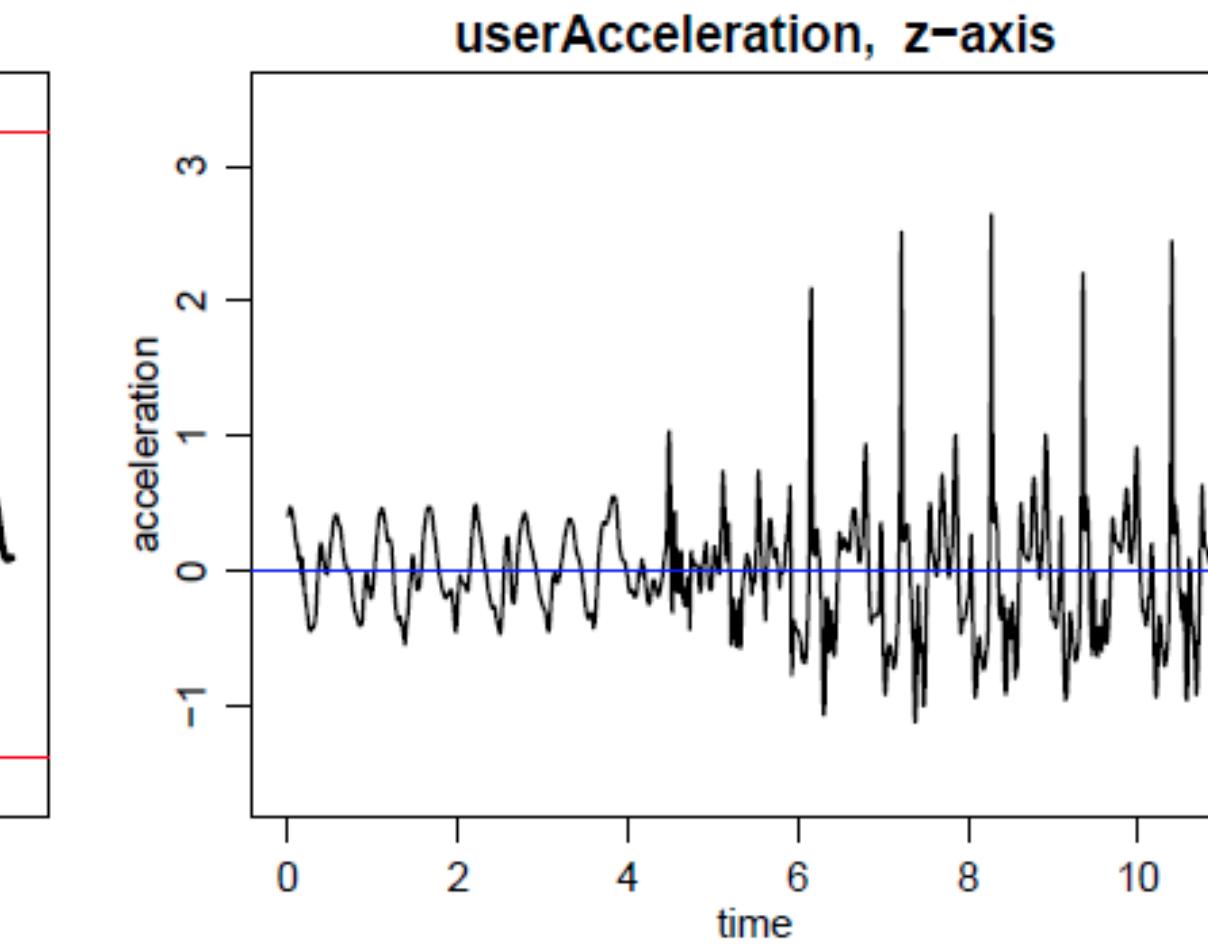
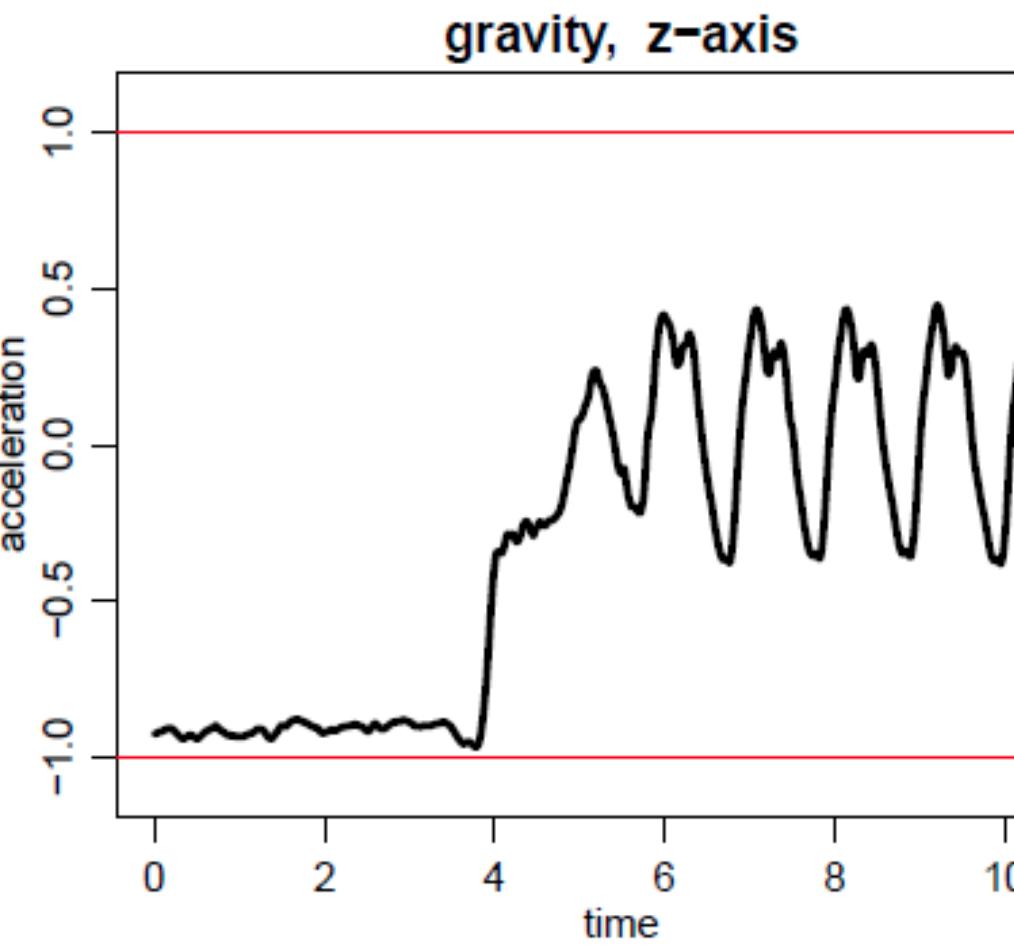
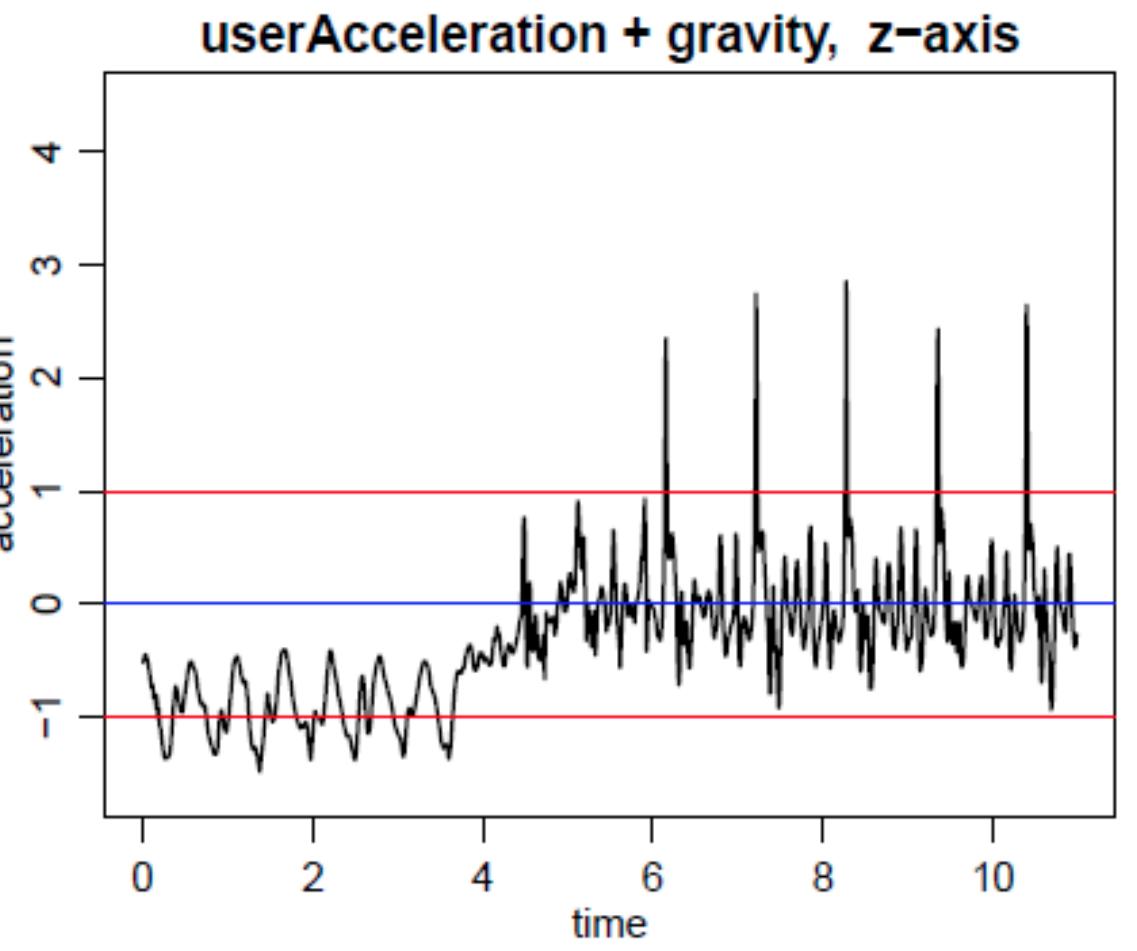
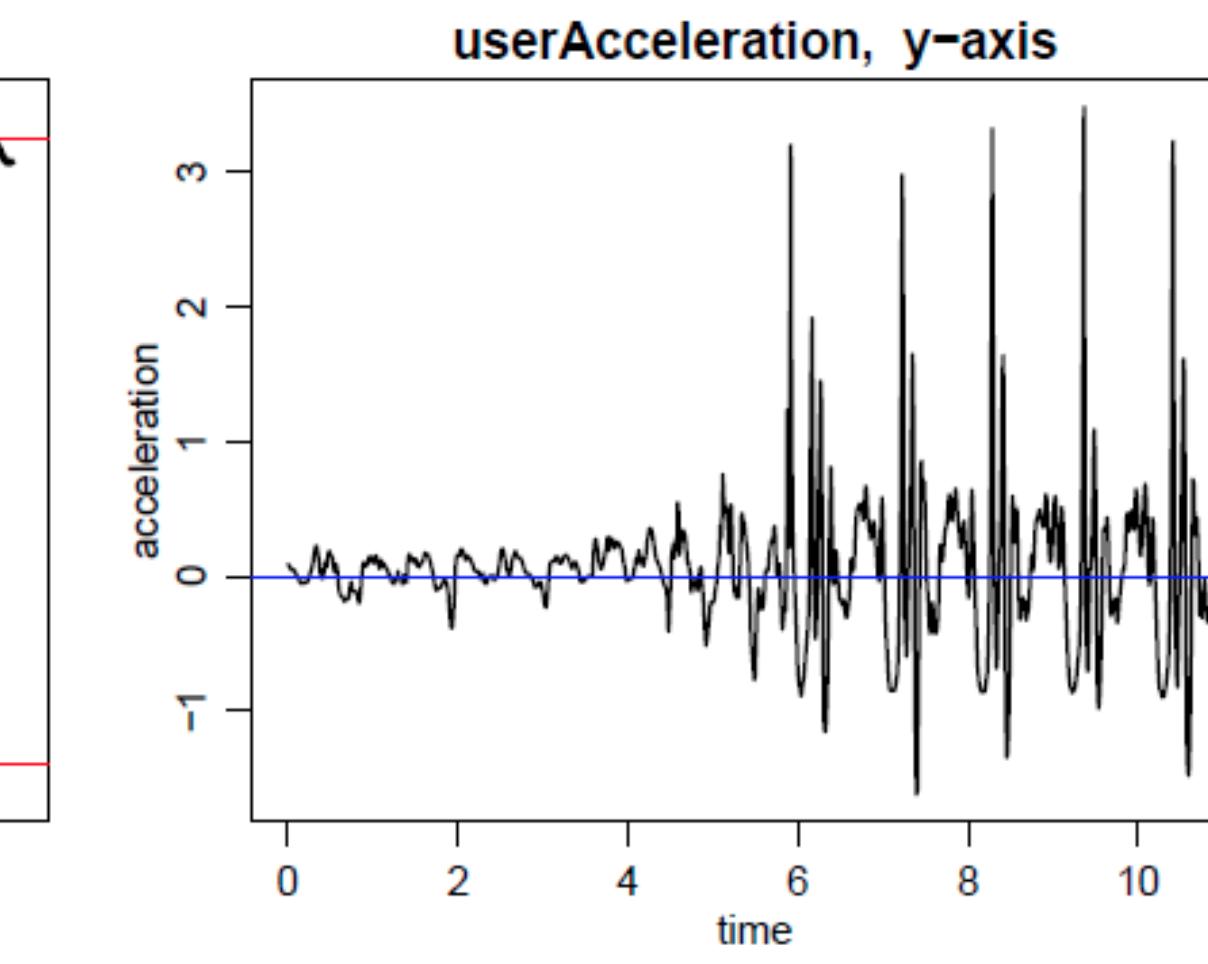
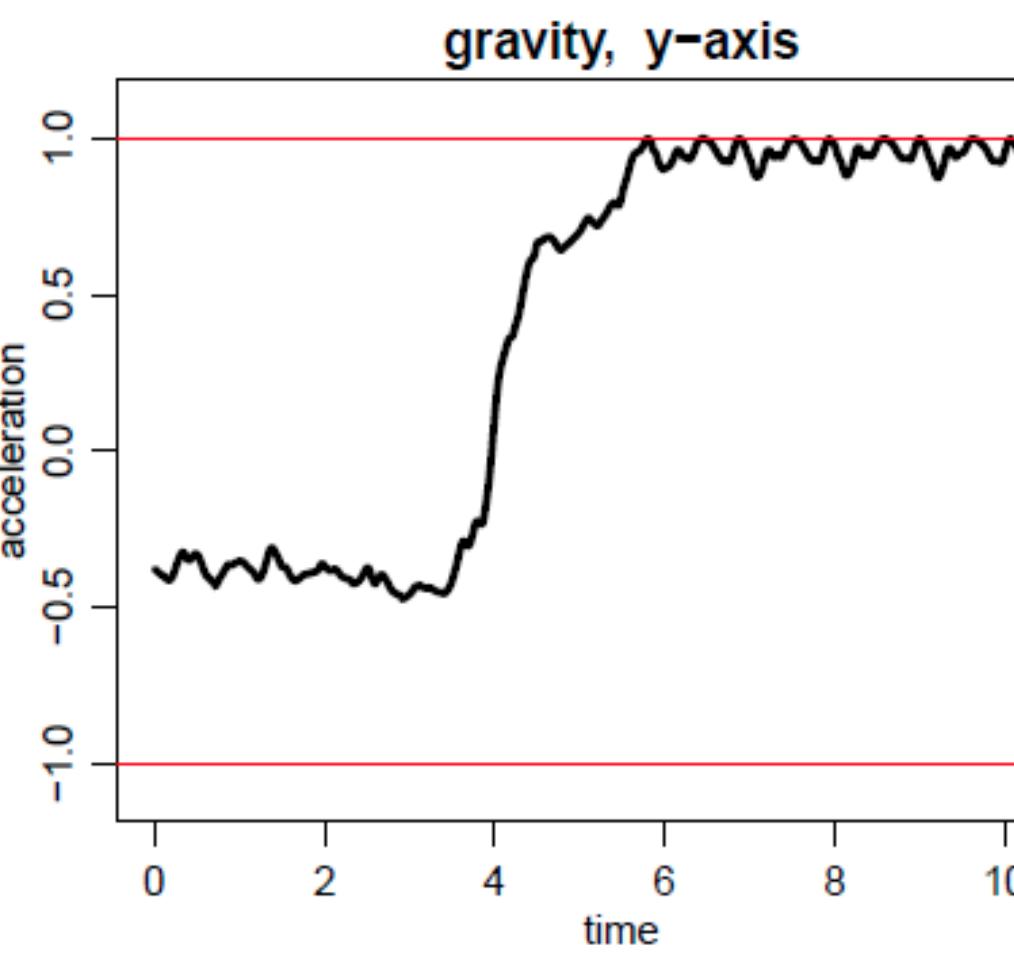
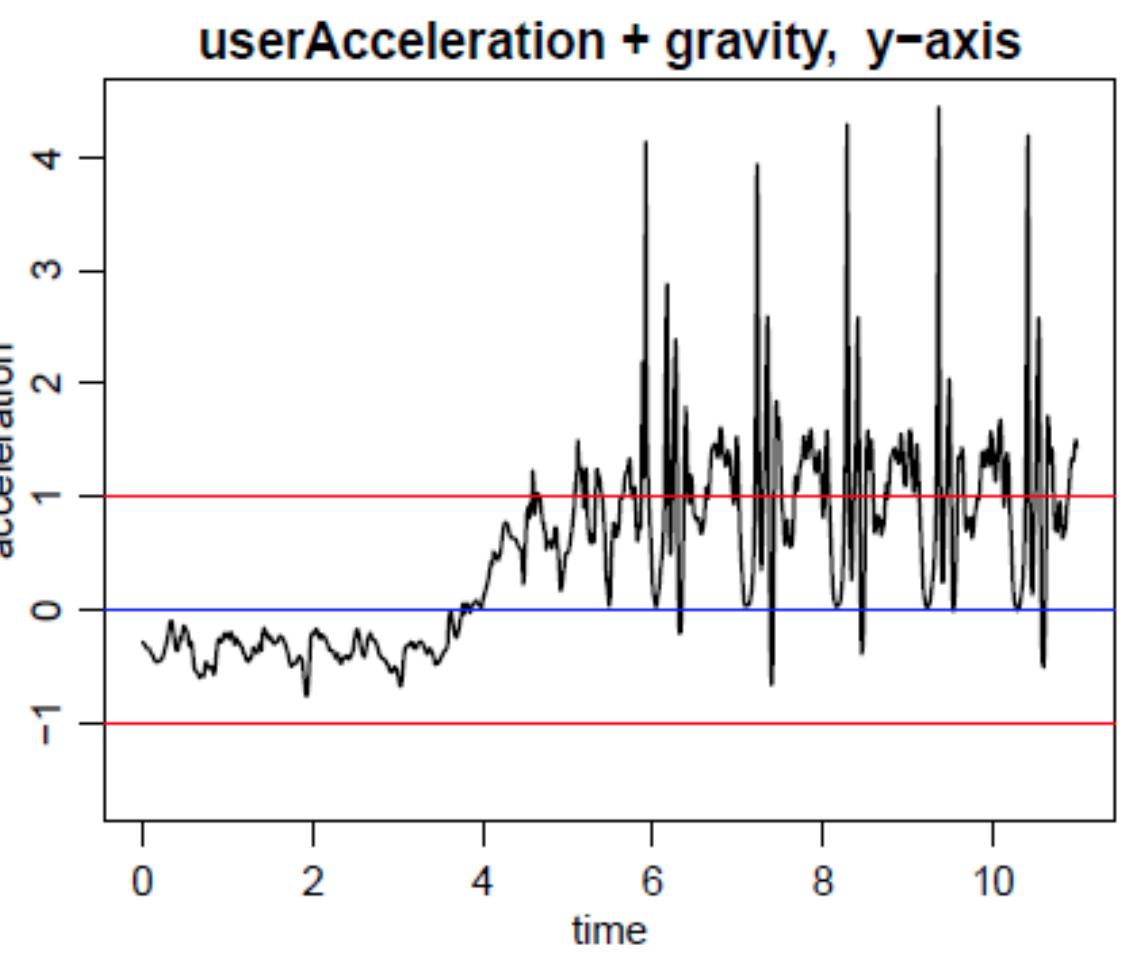
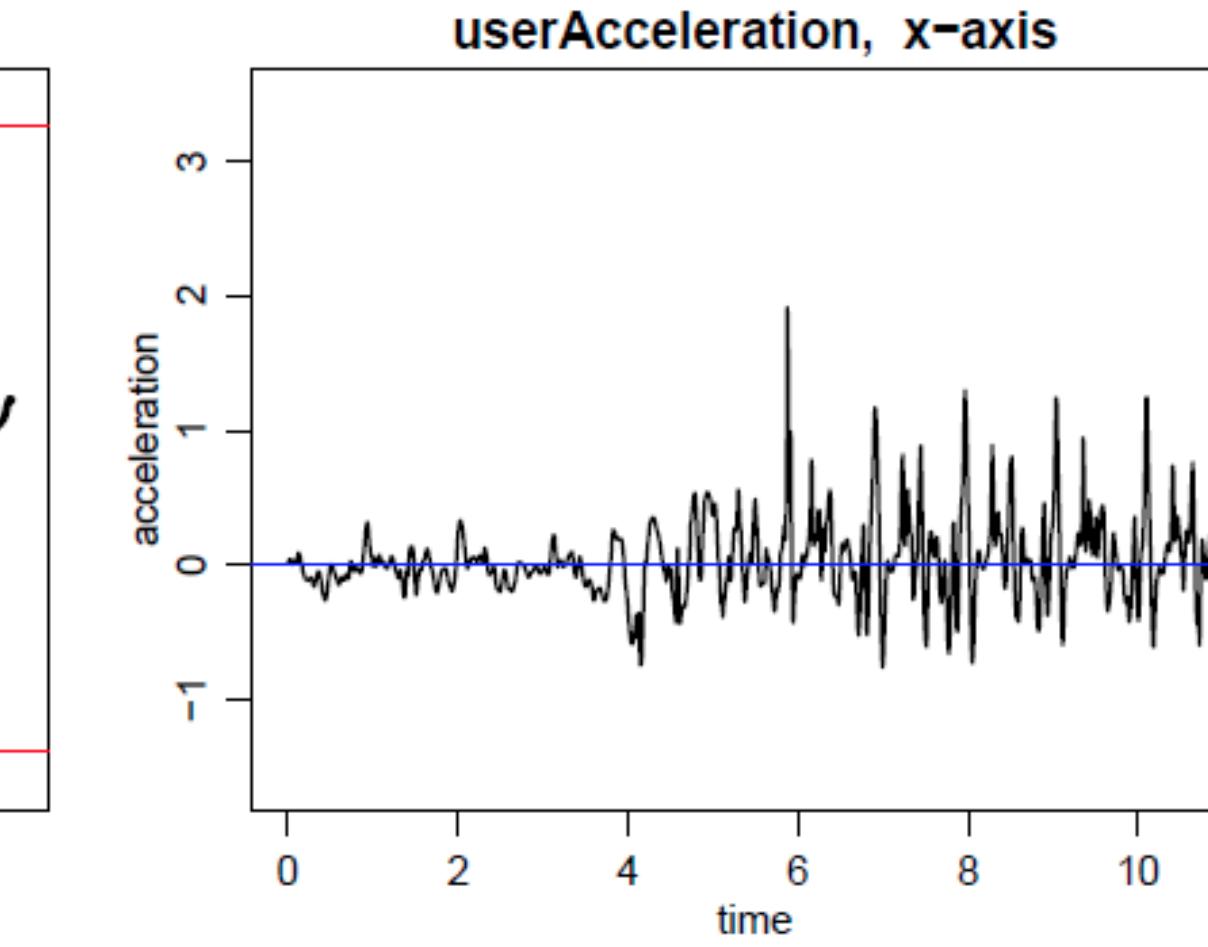
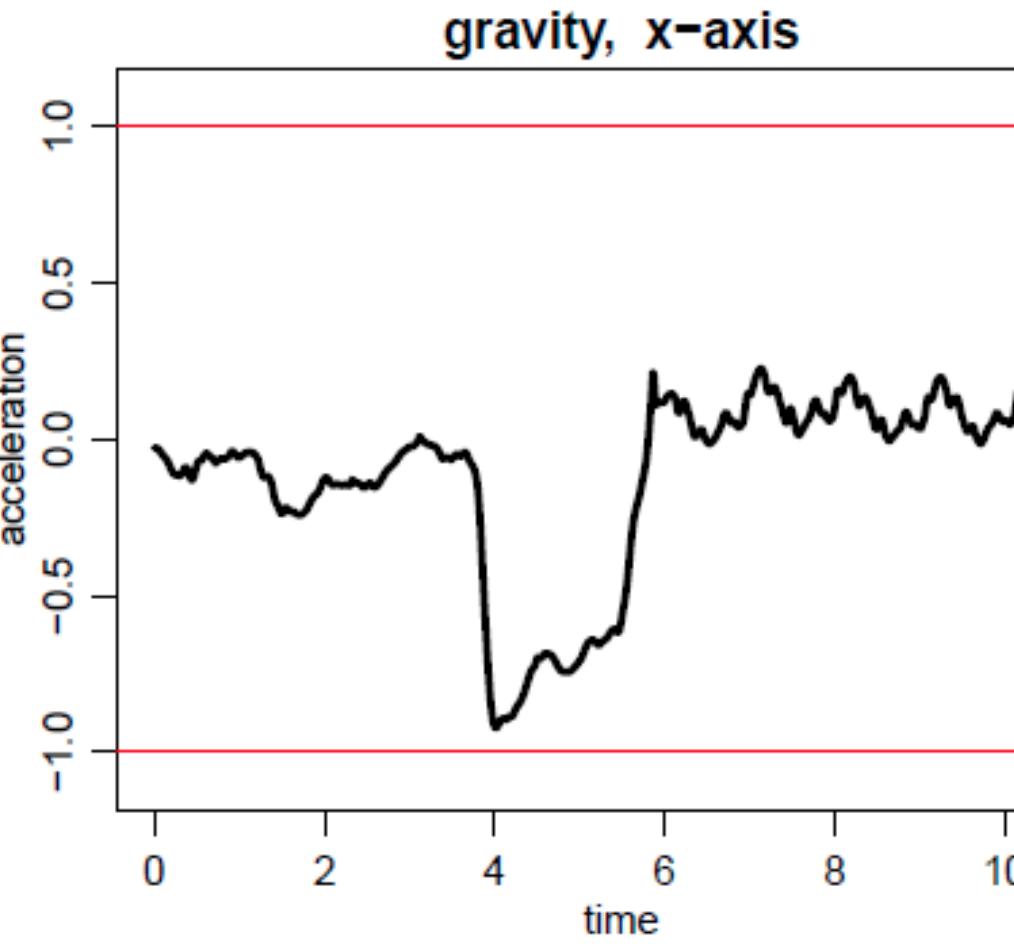
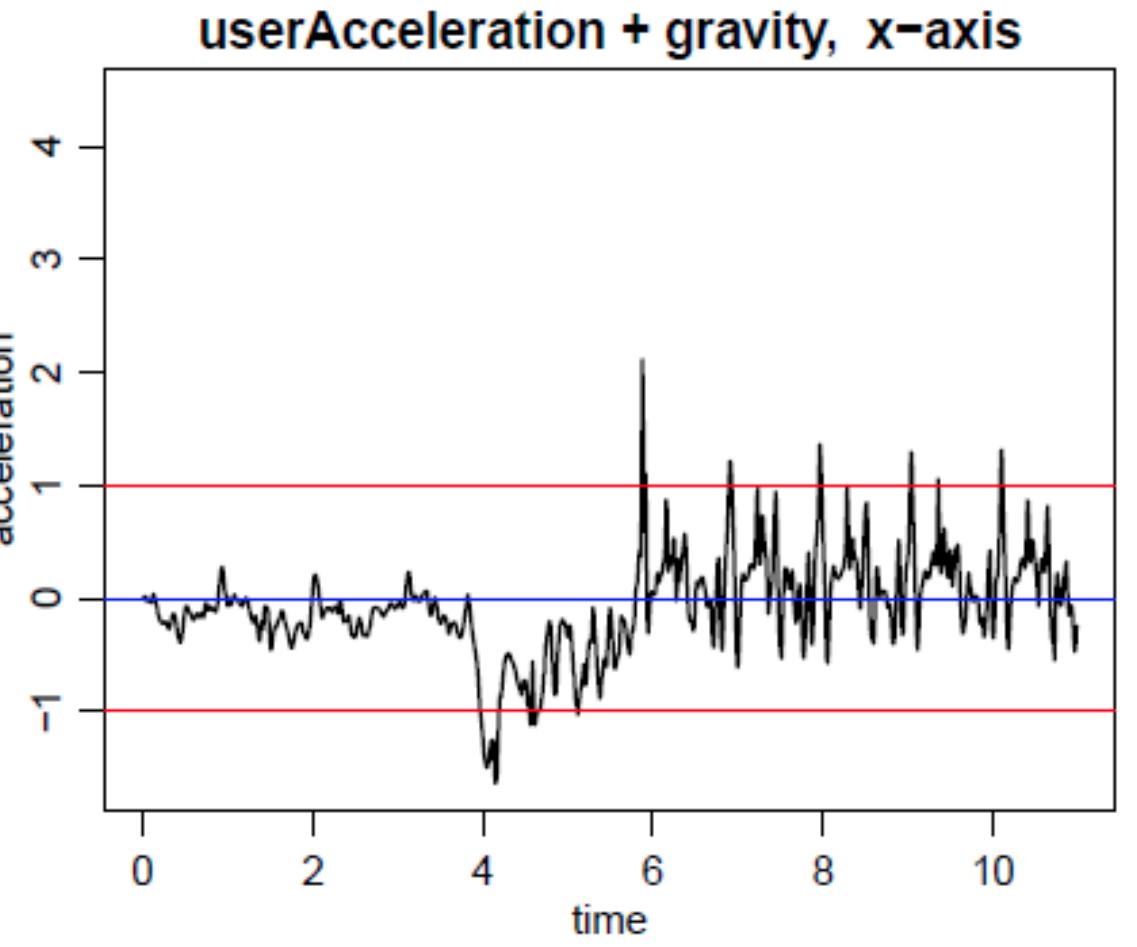
rotationRate

attitude

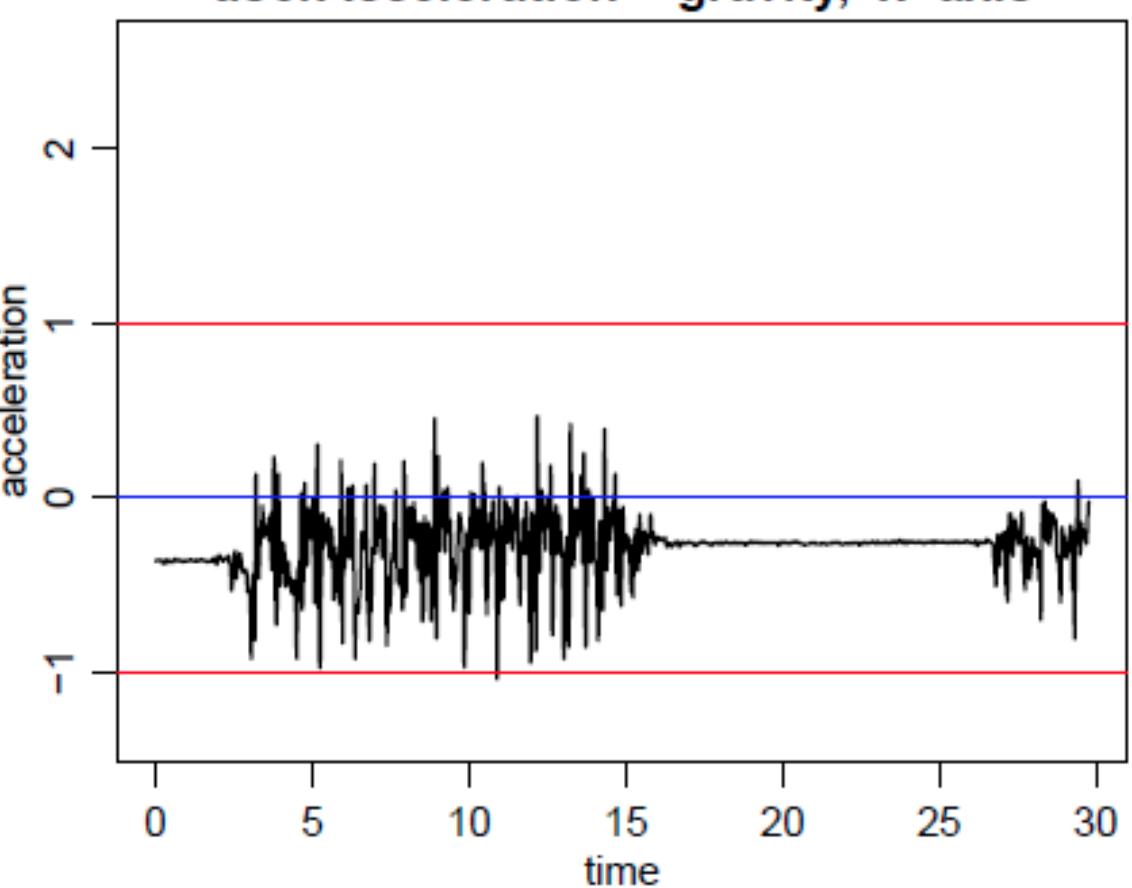


Monty Python's The Flying Circus

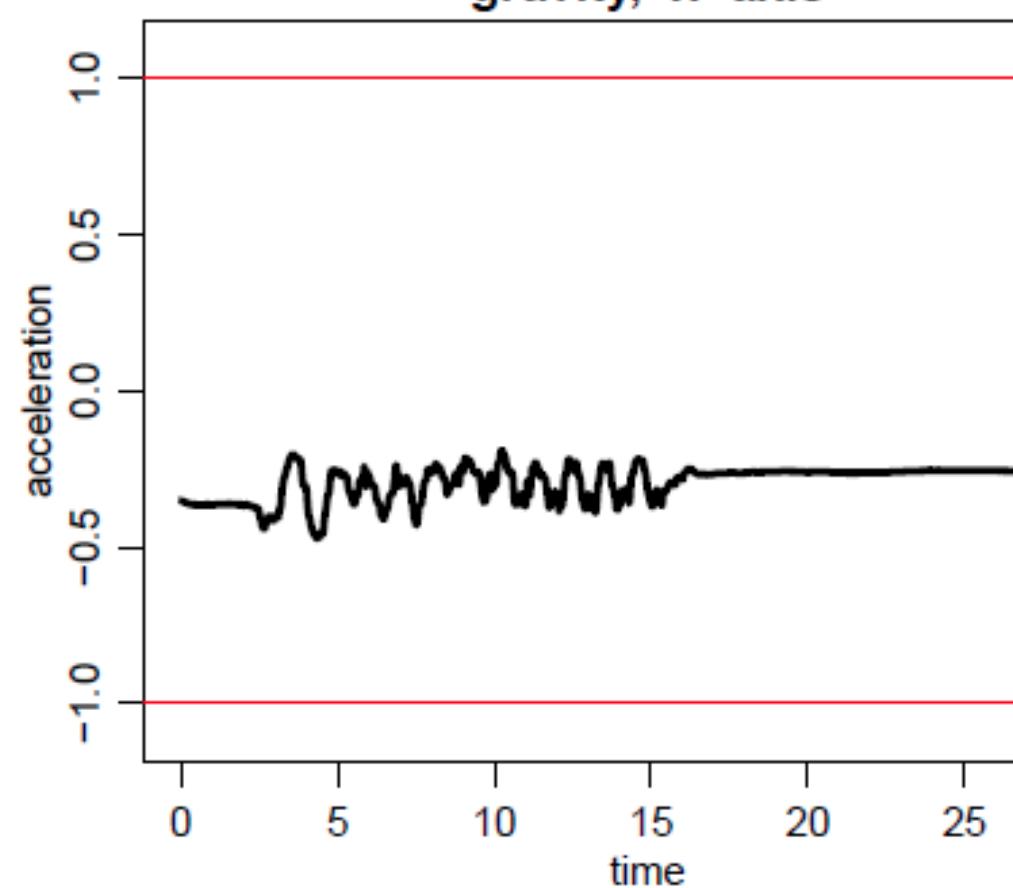




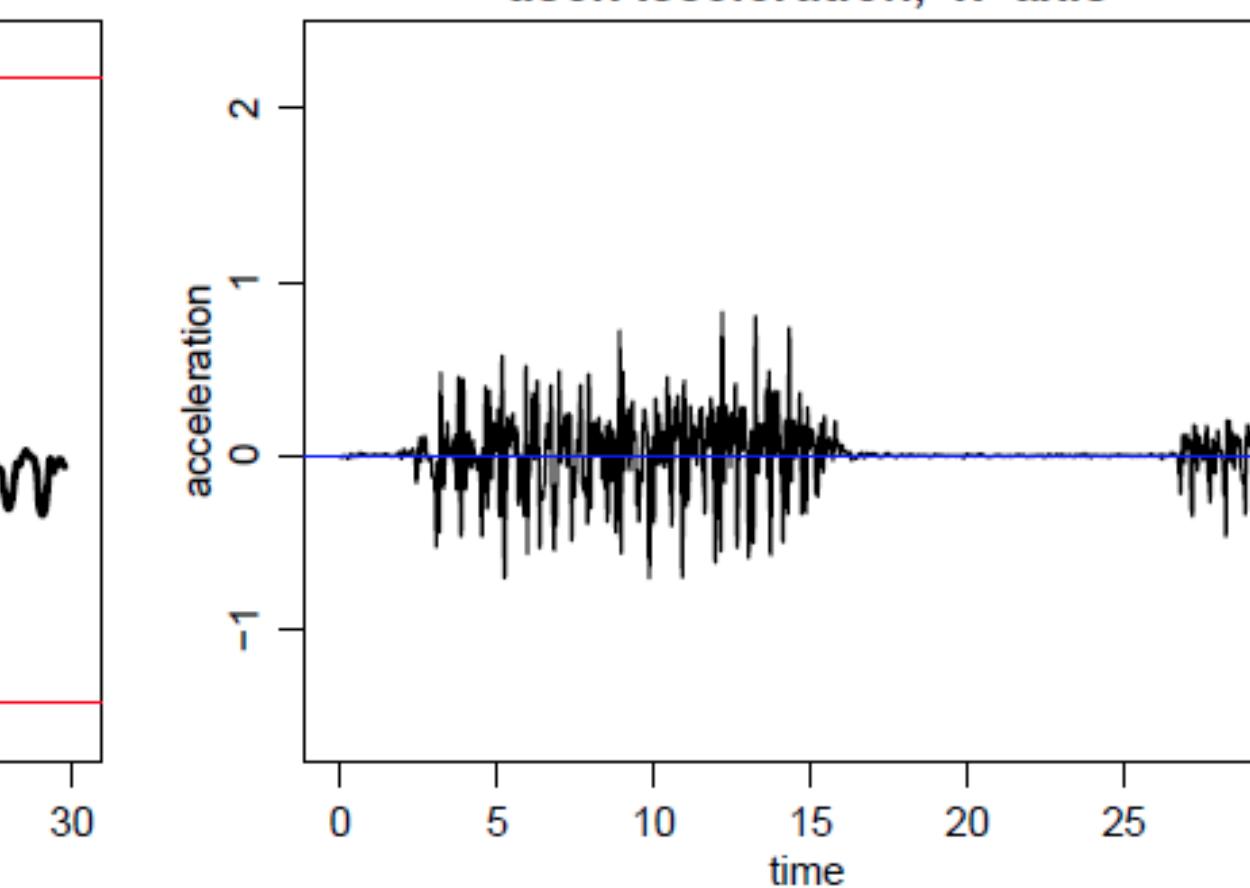
userAcceleration + gravity, x-axis



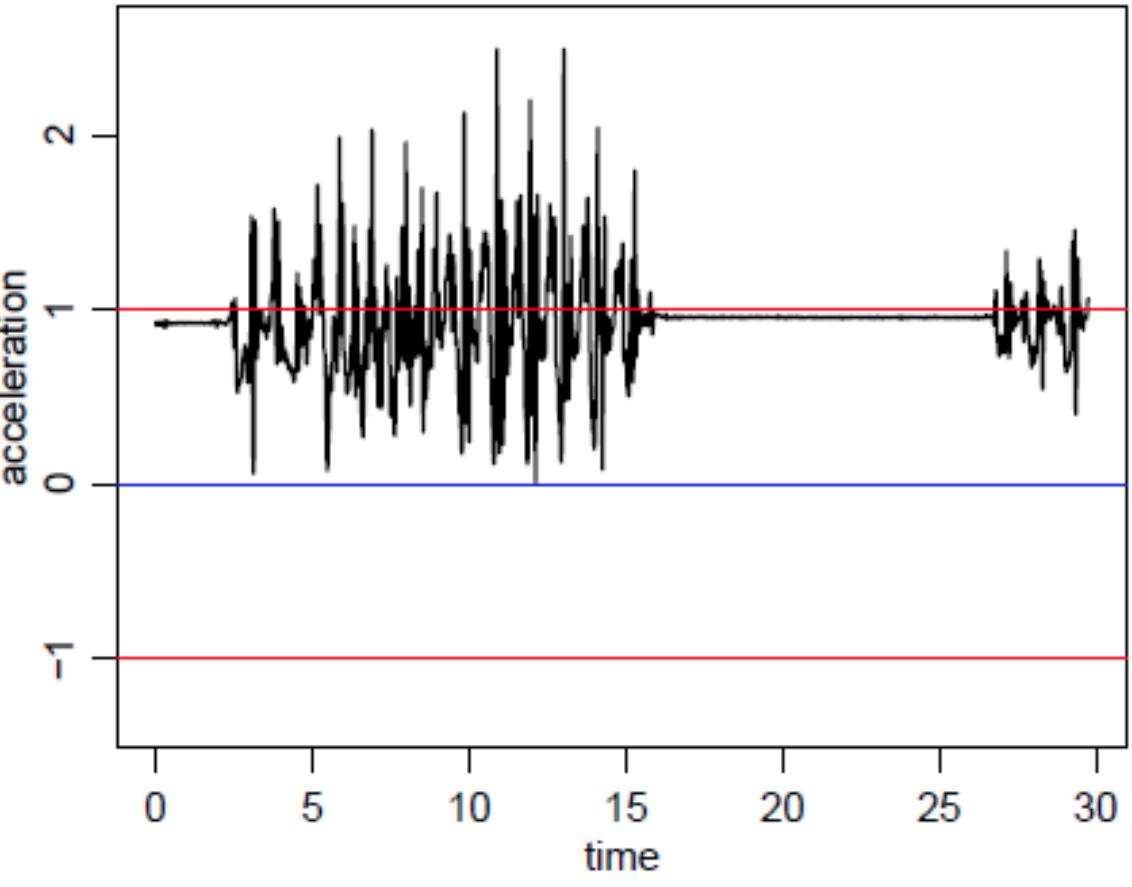
gravity, x-axis



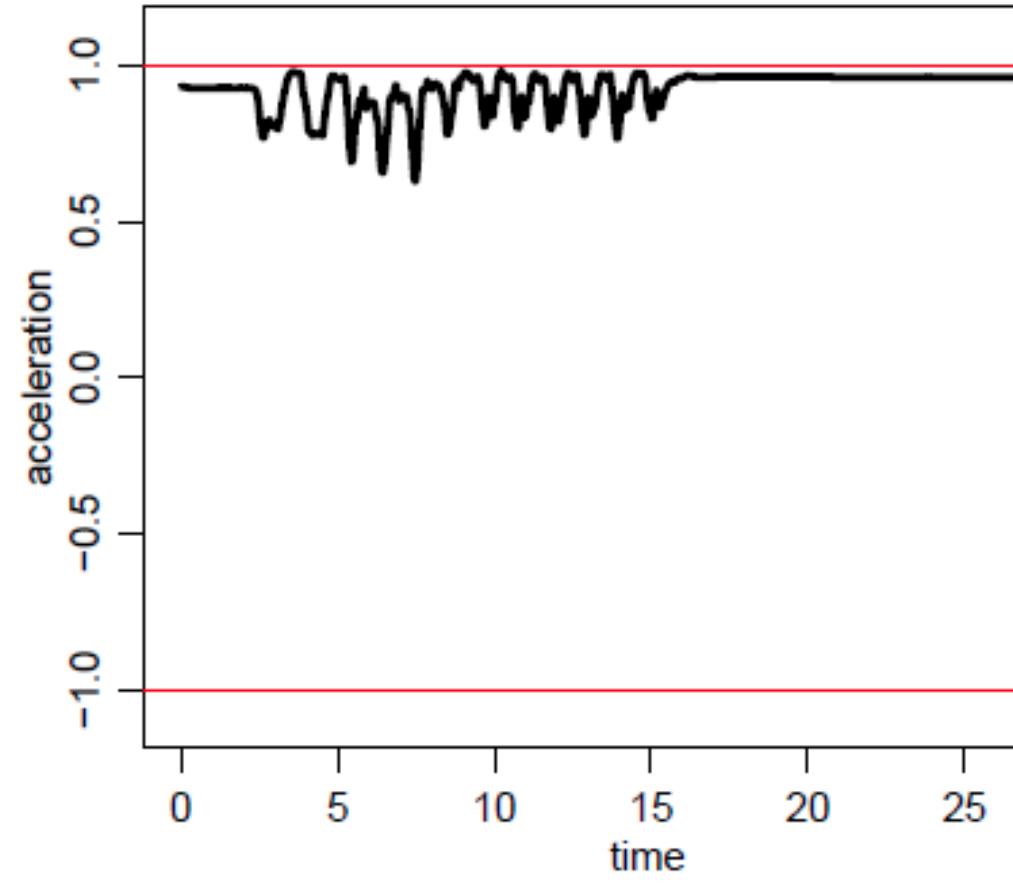
userAcceleration, x-axis



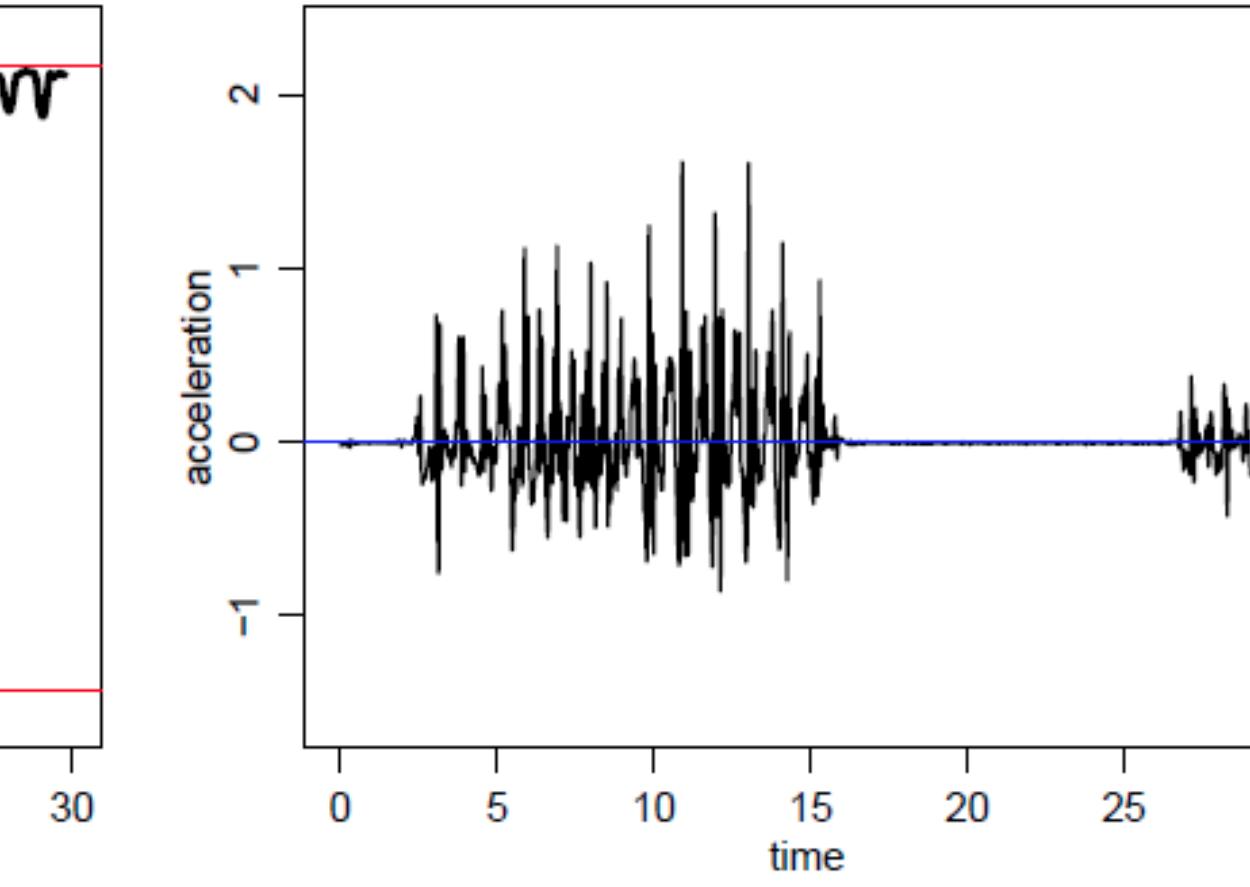
userAcceleration + gravity, y-axis



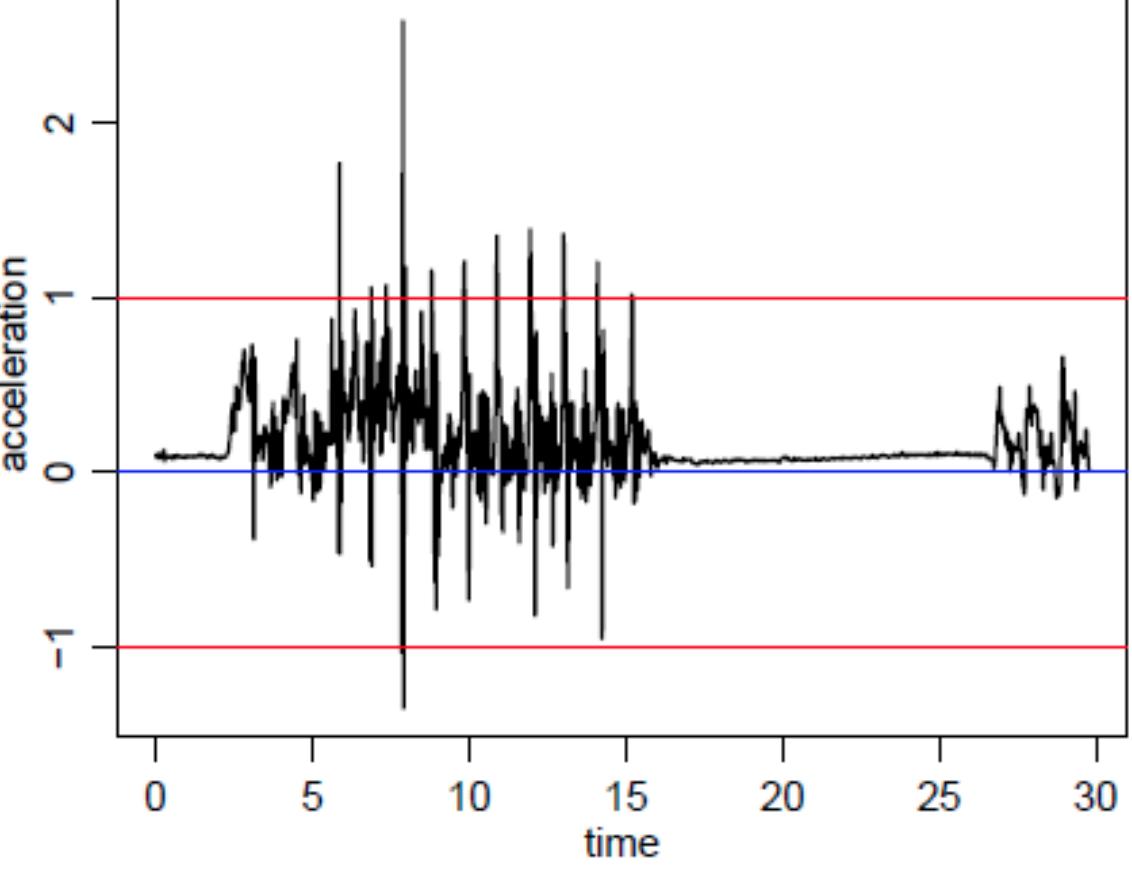
gravity, y-axis



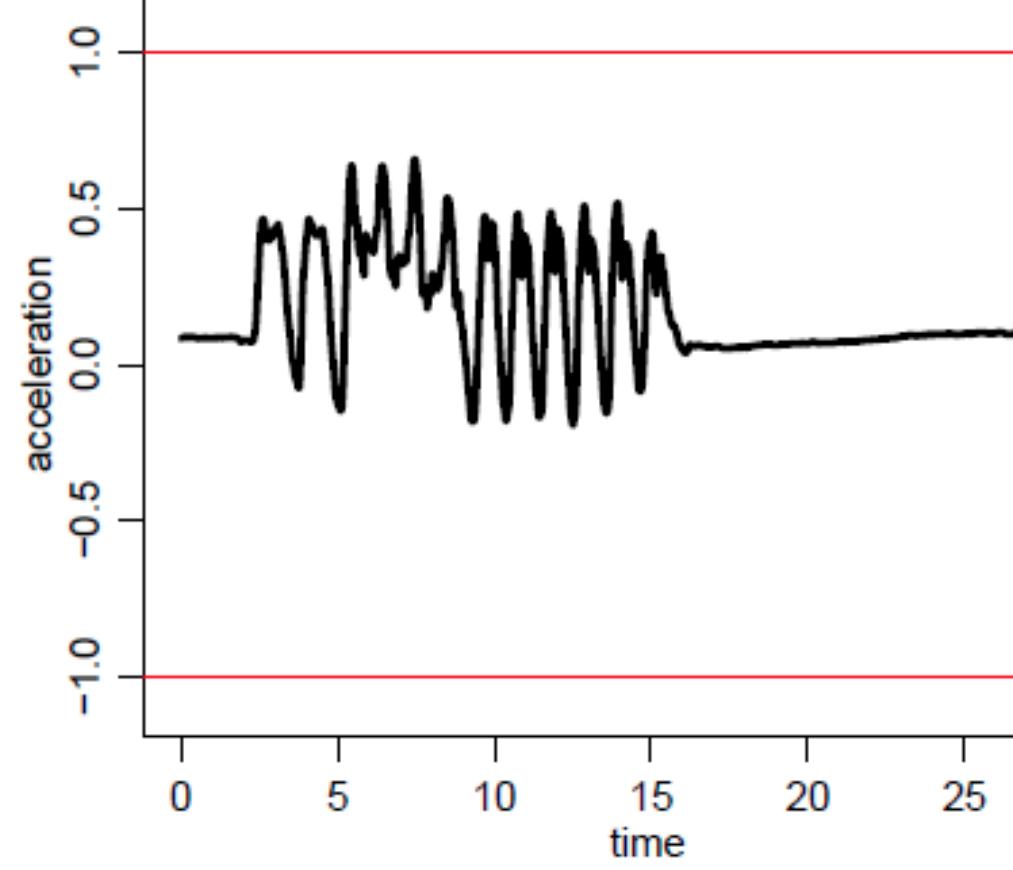
userAcceleration, y-axis



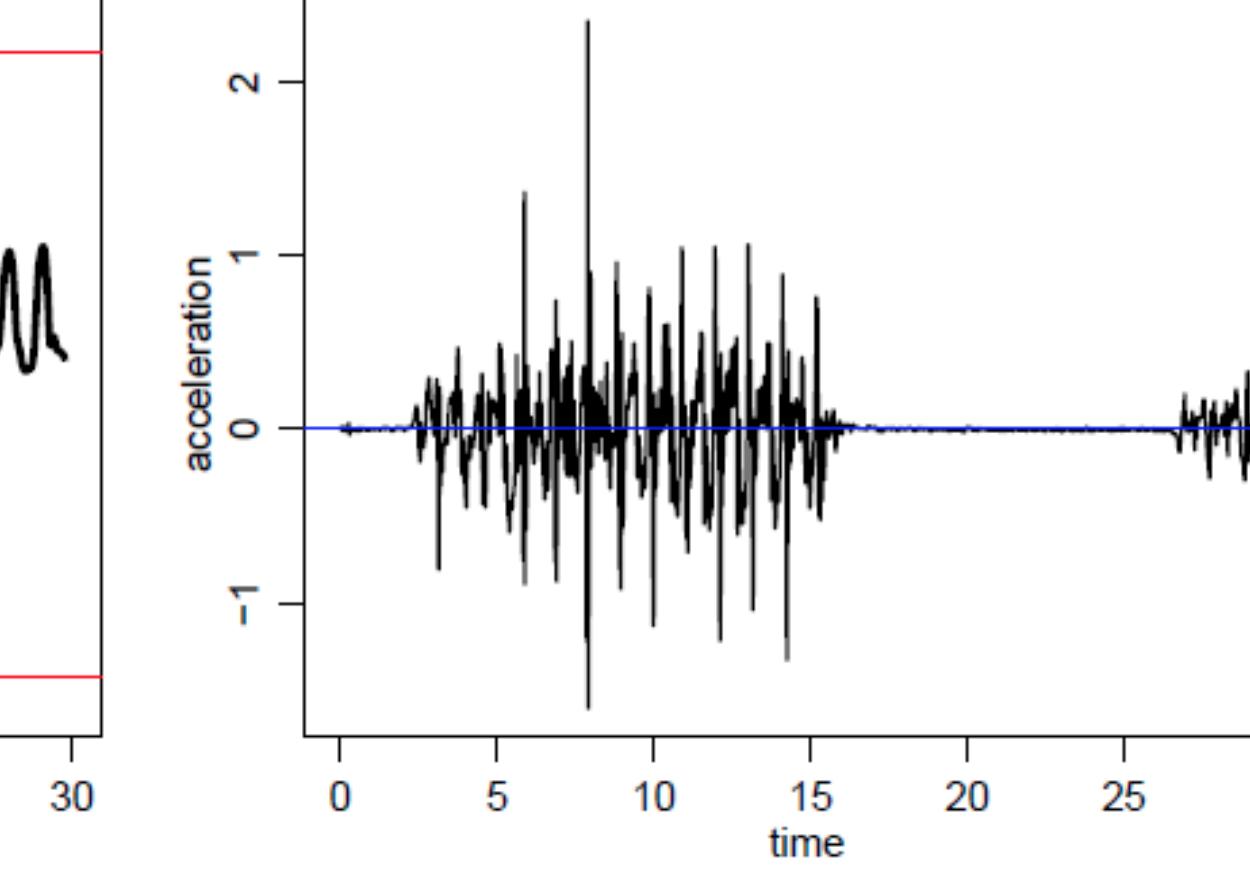
userAcceleration + gravity, z-axis

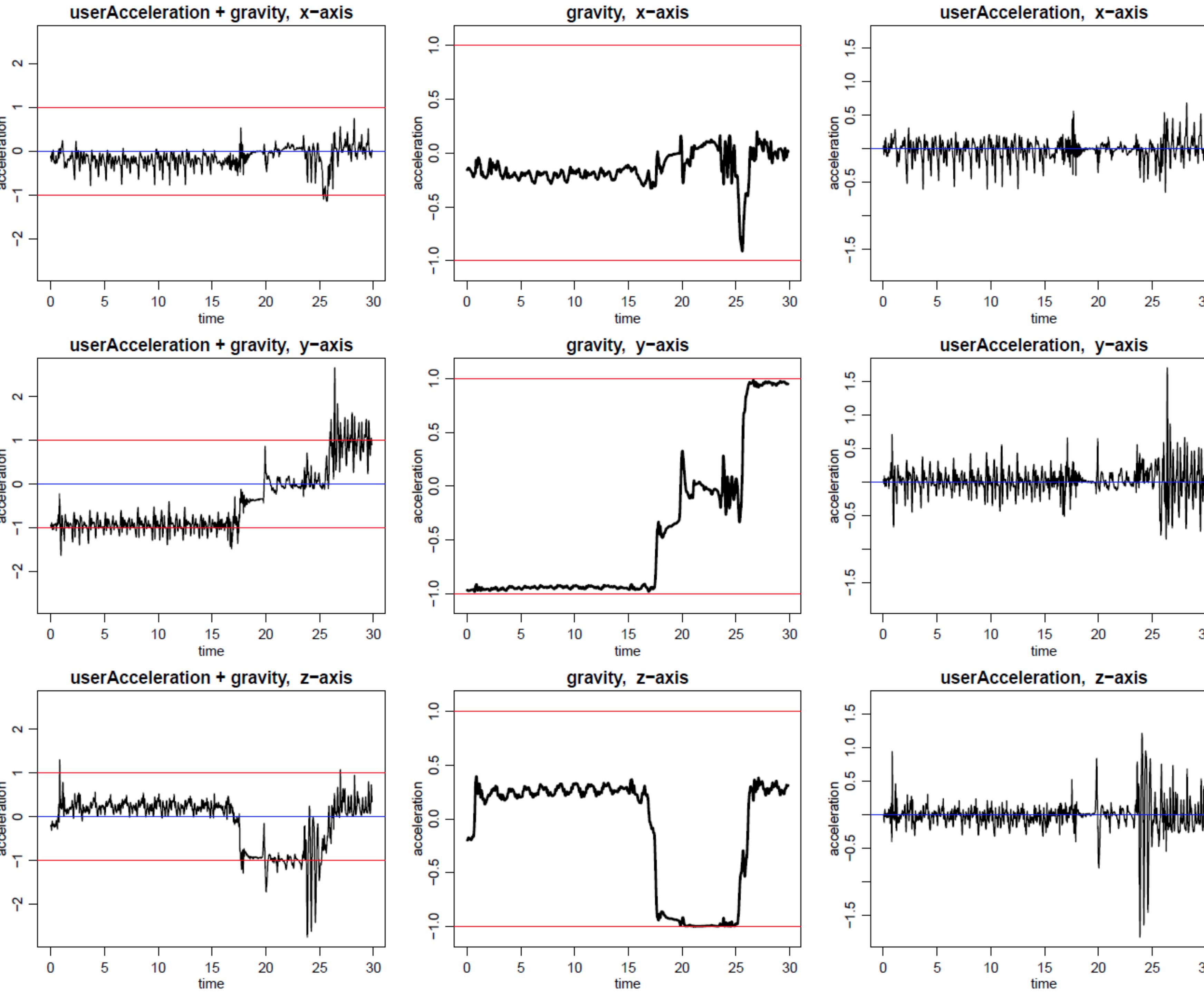


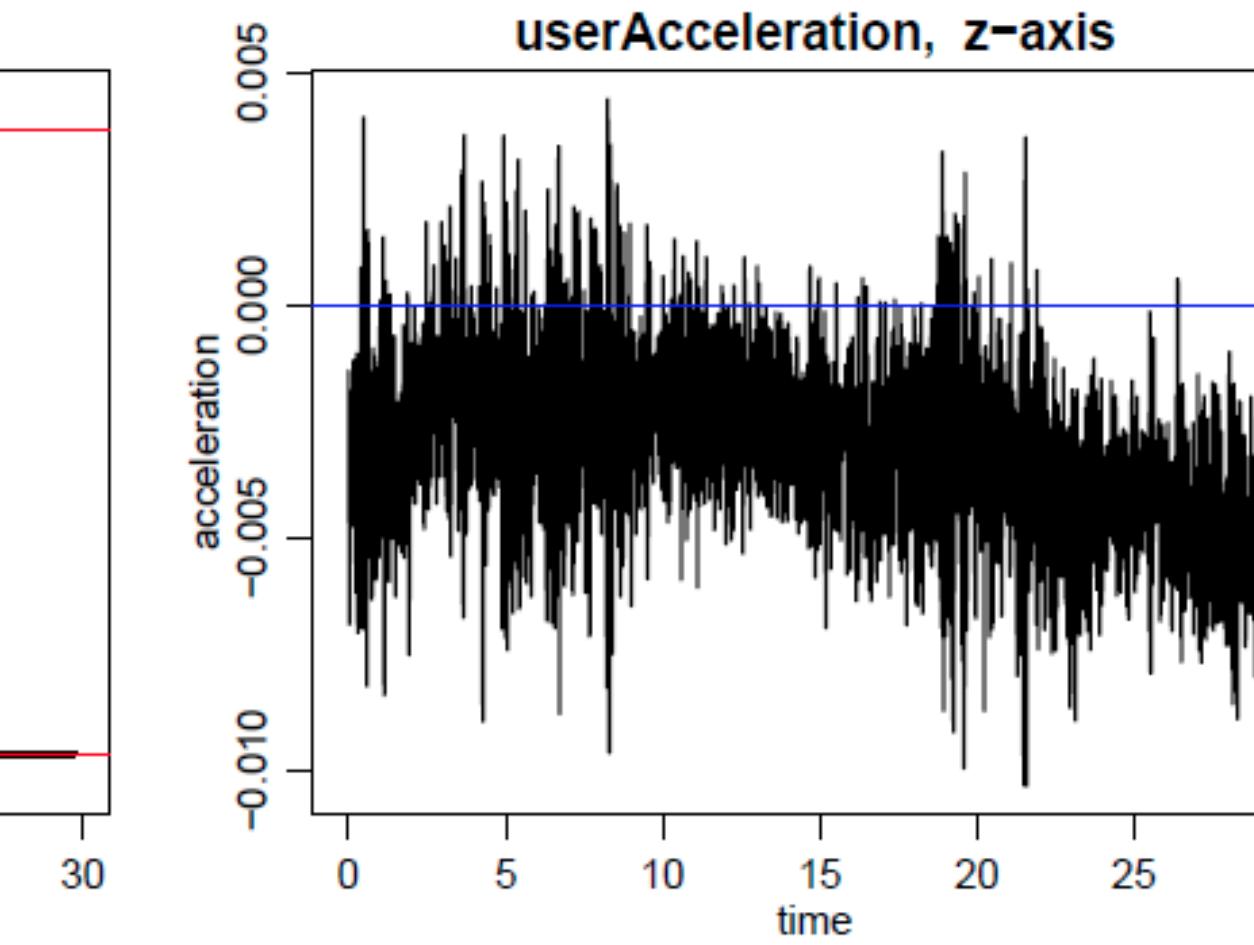
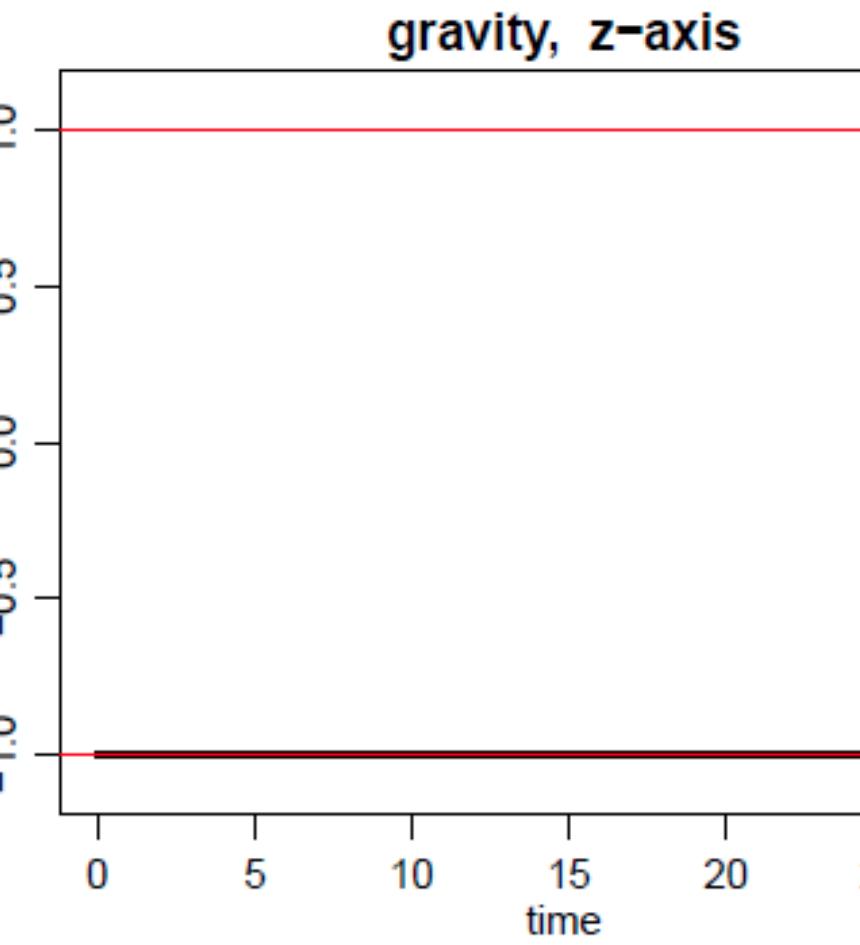
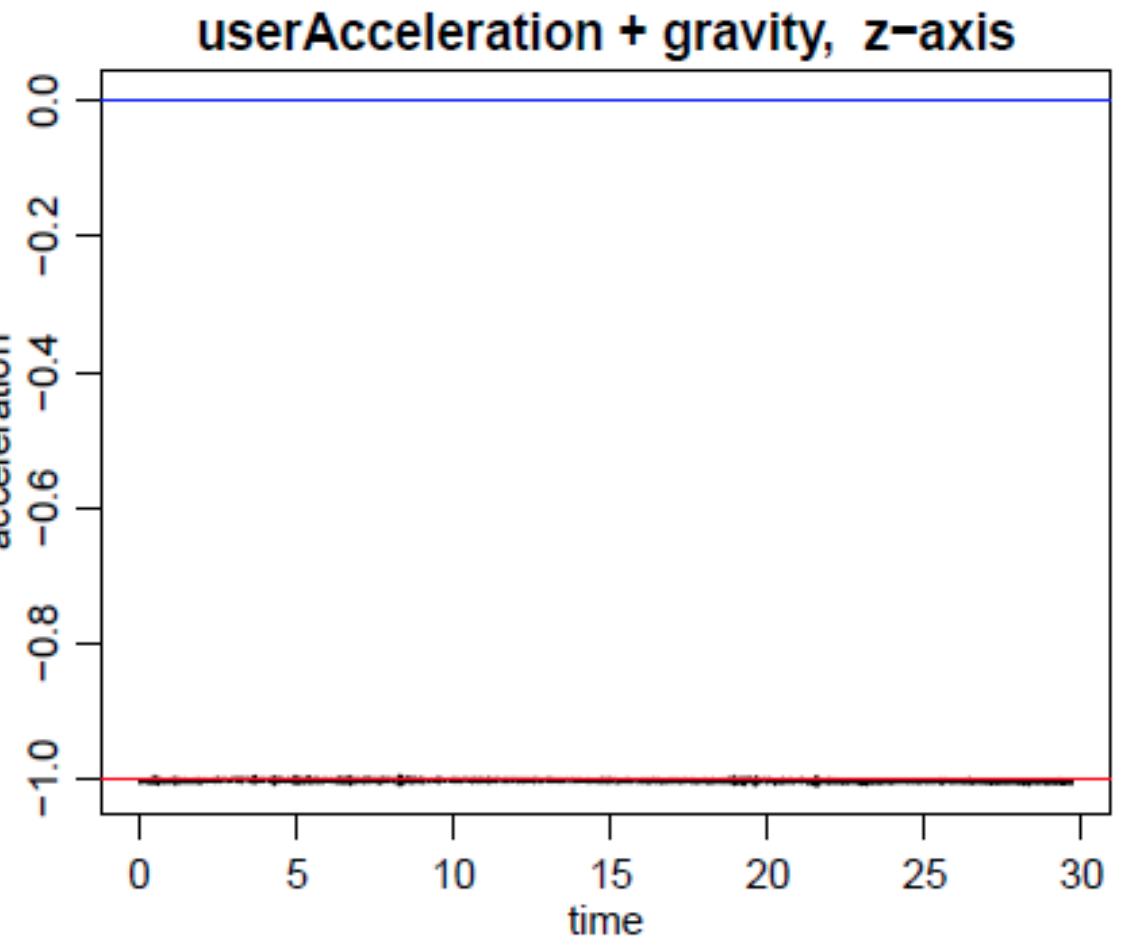
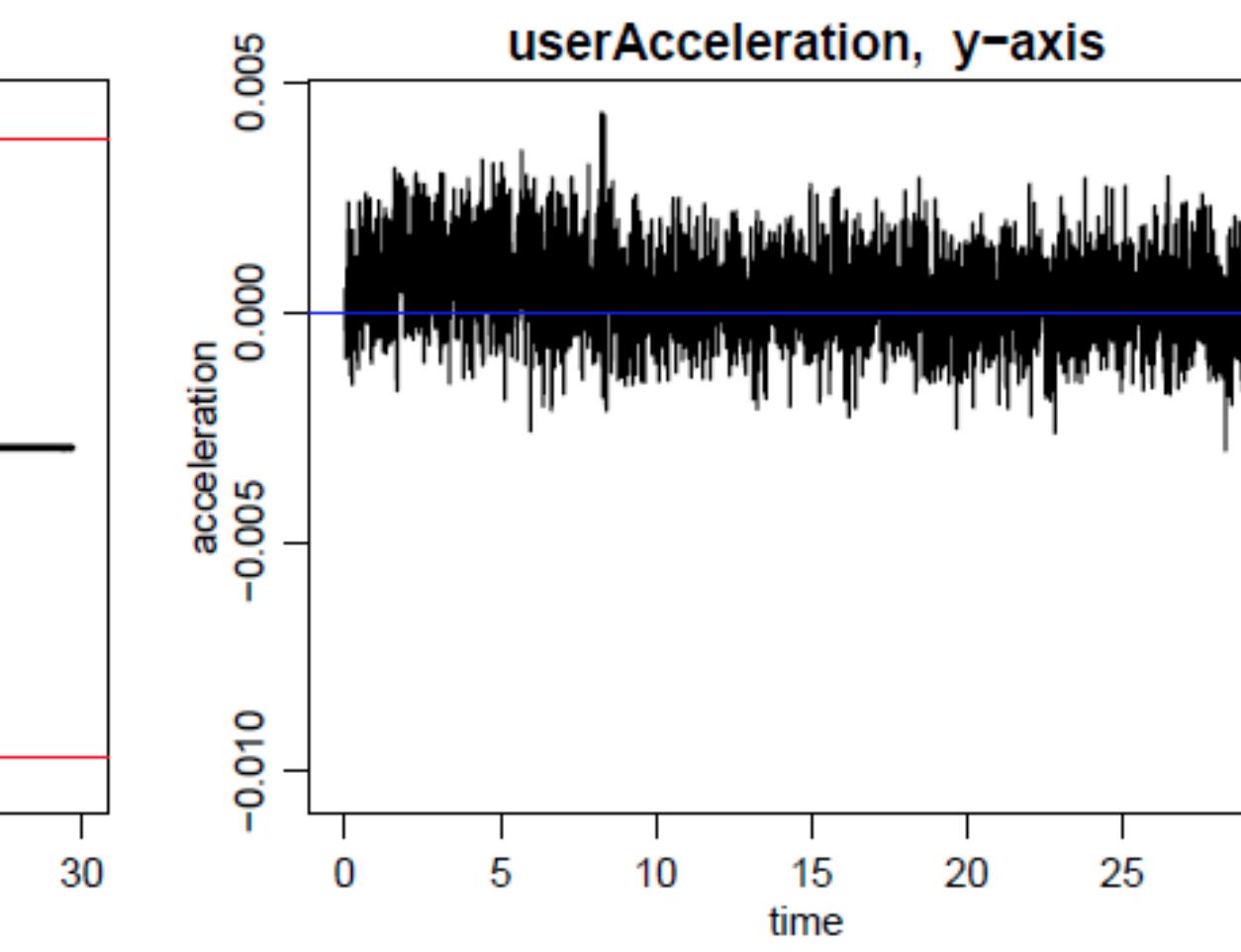
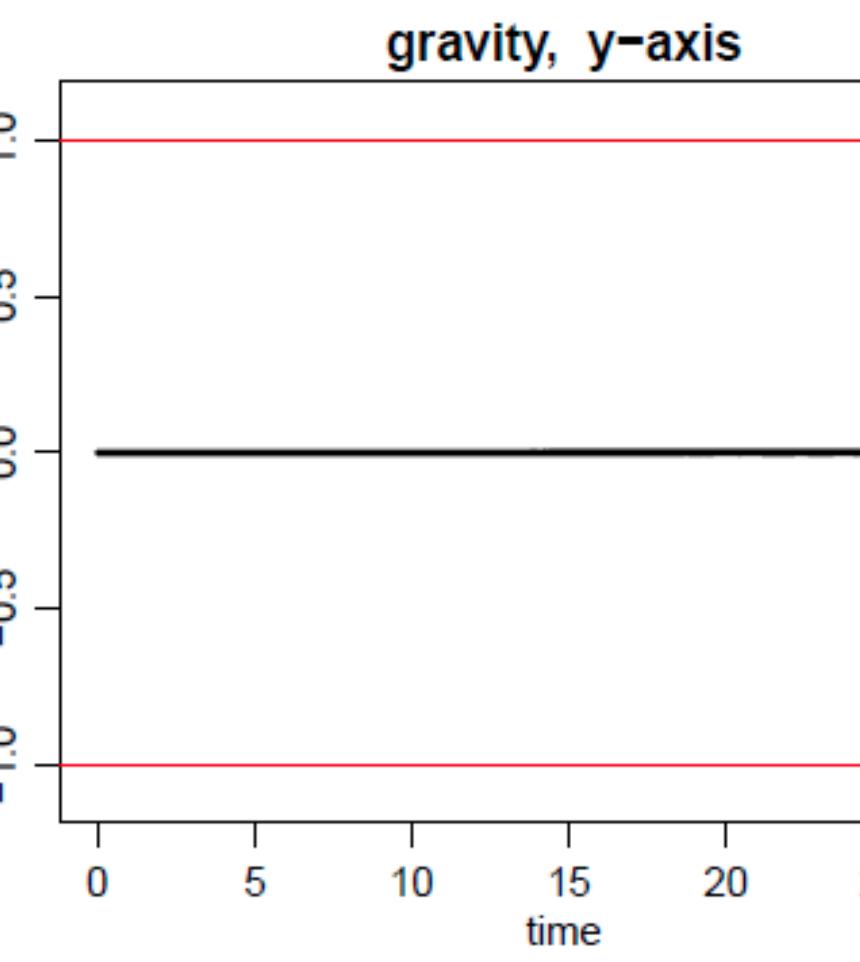
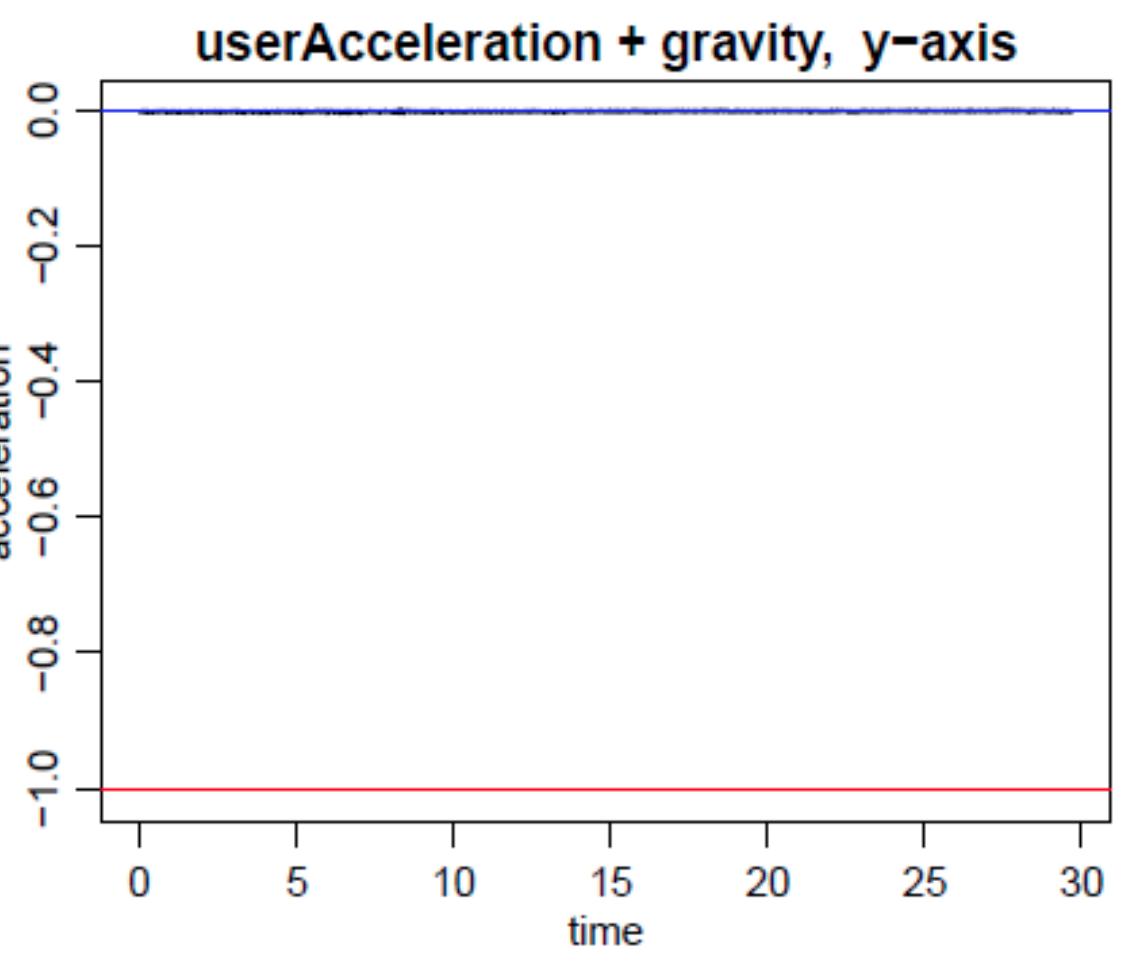
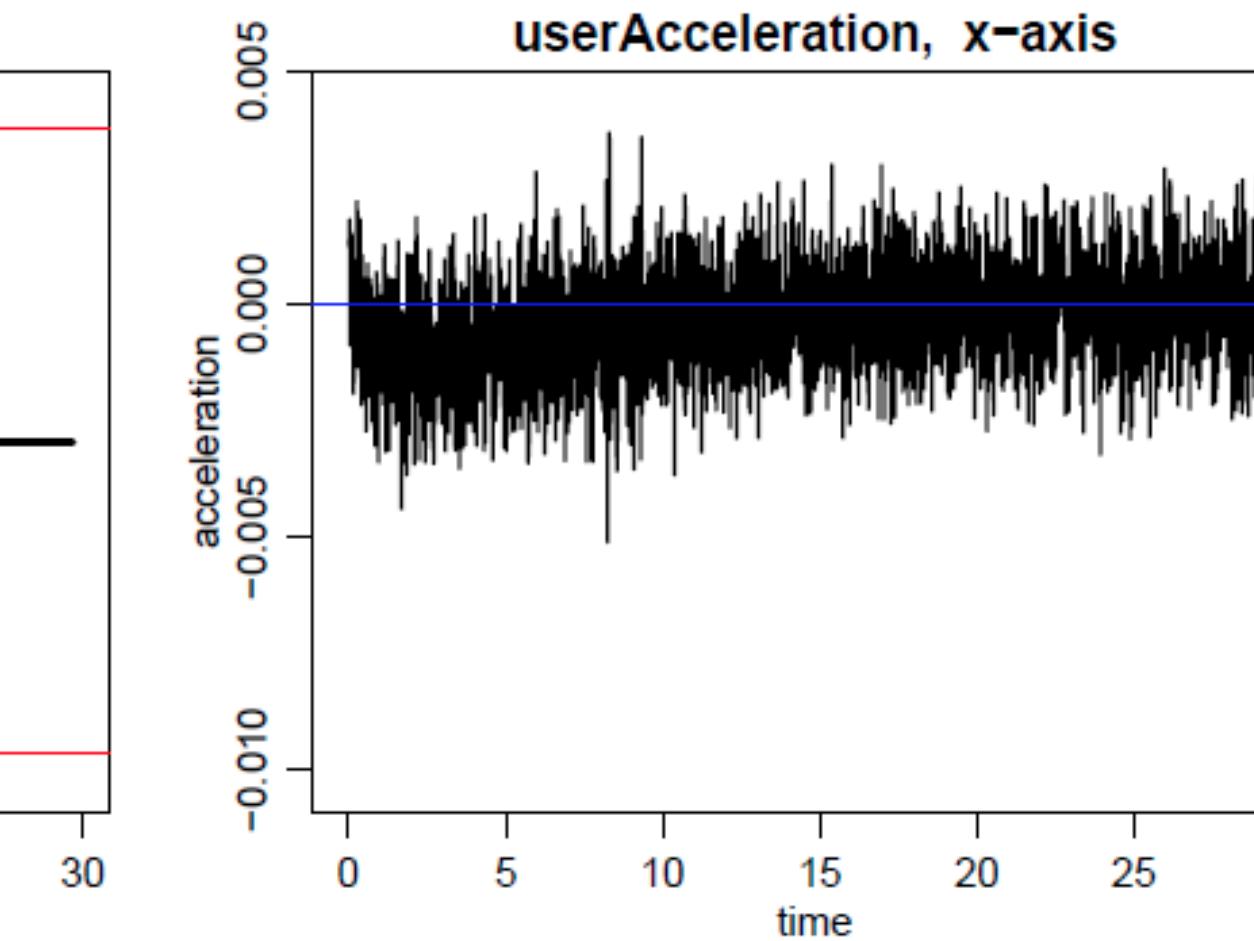
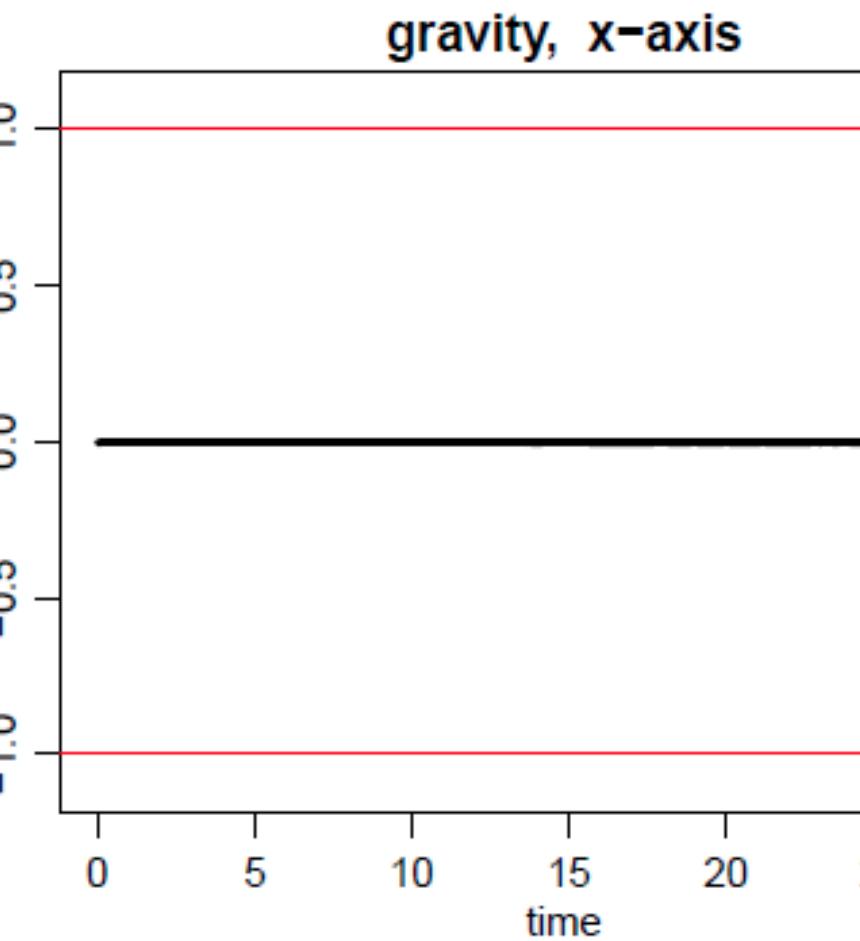
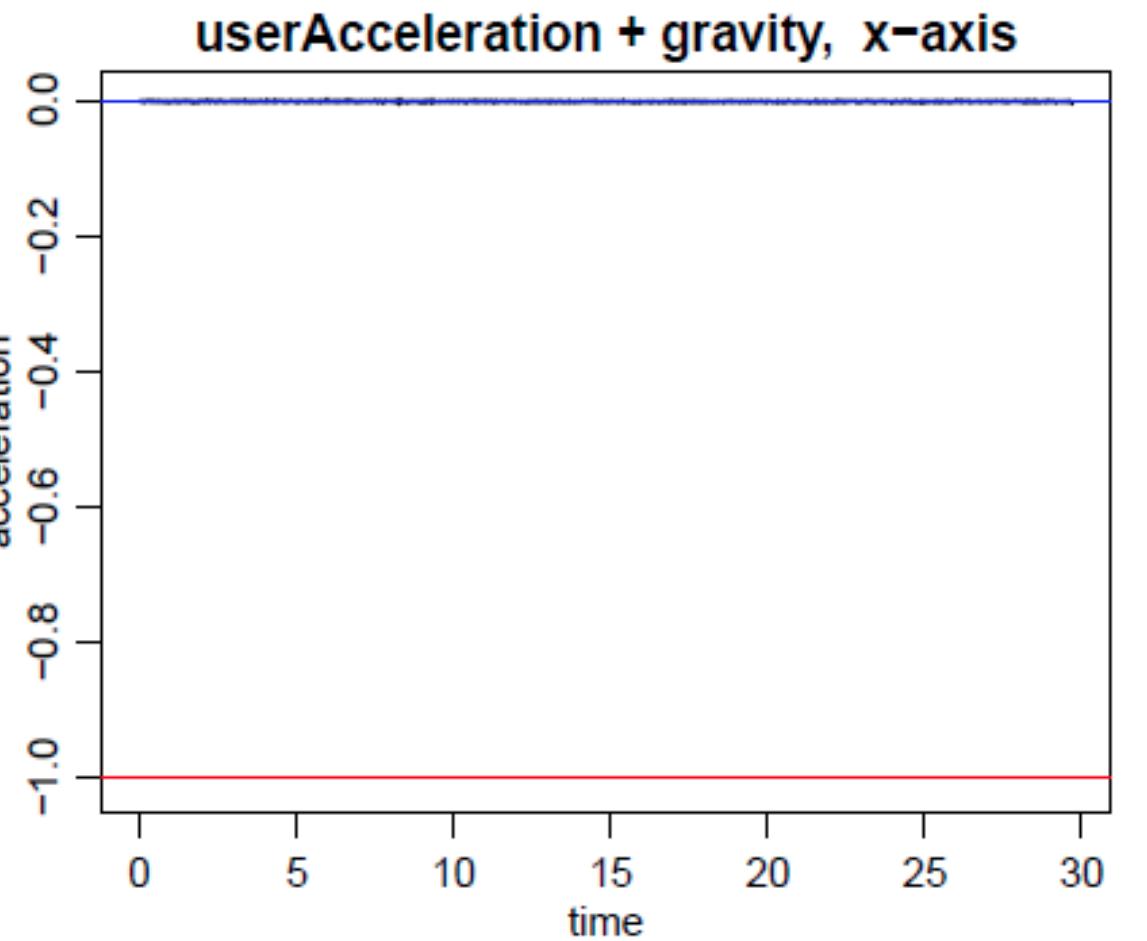
gravity, z-axis



userAcceleration, z-axis







user-selected sharing preferences

A screenshot of a mobile device displaying a research article from **nature biotechnology**. The article title is "First, design for data sharing" by John Wilbanks & Stephen H Friend. It has an Altmetric score of 274 and 11 citations. A "More detail" button is visible. The main text discusses the need for a framework that accounts for choices made by trial participants and researchers.

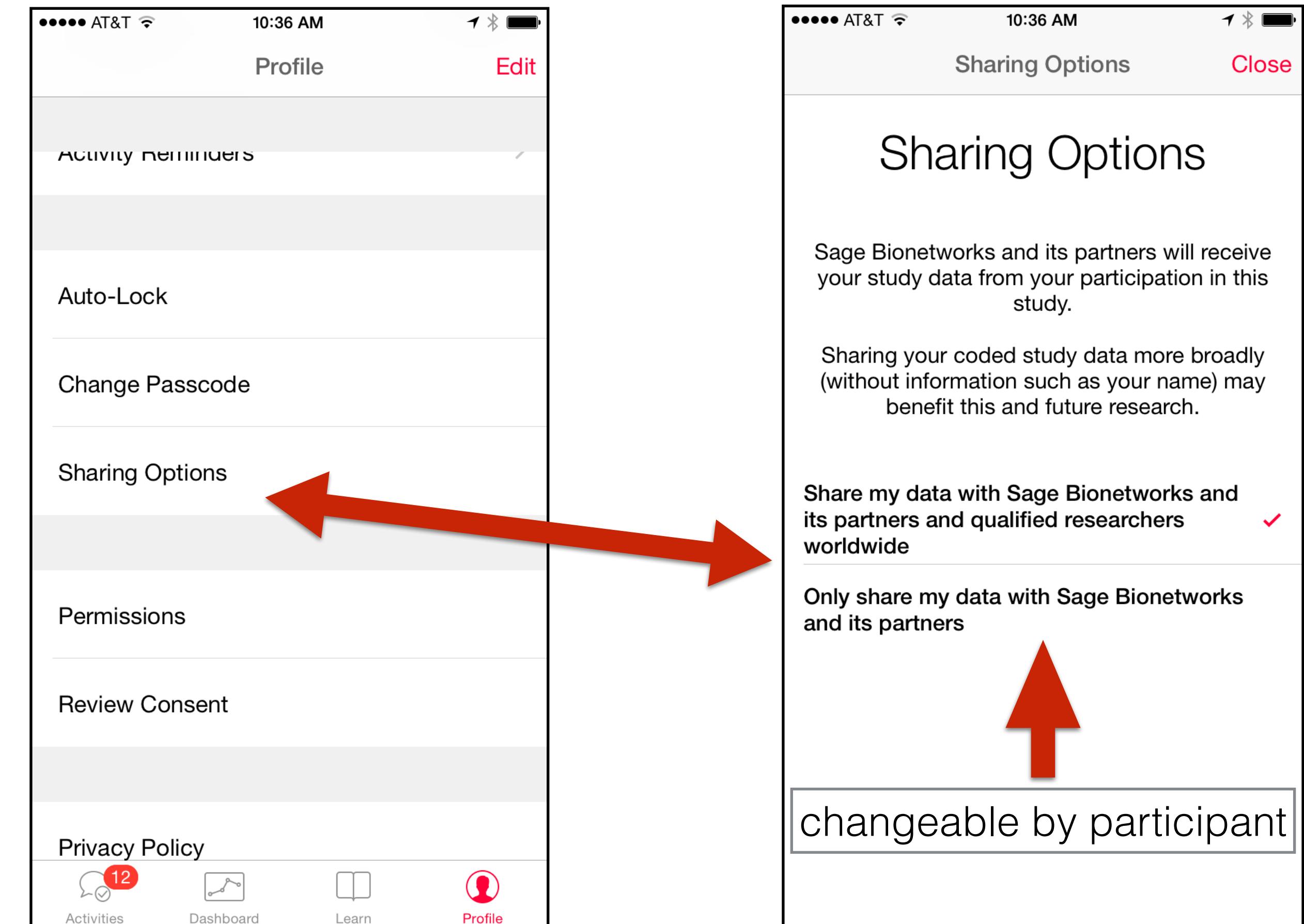
Altmetric: 274 Citations: 11 More detail »

Commentary | OPEN

First, design for data sharing

John Wilbanks & Stephen H Friend ✉

To upend current barriers to sharing clinical data and insights, we need a framework that not only accounts for choices made by trial participants but also qualifies researchers wishing to access and analyze the data.



SCIENTIFIC DATA



Home | Archive | About ▾ | For Authors ▾ | For Referees | Data Policies ▾ | Collections ▾

Home > Data Descriptors > Data Descriptor

SCIENTIFIC DATA | DATA DESCRIPTOR **OPEN**

The mPower study, Parkinson disease mobile data collected using ResearchKit

Brian M. Bot, Christine Suver, Elias Chaibub Neto, Michael Kellen, Arno Klein, Christopher Bare,
Megan Doerr, Abhishek Pratap, John Wilbanks, E. Ray Dorsey, Stephen H. Friend & Andrew D. Trister

promote an ecosystem
where research is conducted
for others to consume

9,520 unique participants

8,320 completed at least one task

1,087 self reported parkinson diagnosis

198,639 total activities and surveys completed

scope of data collection

task name	type of task and schedule	unique participants	unique tasks
demographics	survey - once	6,805	6,805
MDS-UPDRS	survey - monthly	2,024	2,305
PDQ8	survey - monthly	1,334	1,641
memory	activity - t.i.d.	968	8,569
tapping	activity - t.i.d.	8,003	78,887
voice	activity - t.i.d.	5,826	65,022
walking	activity - t.i.d.	3,101	35,410



study data released before any primary publication
150+ independent ‘qualified researchers’

MIT
The Ohio State University
University of Otago
University of Texas Health Science Center
Istanbul Sehir University
University of Iowa
University of Virginia
University of Toronto
Johns Hopkins University
Vanderbilt University
University of Rochester
McGill University
Xi'an Jiaotong University
University of Washington
Harvard University

Parkinson's Disease Foundation
Eli Lilly
AstraZeneca
Apple
Verily
Intel
Infocepts
Posit Science

use case 2: molecular profiling in colorectal cancer

crc subtyping consortium



Subtypes of primary colorectal tumors correlate with response to targeted treatment in colorectal cell lines

Andreas Schlicker¹, Garry Beran³, Christine M Chresta³, Gael McWalter⁴, Alison Pritchard³, Susie Weston⁴, Sarah Runswick⁴, Sara Davenport³, Kerry Heathcote³, Denis Alferez Castro³, George Orphanides³, Tim French^{4*} and Lodewyk FA Wessels^{1,2,5*}



A colorectal cancer classification system that associates cellular phenotype and responses to therapy

Anguraj Sadanandam, Costas A Lyssiotis, Krisztian Homicsko, Eric A Collisson, William J Gibb, Stephan Wullschleger, Liliane C Gonzalez Ostos, William A Lannon, Carsten Grotzinger, Maguy Del Rio, Benoit Lhermitte, Adam B Olshen, Bertram Wiedenmann, Lewis C Cantley, Joe W Gray & Douglas Hanahan



Gene Expression Classification of Colon Cancer into Molecular Subtypes: Characterization, Validation, and Prognostic Value

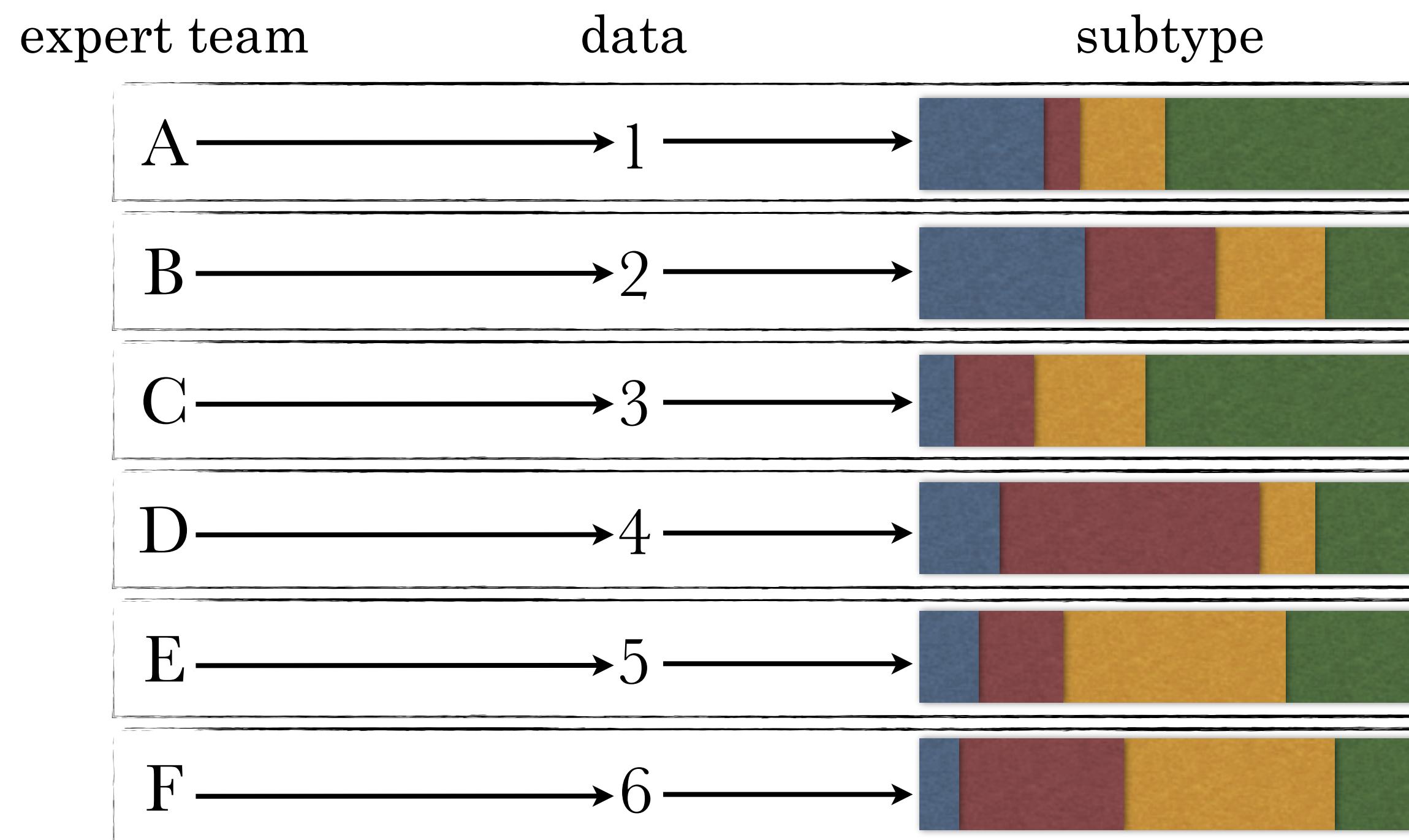
Laetitia Marisa, Aurélien de Reyniès, Alex Duval, Janick Selves, Marie Pierre Gaub, Laure Vescovo, Marie-Christine Etienne-Grimaldi, Renaud Schiappa, Dominique Guenot, Mira Ayadi, Sylvain Kirzin, Maurice Chazal, Jean-François Fléjou, Daniel Benchimol, Anne Berger, Arnaud Lagarde, Erwan Pencreach, Françoise Piard, Dominique Elias, Yann Parc, Sylviane Olschwang, Gérard Milano, Pierre Laurent-Puig , Valérie Boige [view less]



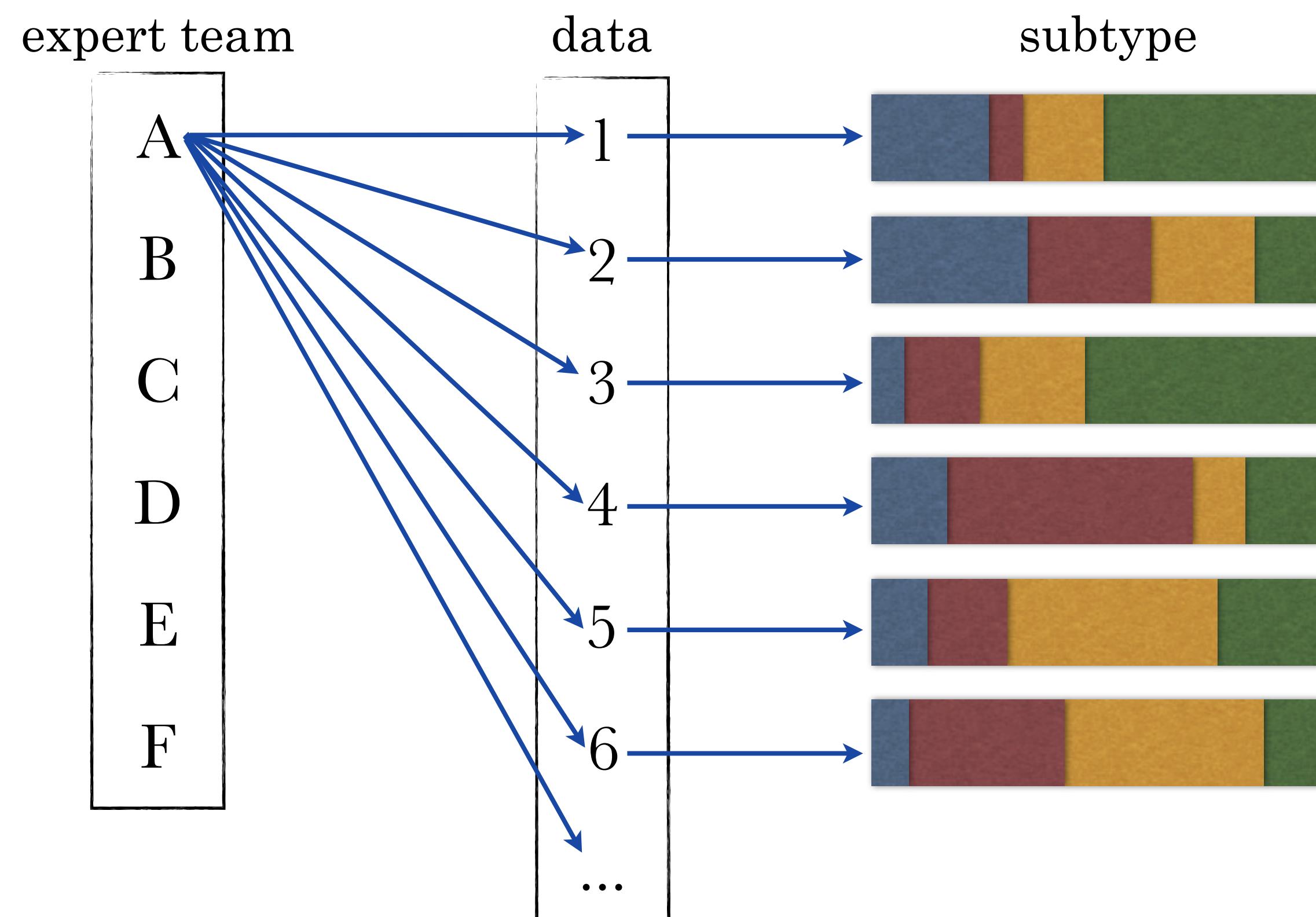
Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer

Eva Budinska,^{1,2,*} Vlad Popovici,^{1,2} Sabine Tejpar,³ Giovanni D'Ario,¹ Nicolas Lapique,¹ Katarzyna Otylia Sikora,¹ Antonio Fabio Di Narzo,¹ Pu Yan,⁴ John Graeme Hodgson,⁵ Scott Weinrich,⁵ Fred Bosman,⁵ Arnaud Roth^{6,7} and Mauro Delorenzi^{1,8}

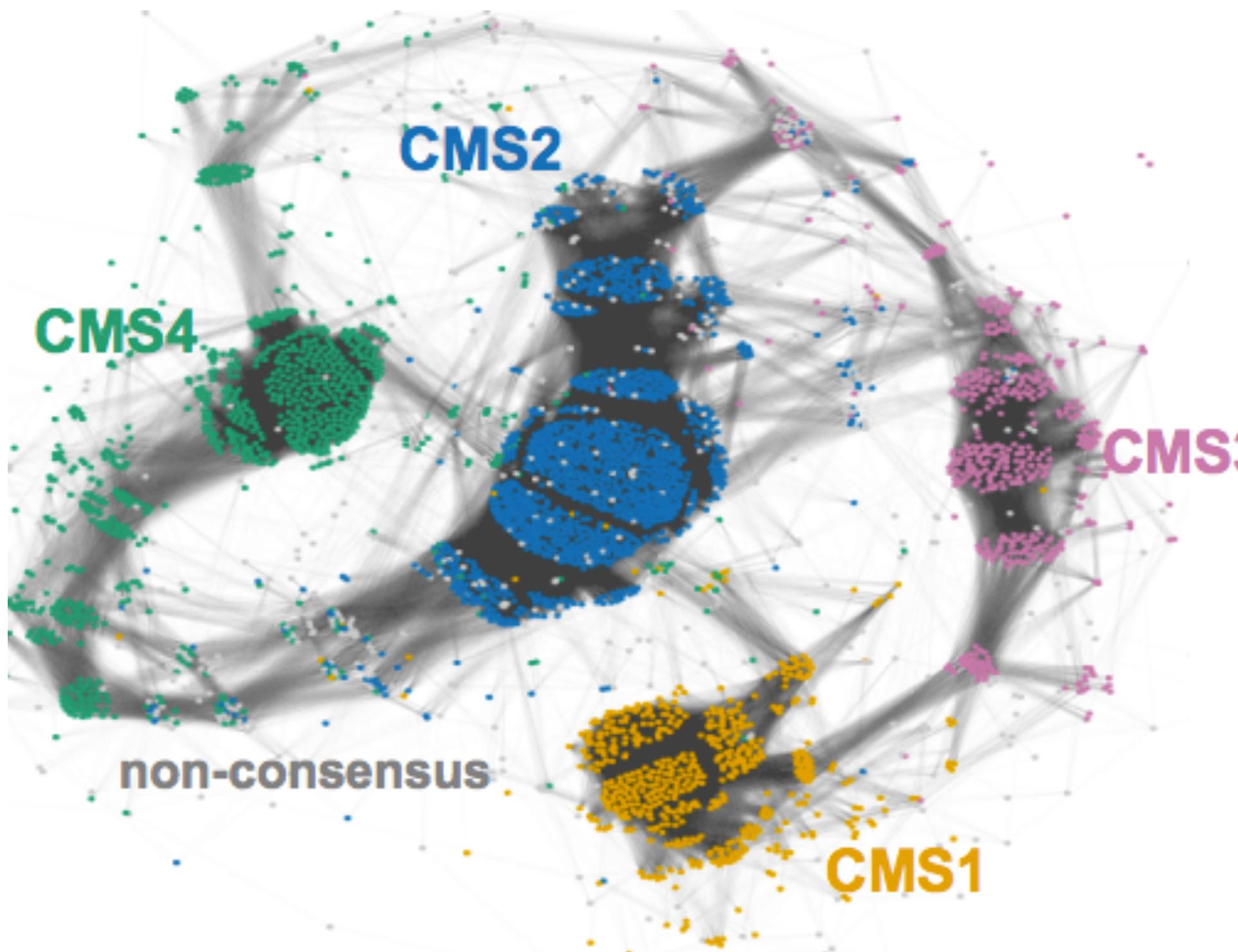
crc subtyping consortium



crc subtyping consortium



crc subtyping consortium

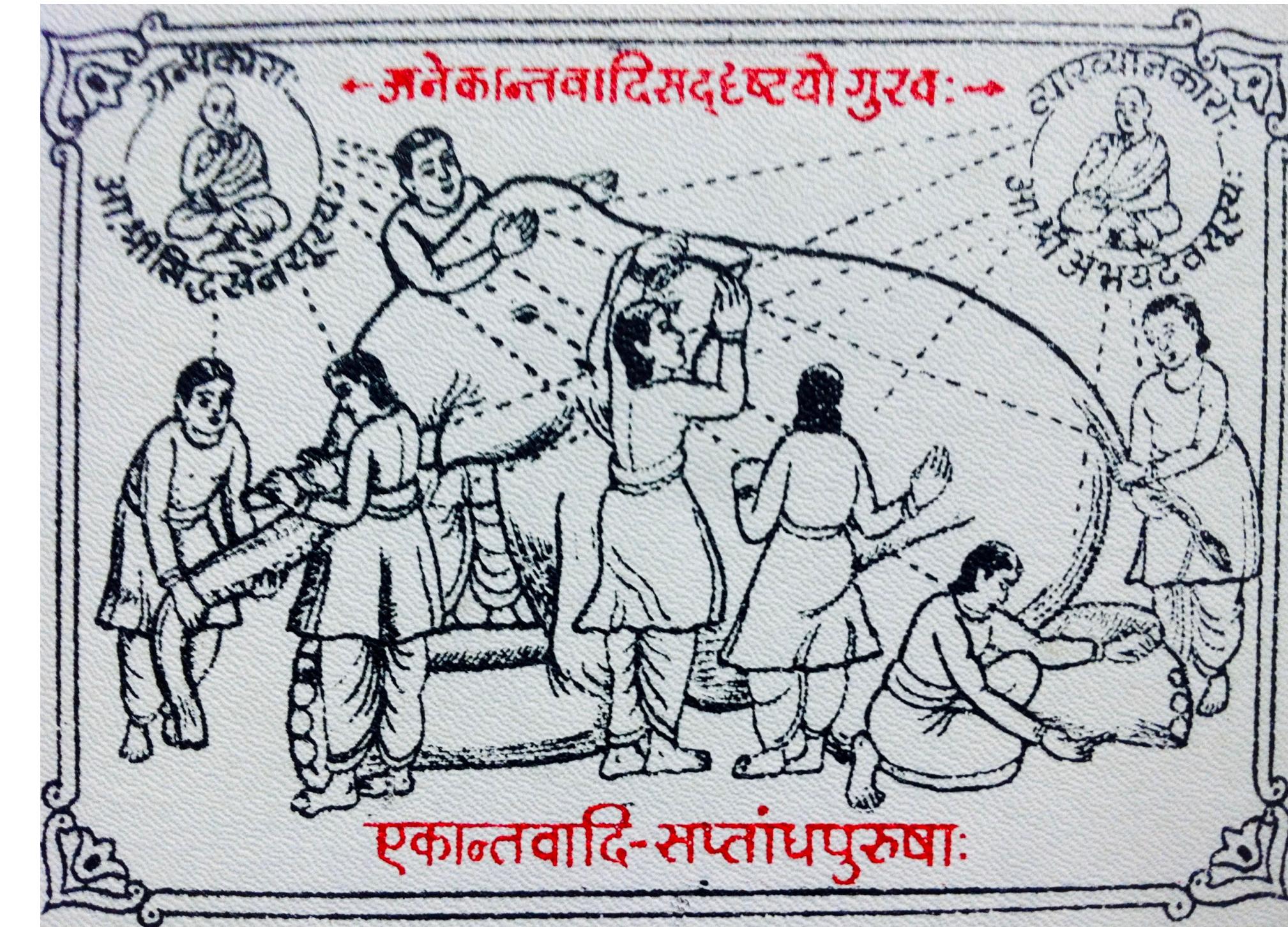


CMS1 MSI immune	CMS2 Canonical	CMS3 Metabolic	CMS4 Mesenchymal
14%	37%	13%	23%
MSI, CIMP high, hypermutation	SCNA high	Mixed MSI status, SCNA low, CIMP low	SCNA high
<i>BRAF</i> mutations		<i>KRAS</i> mutations	
Immune infiltration and activation	WNT and <i>MYC</i> activation	Metabolic deregulation	Stromal infiltration, TGF- β activation, angiogenesis
Worse survival after relapse			Worse relapse-free and overall survival
<small>CIMP, CpG island methylator phenotype; MSI, microsatellite instability; SCNA, somatic copy number alterations.</small>			

guinney et al, *nature medicine* 2015

crc subtyping consortium

parable of
the blind men
and the elephant



romana klee: cc by-sa 2.0

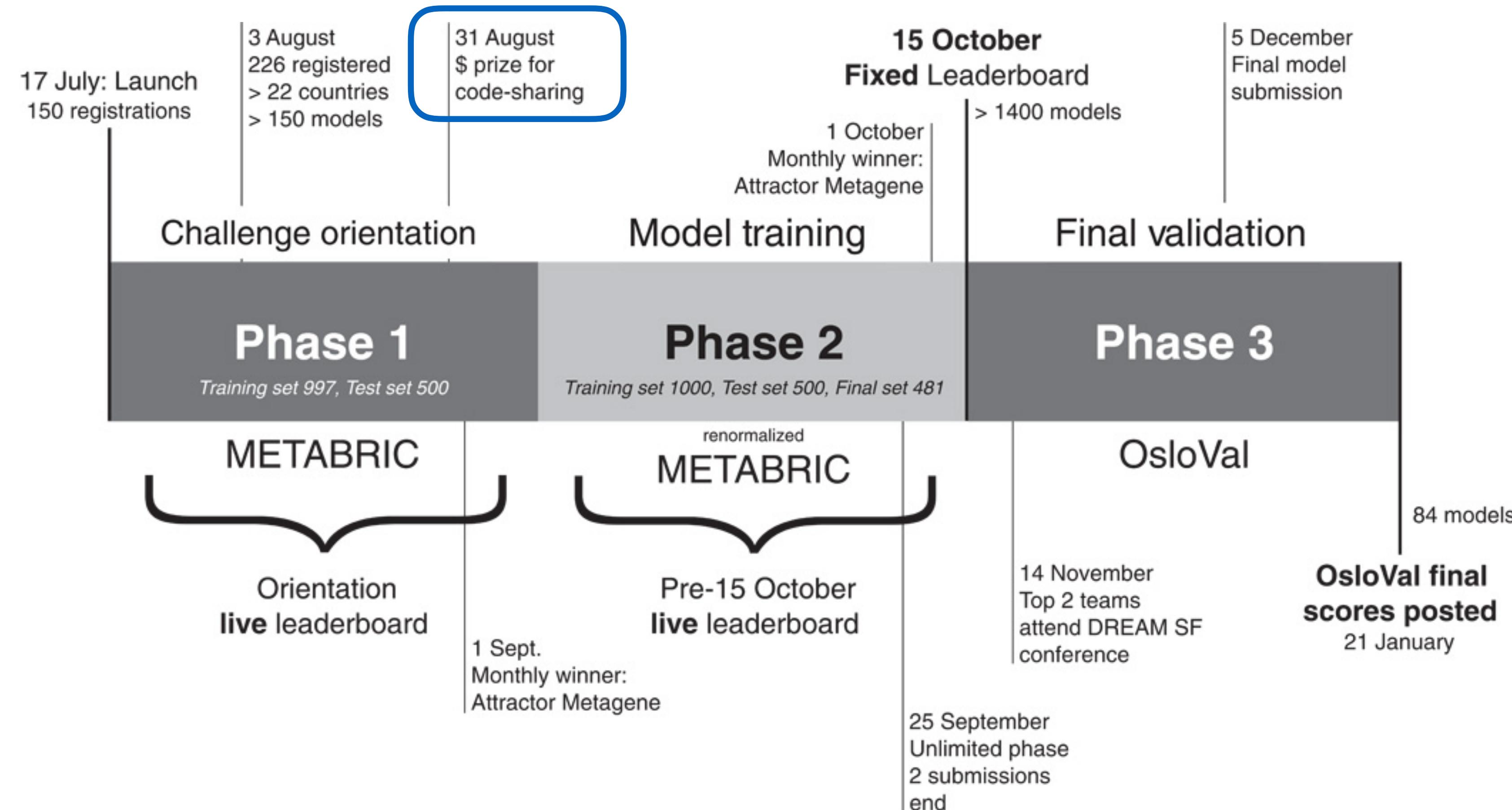
*“while one's subjective experience is true,
it may not be the totality of truth”*

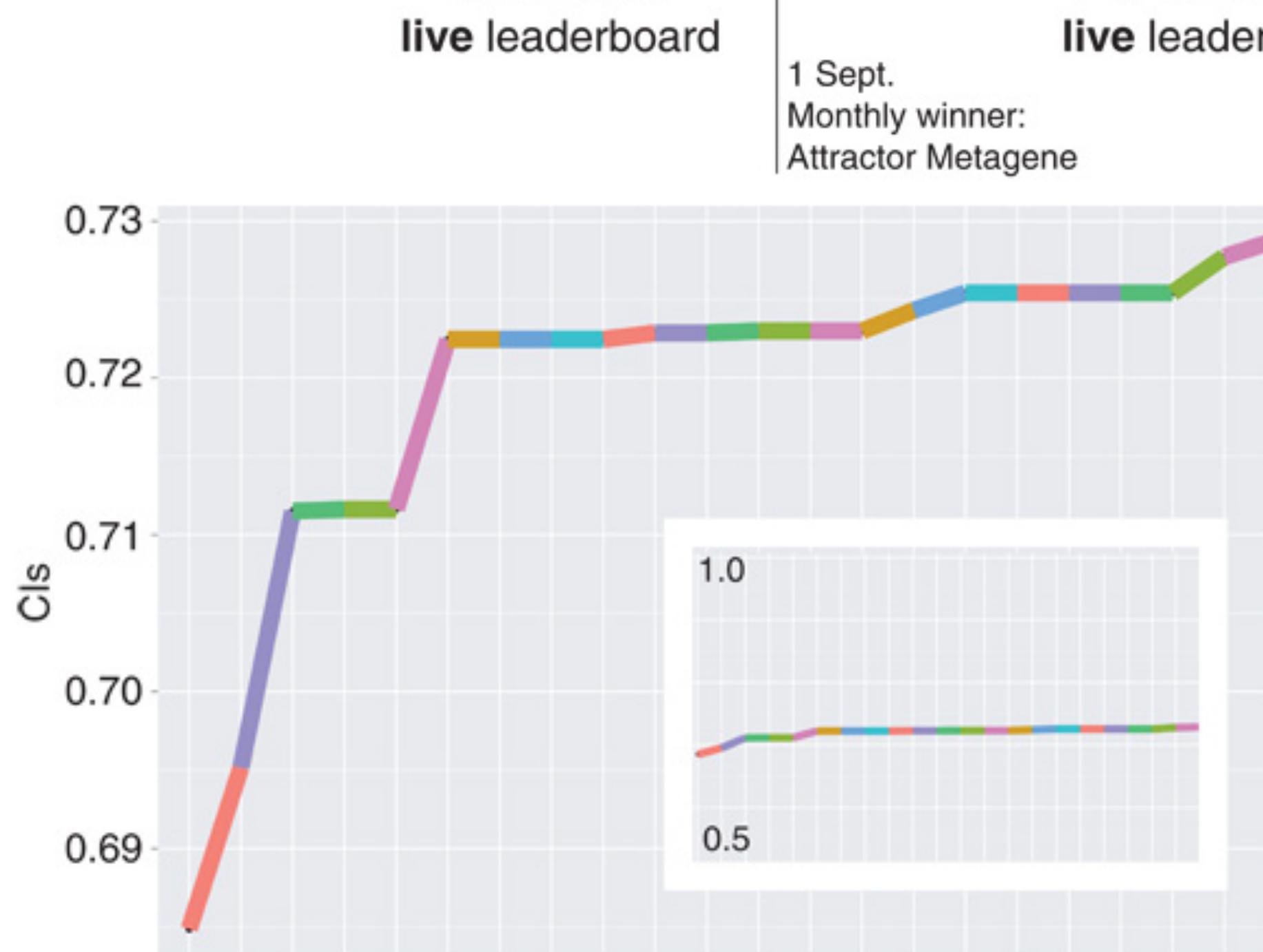
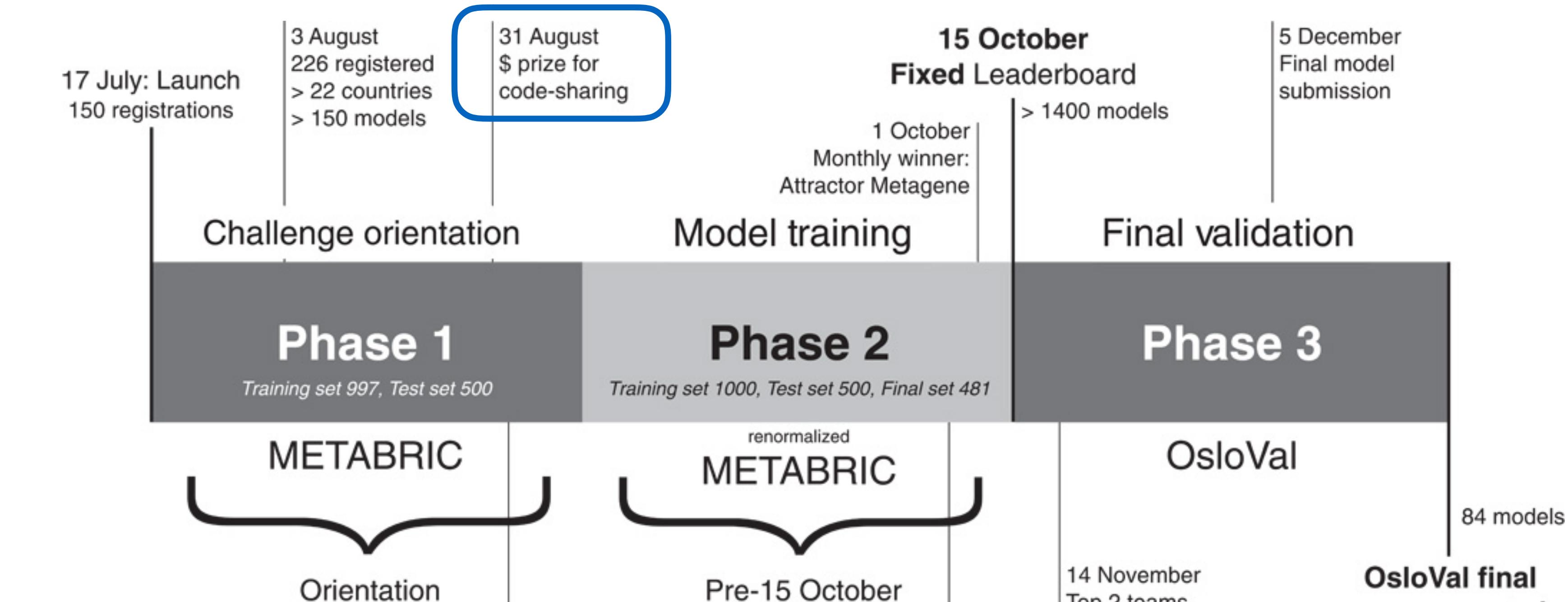
we as scientists (and citizens),
especially amidst the hype of BIG DATA and AI,

need to become comfortable with and
recognize incomplete truths

$$f(\text{imbalance}) = \text{imbalance}$$

use case 3: breast cancer prognosis challenge
(be mindful of when open is appropriate)





questions?