Big data: a new era in sleep medicine?

H. Woehrle
Sleep and Ventilation Center Blaubeuren/
Lung Center Ulm
hwoehrle@lungenzentrum-ulm.de

Disclosures

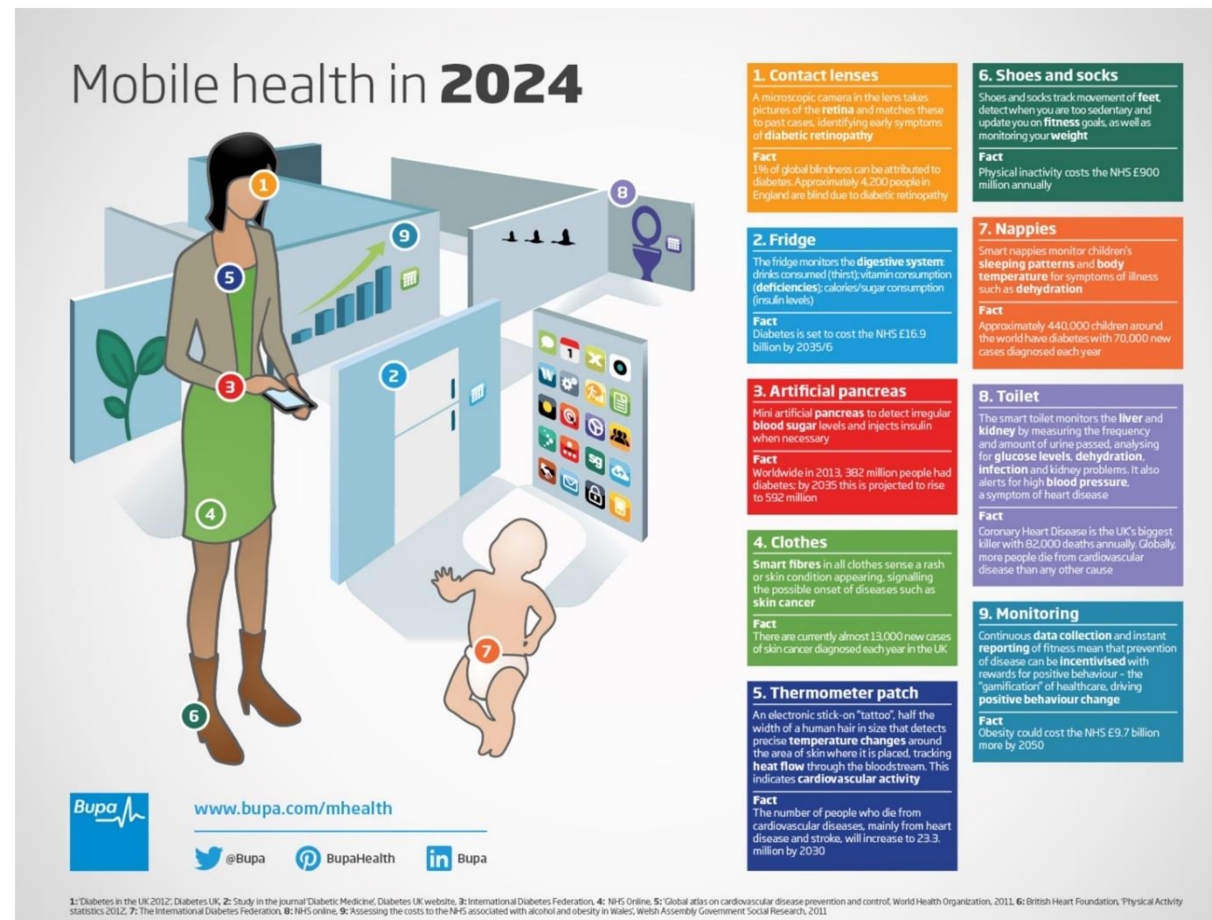
- Lecture Fees/Consulting: Astra Zeneca, Bayer, Boehringer Ingelheim, GSK, Inspire, ResMed, Vital Air, Weinmann
- Research Grants: Weinmann, MAP, ResMed
- Employment as Medical Director (SERVE-HF study management) by ResMed 2006-16

Quelle: TheAtomicMoose.ca

Digitialization = Data storage and analysis

Quellen: duden.de; deacademic.com

mHealth



Quellen: redmondpie.com; mhealthwatch.com

Big Data

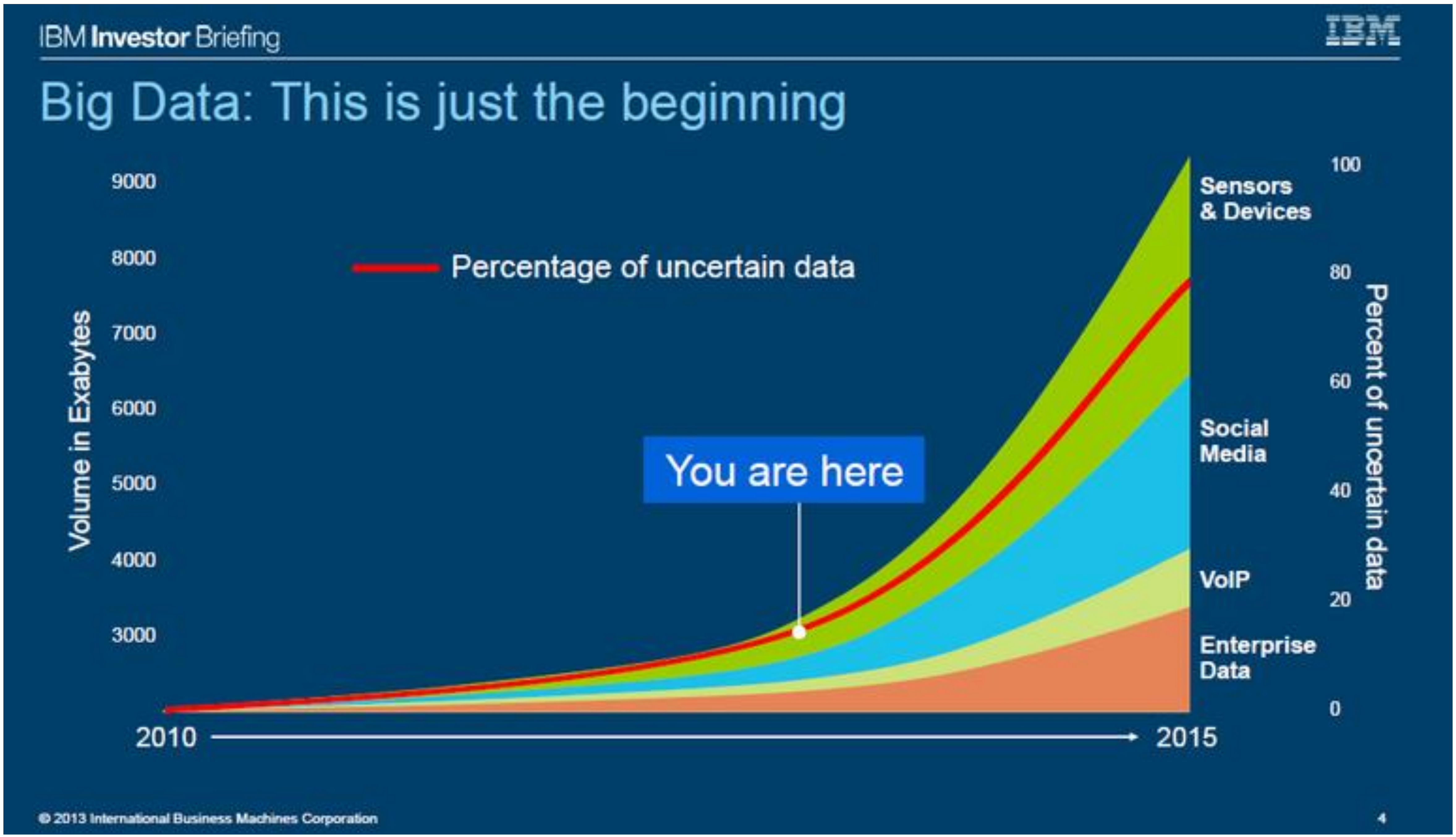
Big data is data sets that are so **voluminous** and **complex** that traditional data processing application software are inadequate to deal with them.

Wikipedia
The Free Encyclopedia

The term "big data" refers not only to large data sets, but also to the frameworks, techniques, and tools used to analyze it. It can be collected through any data-generating process such as social media, public utility infrastructure, and search engines. Big data may be either semi-structured, structured, or unstructured.

CIKLUM
EMPOWERING COLLABORATION

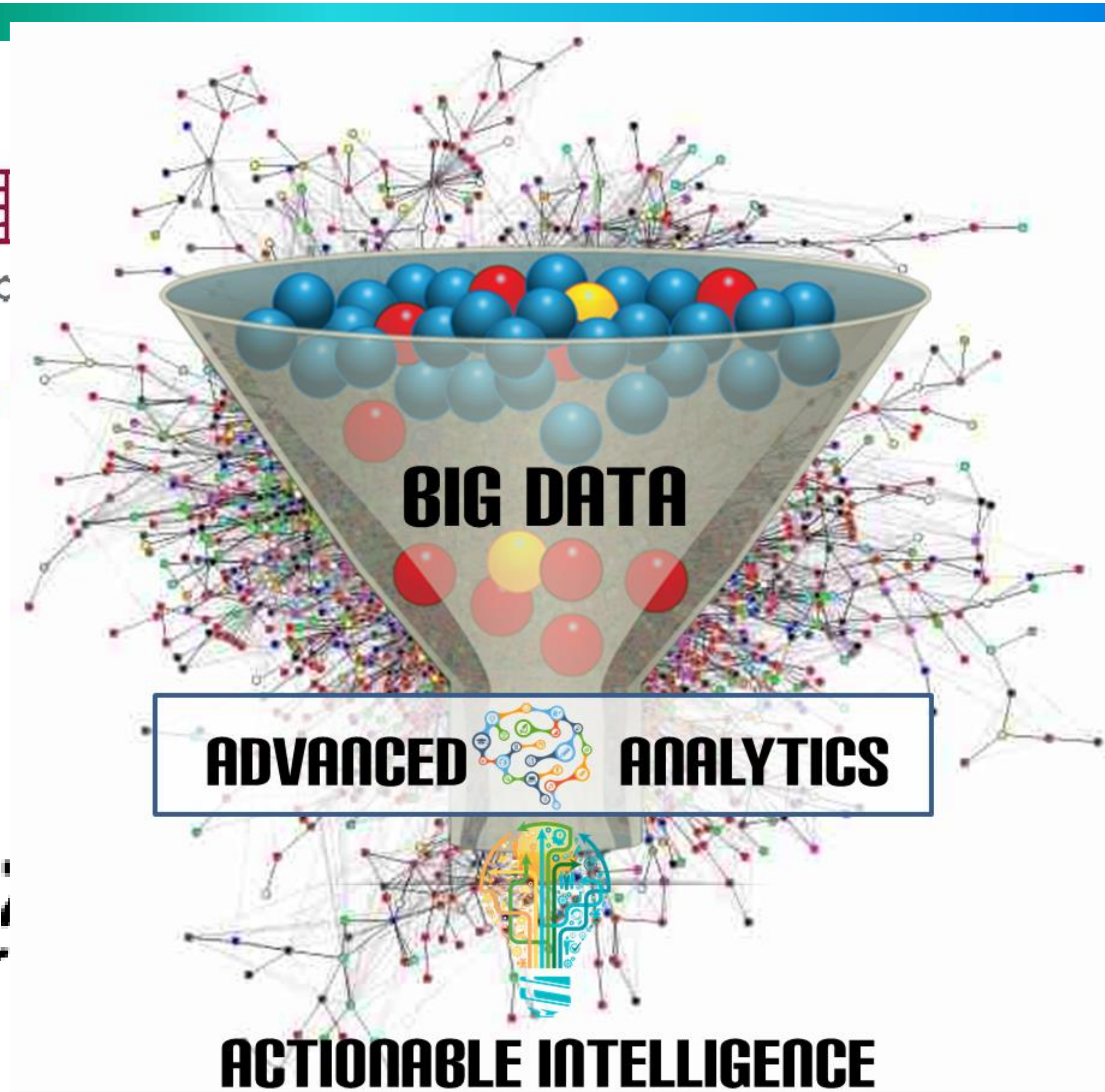
Big Data growing fast



The Five V‘s of Big Data: 4 + value

https://www.mosaiq.com; http://onlinembapage.com/data-analytics-mba/; http://www.phonecruncher.com; resmed.com

Big Data



Source: blog.ness-ses.com; http://www.enterrasolutions.com

Validity of Machine Learning

https://www.reddit.com; https://twitter.com/teenybiscuit/status/707727863571582978/photo/1

Big Data

Pros

- Insights (e.g., phenotyping, precision medicine)
- Safety (long-term, real time; interactions)
- Trends, trajectories
- Monitoring

Cons

- Inherent biases in how all data are collected and interpreted.
- Objections to the idea that data mining can replace hypothesis driven theory by content experts.
- The observation that the larger the data set the more likely that spurious correlations that are not useful will be identified

Bottles et al; PEJ JULY•AUGUST/2014

Big data – data ownership?

Who Owns the Data? Open Data for Healthcare

Patty Kostkova[1], Helen Brewer[2], Simon de Lusignan[3], Edward Fottrell[4], Ben Goldacre[5], Graham Hart[4], Phil Koczan[6], Peter Knight[7], Corinne Marsolier[8], Rachel A. McKendry[9], Emma Ross[10], Angela Sasse[1], Ralph Sullivan[11], Sarah Chaytor[12], Olivia Stevenson[12], Raquel Velho[13] and John Tooke[14]*

frontiers in Public Health
February 2016 | Volume 4 | Article 7

By way of contrast, through increasing popularity of social media, GPS-enabled mobile apps and tracking/wearable devices, the IT industry and MedTech giants are pursuing new projects without clear public and policy discussion about ownership and responsibility for user-generated data.

In lights of the risk-adverse pro-business policy making attitudes in this domain, it is the golden opportunity and professional responsibility of the research community to challenge policymakers and regulatory bodies authorities and actively lead on the complex multi-stakeholder processes of establishing this new agenda.

The NEW ENGLAND JOURNAL of MEDICINE

OCCASIONAL NOTES

Chocolate Consumption, Traffic Accidents and Serial Killers

James R. Winters and Seán G. Roberts

School of Philosophy, Psychology & Language Sciences,
University of Edinburgh,
Dugald Stewart Building,
3 Charles Street,
Edinburgh,
EH8 9AD

J.R.Winters@sms.ed.ac.uk

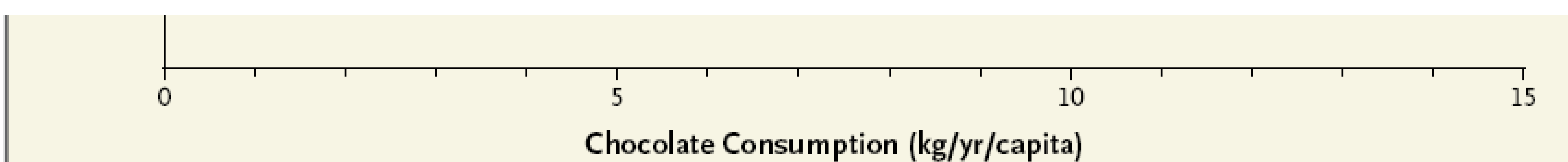
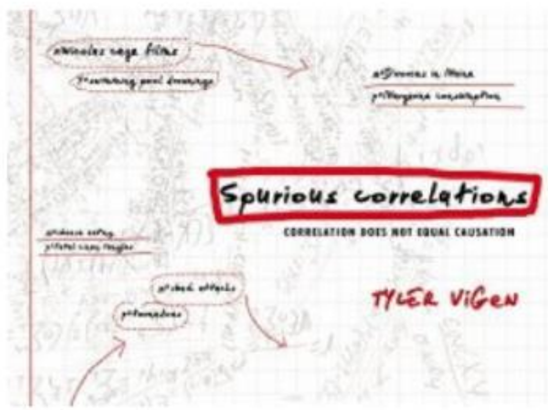


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

Big data – spurious correlations

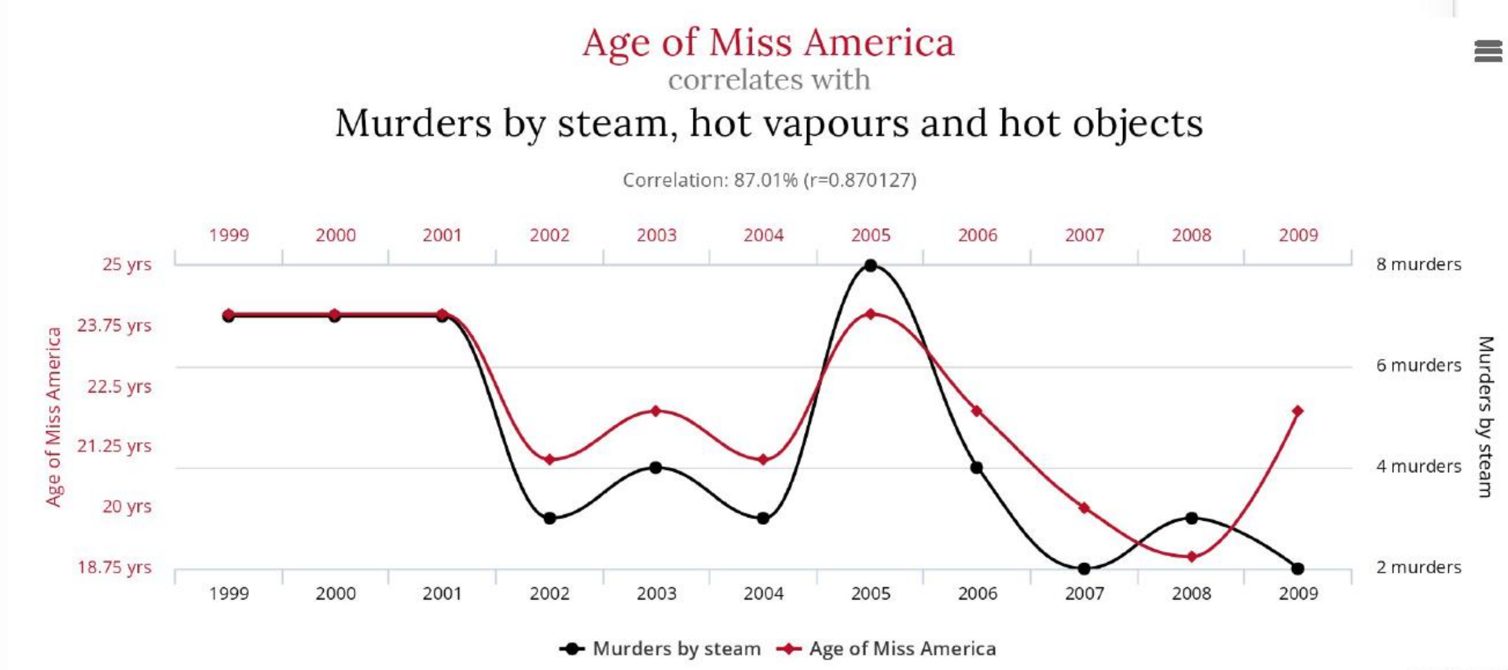
Spurious correlations



Now a ridiculous book!

- Spurious charts
- Fascinating factoids
- Commentary in the footnotes

Amazon | Barnes & Noble | Indie Bound



Data sources: Wikipedia and Centers for Disease Control & Prevention

Machine Learning in Medicine

LETTER RESEARCH

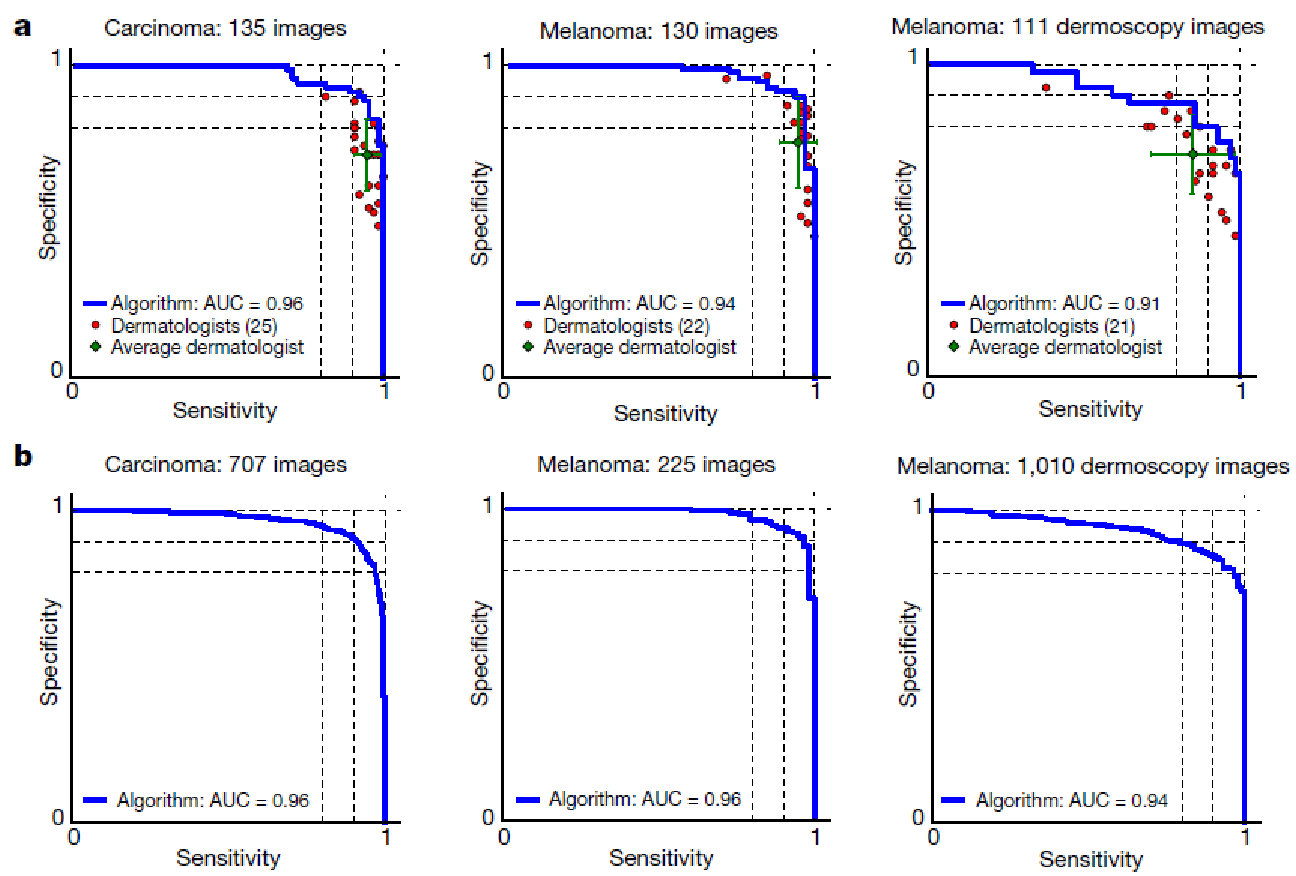


Figure 3 | Skin cancer classification performance of the CNN and dermatologists. a, The deep learning CNN outperforms the average of the dermatologists at skin cancer classification using photographic and dermoscopic images. Our CNN is tested against at least 21 dermatologists at keratinocyte carcinoma and melanoma recognition. For each test, previously unseen, biopsy-proven images of lesions are displayed, and dermatologists are asked if they would: biopsy/treat the lesion or reassure the patient. Sensitivity, the true positive rate, and specificity, the true negative rate, measure performance. A dermatologist outputs a single prediction per image and is thus represented by a single red point. The green points are the average of the dermatologists for each task, with error bars denoting one standard deviation (calculated from $n = 25$, 22 and 21 tested dermatologists for keratinocyte carcinoma, melanoma and melanoma under dermoscopy, respectively). The CNN outputs a malignancy probability P per image. We fix a threshold probability t such that the prediction $\hat{y}$ for any image is $\hat{y} = P \geq t$, and the blue curve is drawn by sweeping t in the interval 0–1. The AUC is the CNN's measure of performance, with a maximum value of 1. The CNN achieves superior performance to a dermatologist if the sensitivity–specificity point of the dermatologist lies below the blue curve, which most do. Epidermal test: 65 keratinocyte carcinomas and 70 benign seborrheic keratoses. Melanocytic test: 33 malignant melanomas and 97 benign nevi. A second melanocytic test using dermoscopic images is displayed for comparison: 71 malignant and 40 benign. The slight performance decrease reflects differences in the difficulty of the images tested rather than the diagnostic accuracies of visual versus dermoscopic examination. **b**, The deep learning CNN exhibits reliable cancer classification when tested on a larger dataset. We tested the CNN on more images to demonstrate robust and reliable cancer classification. The CNN's curves are smoother owing to the larger test set.

Big Data and Medicine

TECH TWITTER

How Twitter Knows When You're Depressed

Sam Frizell @Sam_Frizell | Jan. 27, 2014

A team of researchers have developed a way to scan your tweets and determine whether or not you're depressed with a claimed accuracy rate of 70%, though the scientists admit their model is a long way from perfect

David Paul Morris—Bloomberg/Getty Images

With its 230 million regular users, Twitter has become such a broad stream of personal expression that researchers are beginning to use it as a tool to dig into public health problems. Believe it or not, a scientist out there might actually *care* about the sandwich you ate for lunch—even if most of your followers don't.

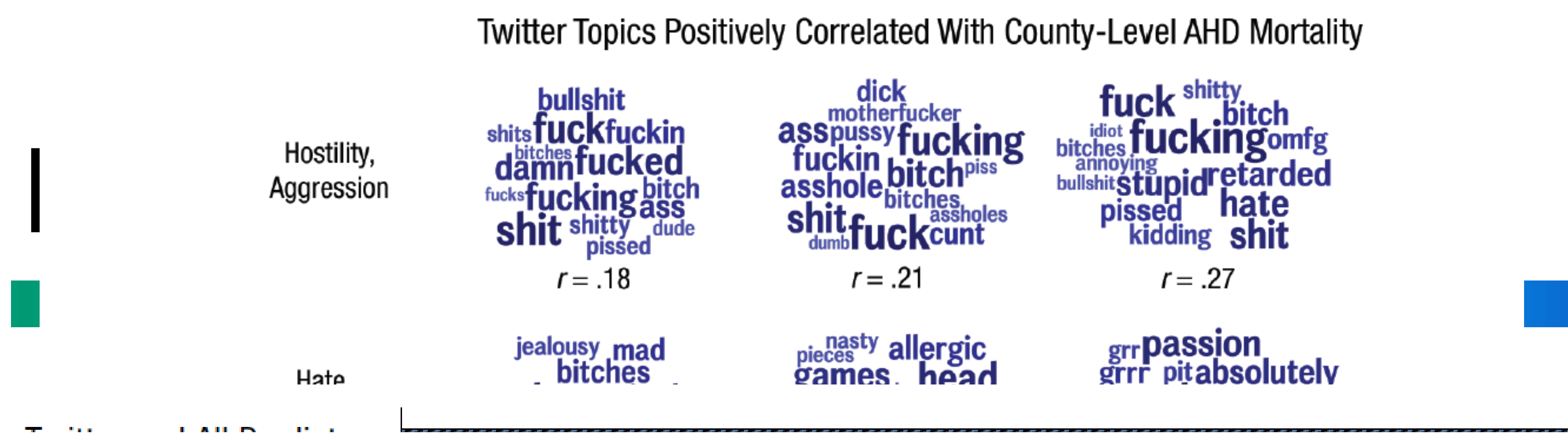
"Our attitude is that Twitter is the largest observation ever known, and we're working very hard to take adv McCormick of the Center for Statistics and the Social Sciences at the University of Washington.

What if, for example, an artificial intelligence model could scan your Twitter feed and tell you if you're at risk for depression? And what if you could receive notices from third parties, for instance, that warned you that you may want to seek help, just based on an automated scan of your tweets? Eric Horvitz, co-director of Microsoft Research Redmond has helped pioneer research on Twitter and depression. He says that could one day be a possibility.

"We wondered if we could actually build measures that might be able to detect if someone is severely depressed, just in publicly posted media. What are people telling the world in public spaces?" asks Horvitz. "You might imagine tools that could make people aware of a swing in mood, even before they can feel it themselves."

Horvitz and a team of researchers helped develop a model that can scan tweets and predict depression in Twitter users, with an accuracy they claim to be 70%. Researchers say the system is still far from perfect. When the model scans your tweets, it misses some signals and doesn't diagnose many people—about 30%—who really will get depression. And the system has a "false positive" issue, Horvitz said, causing it to incorrectly predict that healthy Twitter users will get depression in about 10% of cases.

The Microsoft team found 476 Twitter users, 171 of whom were seriously depressed. They went back into users' Twitter histories as far as a year in advance of their depression



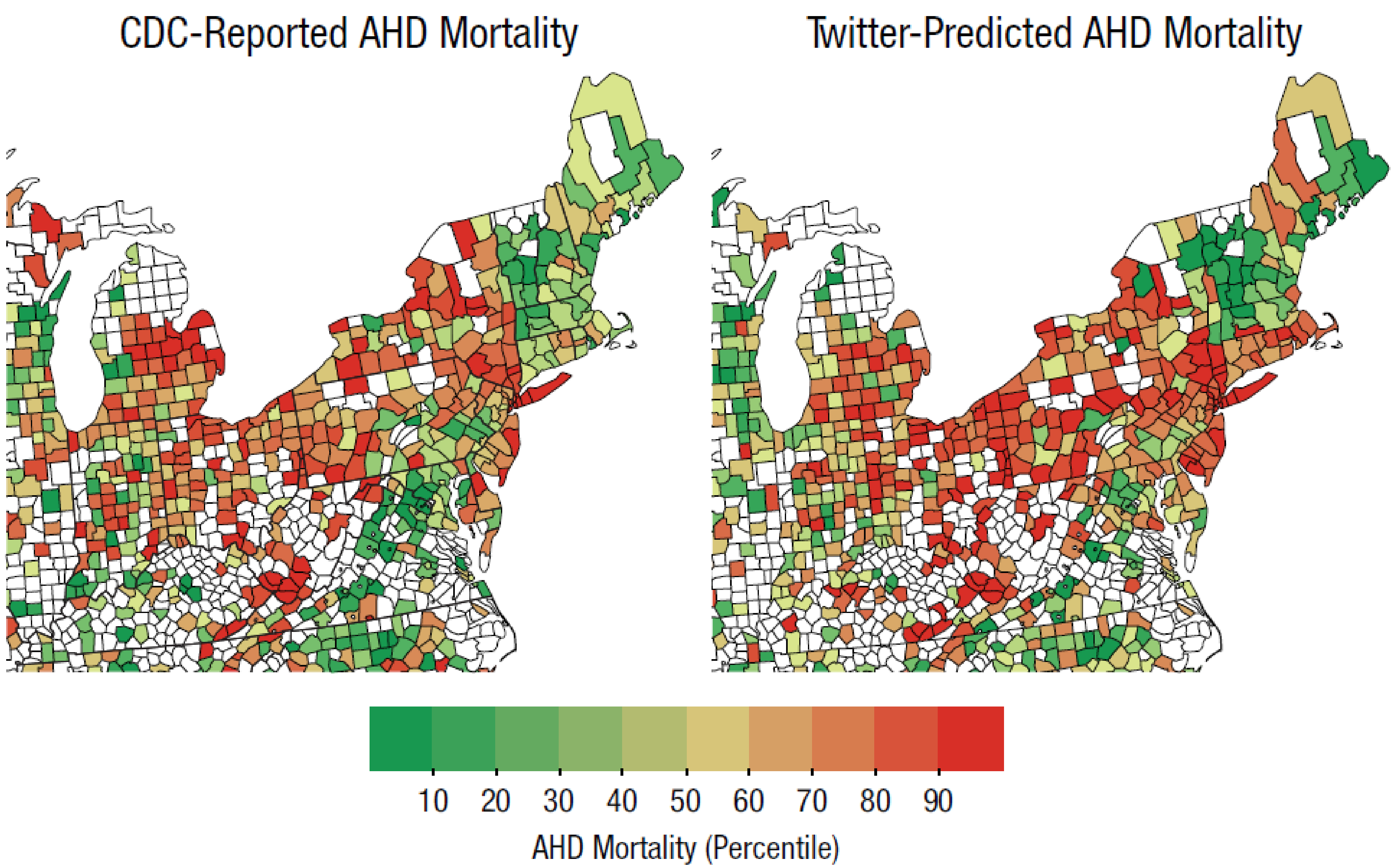


Fig. 3. Map of counties in the northeastern United States showing age-adjusted mortality from atherosclerotic heart disease (AHD) as reported by the Centers for Disease Control and Prevention (CDC; left) and as estimated through the Twitter-language-only prediction model (right). The out-of-sample predictions shown were obtained from the cross-validation process described in the text. Counties for which reliable CDC or Twitter language data were unavailable are shown in white.

topics with negative correlations (bottom) have each been grouped into sets, which are labeled at the left. The size of the word represents its prevalence relative to all words within a given topic (larger = more frequent; for details, see the Supplemental Method).

Telehealth in sleep apnea management

Journal of
Clinical Medicine

MDPI

Review

Technical Developments and Clinical Use of Telemedicine in Sleep Medicine

Marie Bruyneel

Chest Service, Saint-Pierre University Hospital, B-1000 Brussels, Belgium;
Marie_Bruyneel@stpierre-bru.be; Tel.: +32-2-535-4219; Fax: +32-2-535-4174

Academic Editors: Yolanda Blanco, Núria Solà-Valls, Rajender Gattu and Richard Lichenstein
Received: 23 October 2016; Accepted: 6 December 2016; Published: 13 December 2016

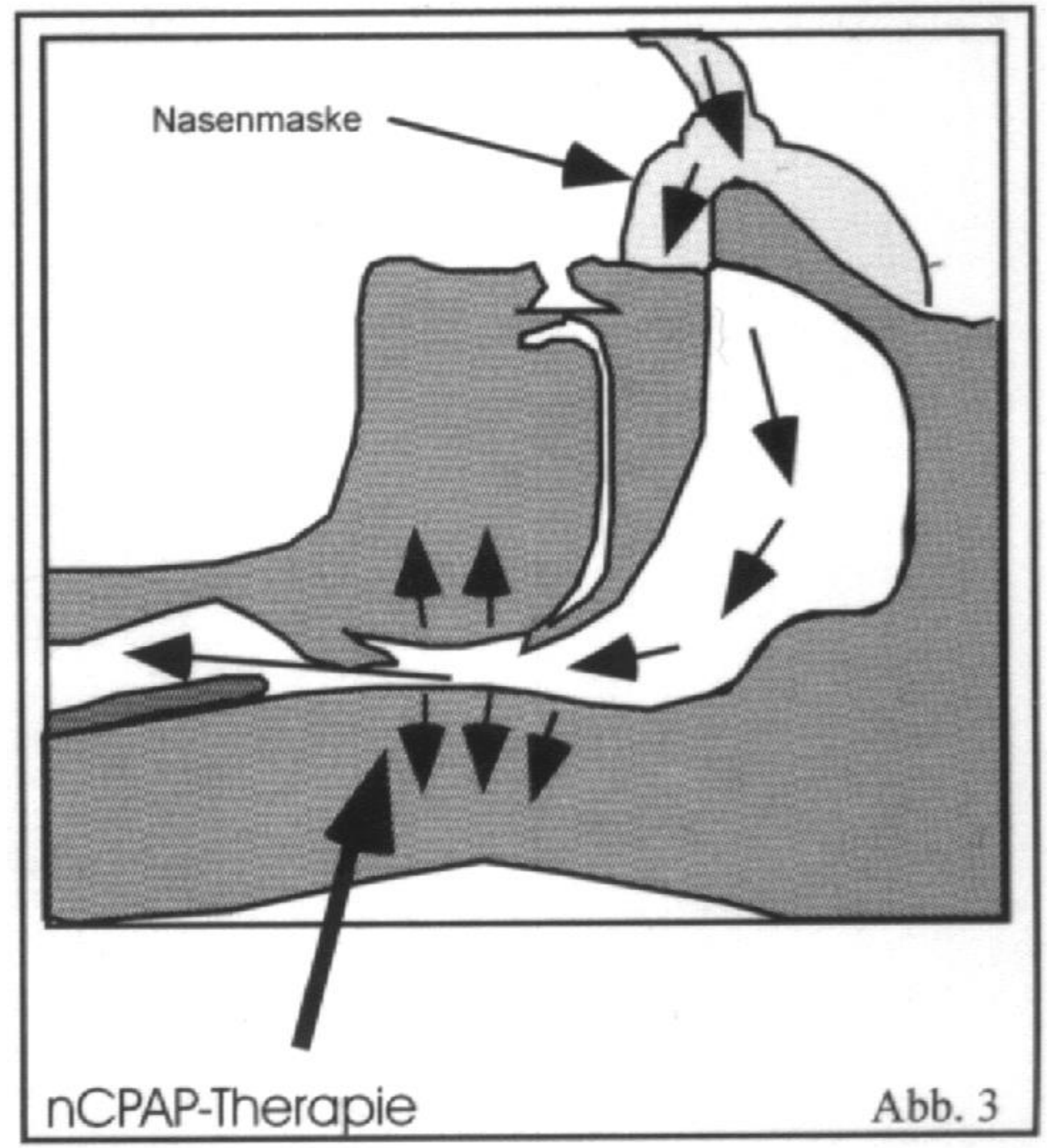
1. Telediagnostics
2. Teleconsultation
3. Teletherapy
4. Telemonitoring

Academic industrial collaboration on big data

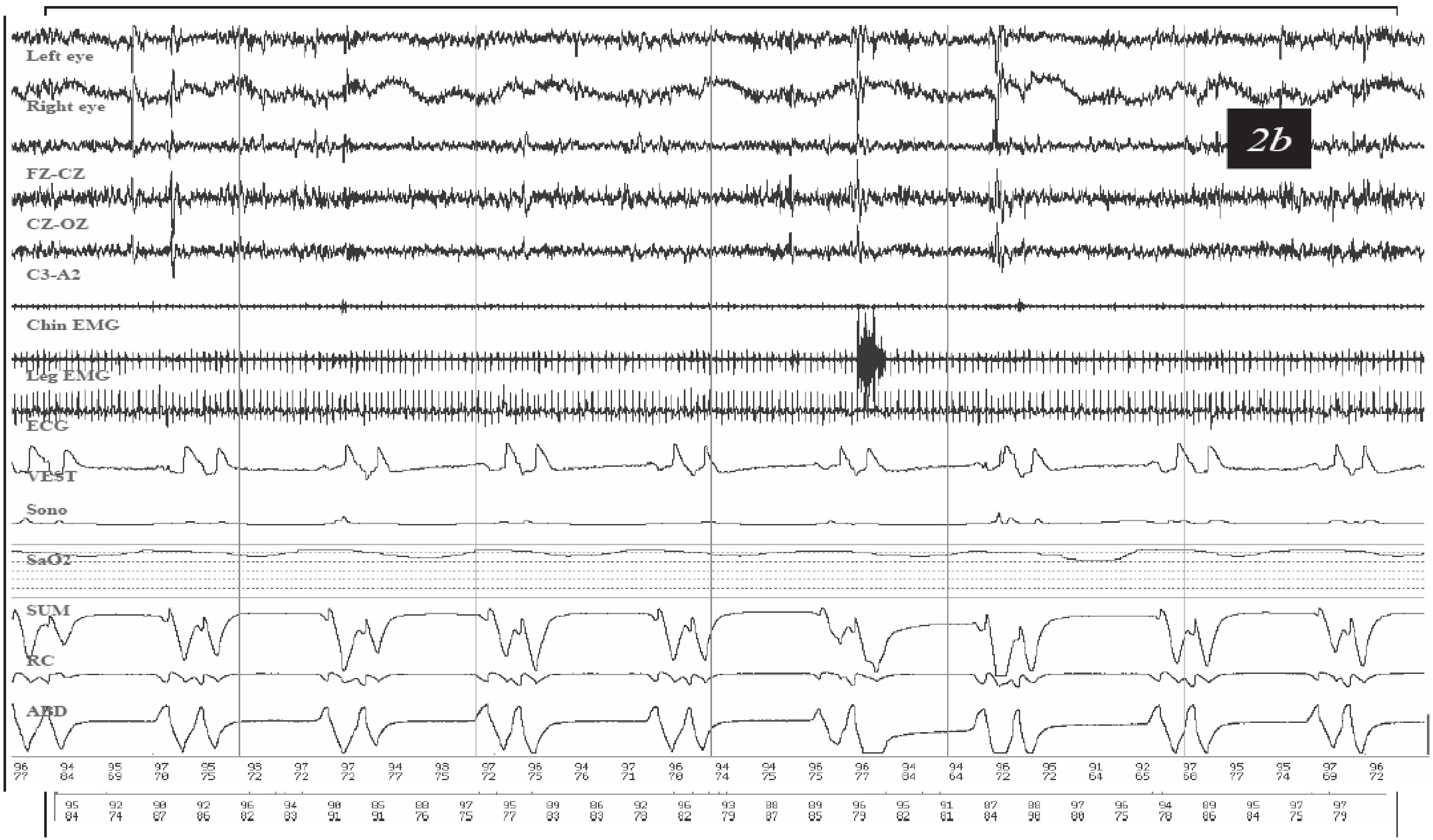
CPAP-Therapy

VORTEX BLOWER and SOFT MASK

(Sullivan) 1981 - 87



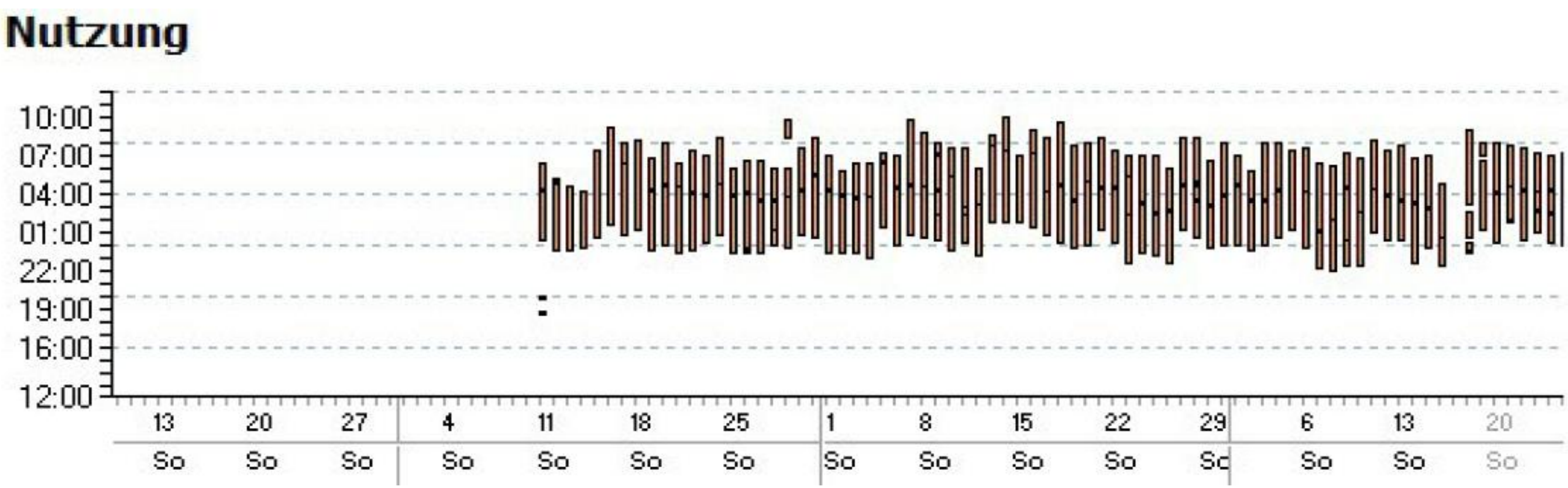
Emerging central sleep apnea during CPAP

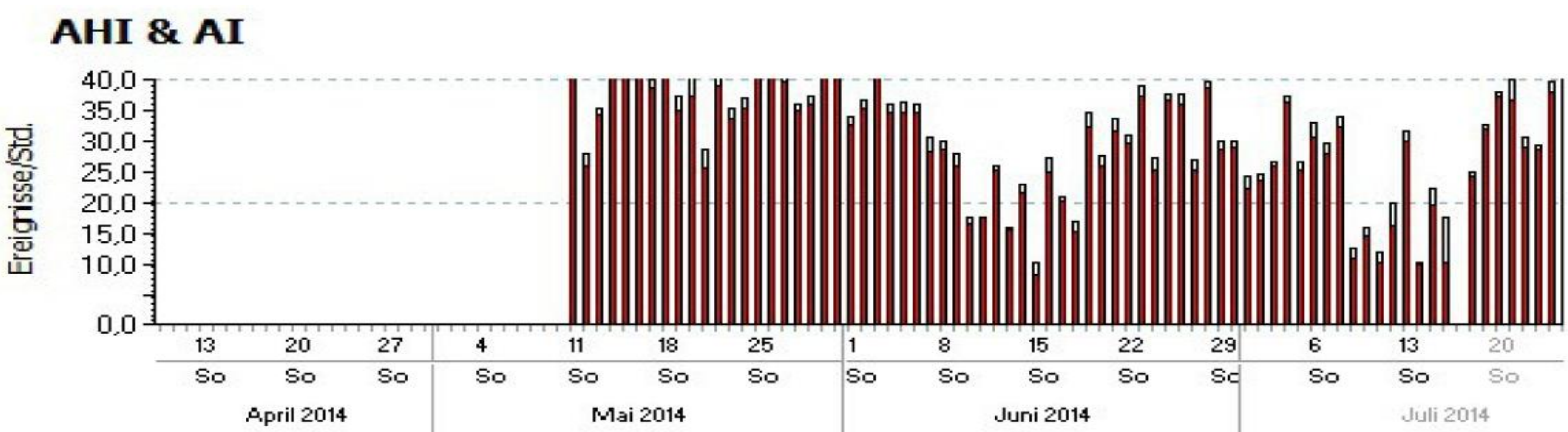


CPAP mit p=8cm H20, cAI=28/h

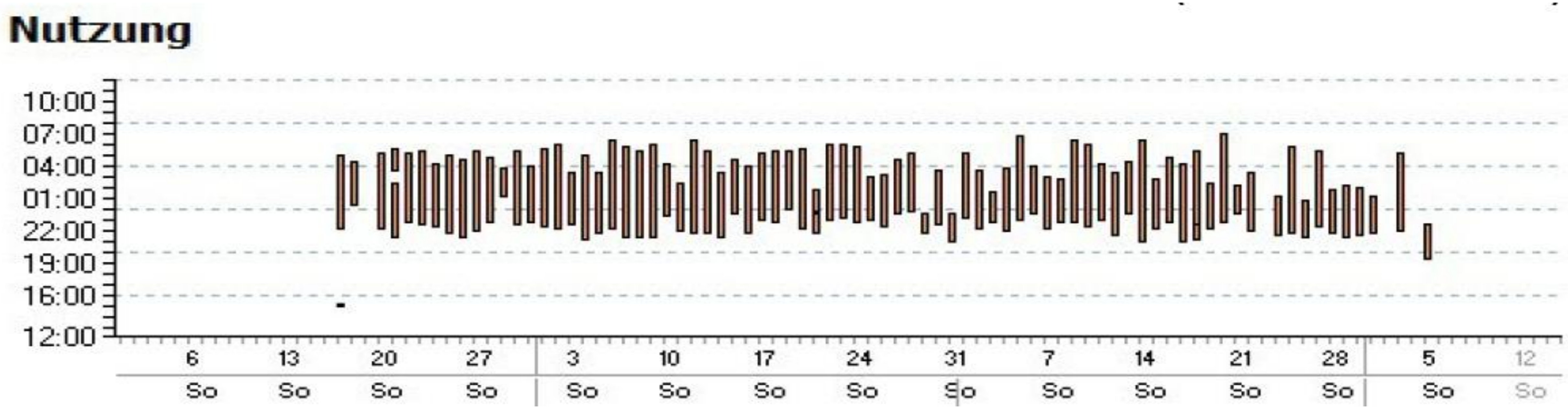
Morgenthaler, Sleep 2006

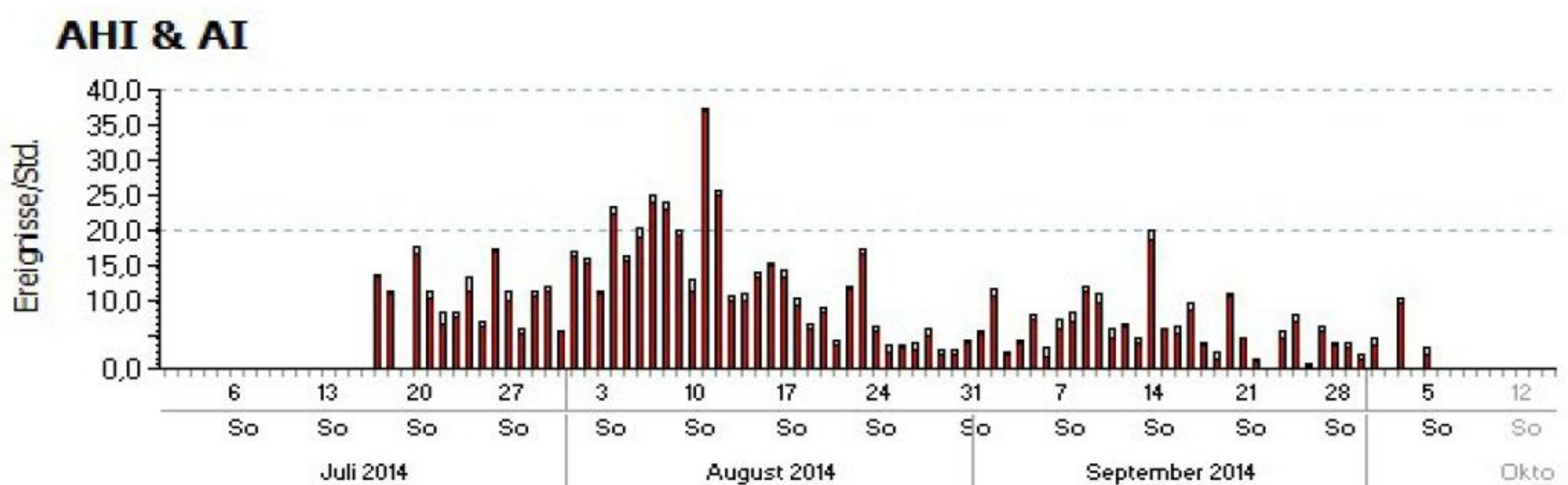
Case study – CSA persistent



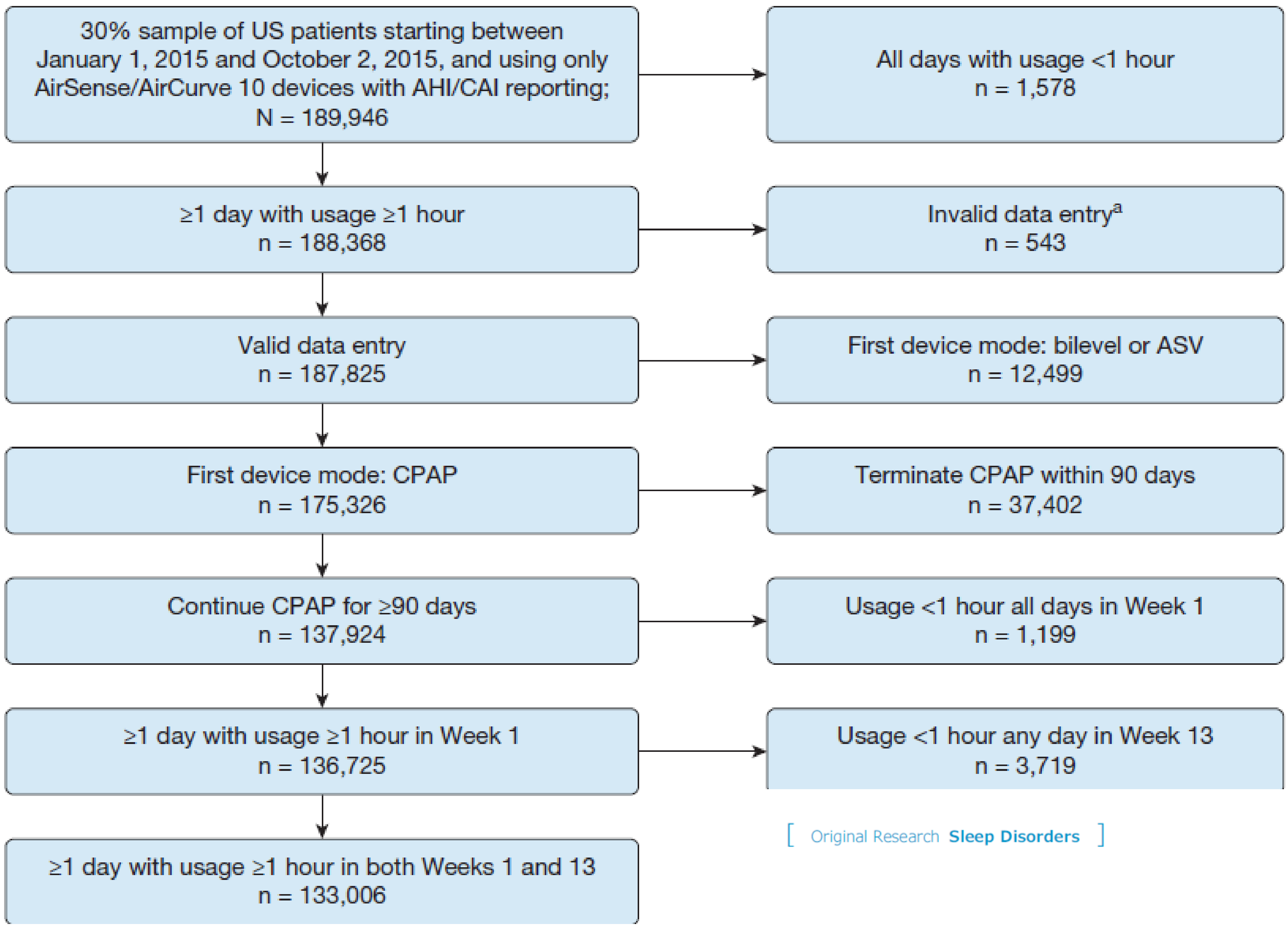


Case study – CSA transient





Patient selection – US sample, IRB waiver



[Original Research **Sleep Disorders**] CHEST

Trajectories of Emergent Central Sleep Apnea During CPAP Therapy

Dongquan Liu, PhD; Jeff Armitstead, PhD; Adam Benjafield, PhD; Shiyun Shao, PhD; Atul Malhotra, MD; Peter A. Cistulli, MD, PhD; Jean-Louis Pepin, MD, PhD; and Holger Woehrle, MD

Trajectories of CSA

[Original Research **Sleep Disorders**]

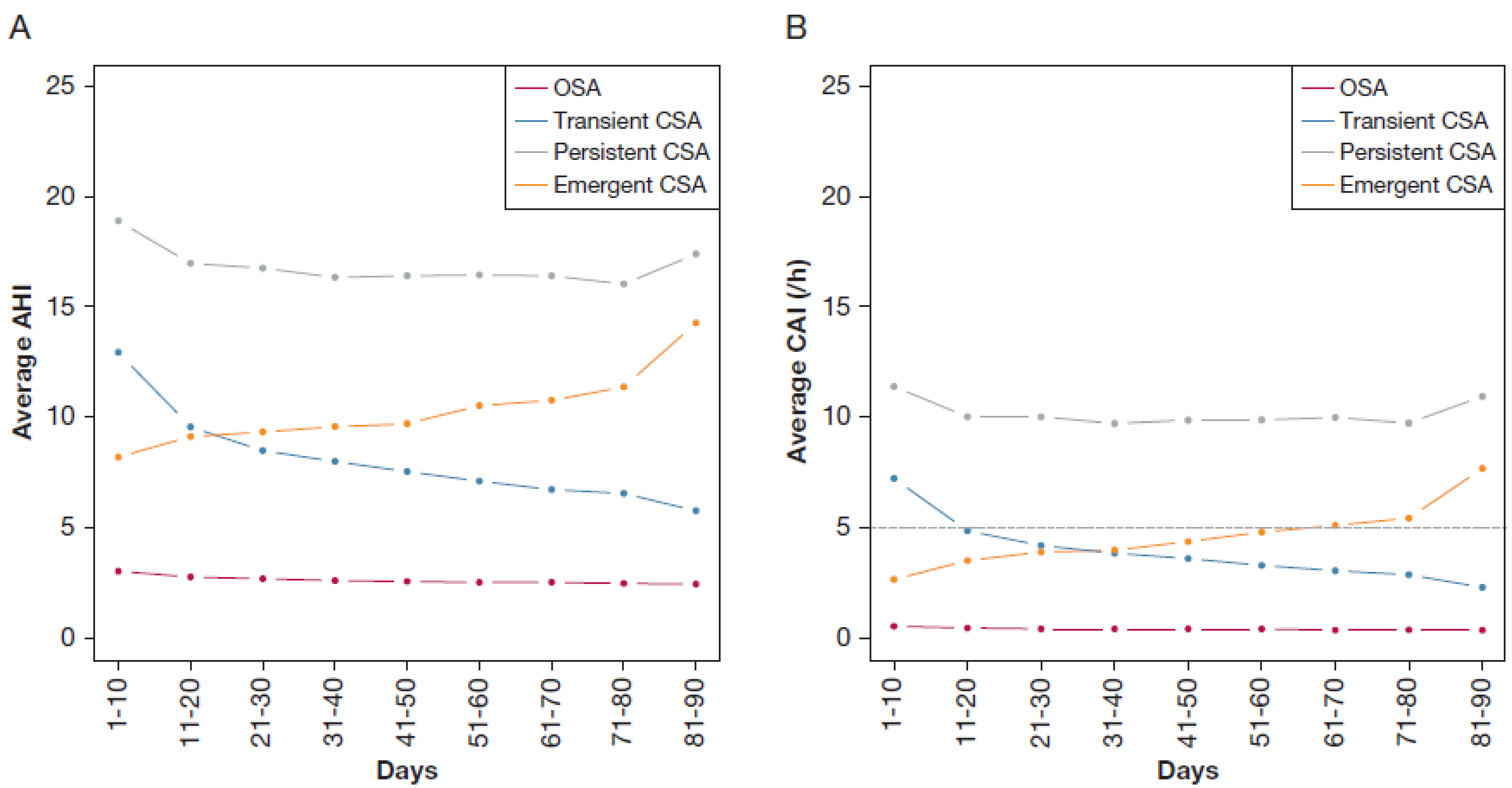
CHEST

Trajectories of Emergent Central Sleep Apnea During CPAP Therapy

CrossMark

Dongquan Liu, PhD; Jeff Armitstead, PhD; Adam Benjafield, PhD; Shiyun Shao, PhD; Atul Malhotra, MD; Peter A. Cistulli, MD, PhD; Jean-Louis Pepin, MD, PhD; and Holger Woehrle, MD

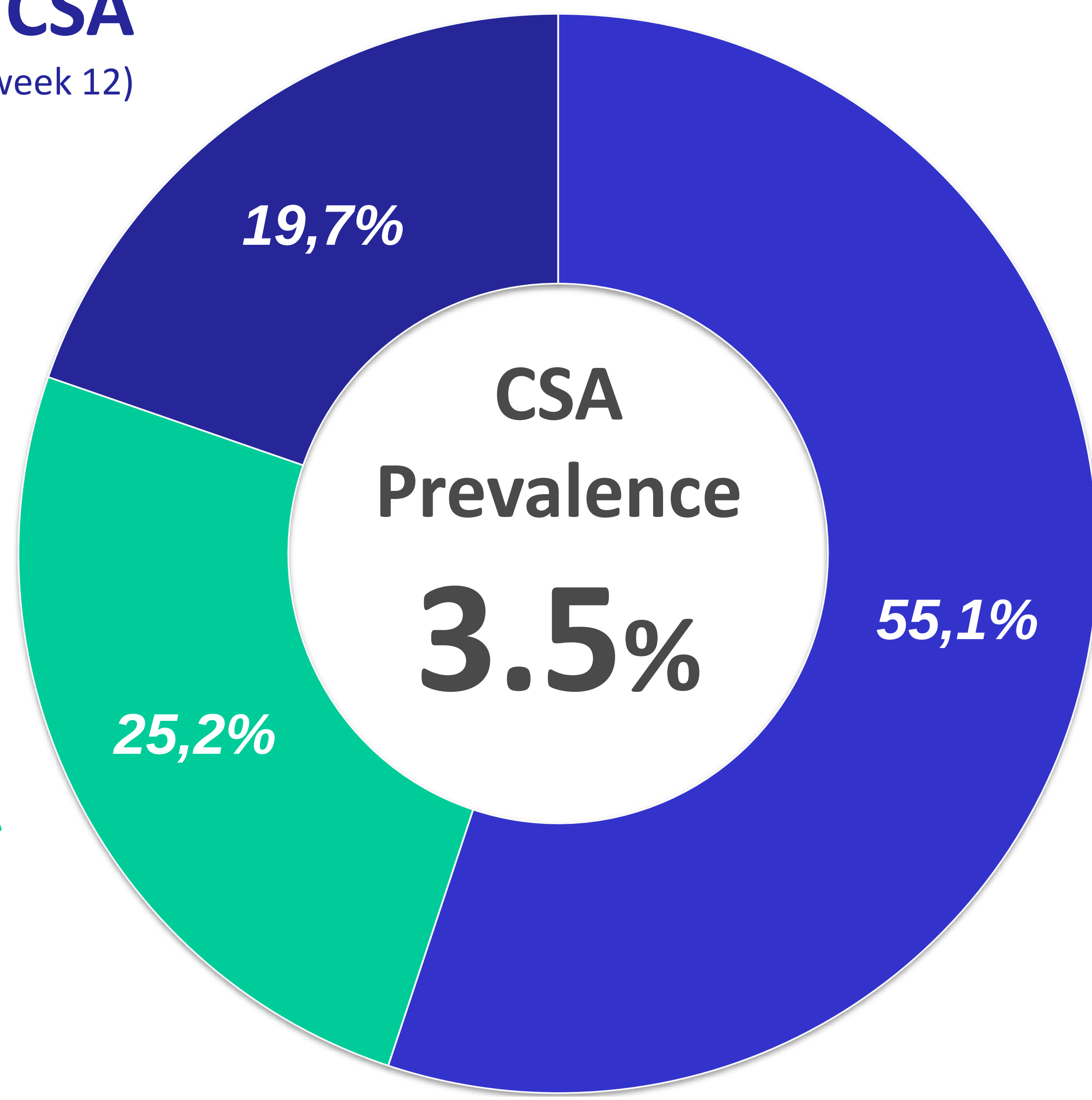
n=133.006



"CSA Trajectories"

Emergent CSA
(CAI <5/h week 1, ≥5/h week 12)

n = 133,006
90 days therapy data



Persistent CSA
(CAI ≥ 5/h week 1, ≥5/h week 12)

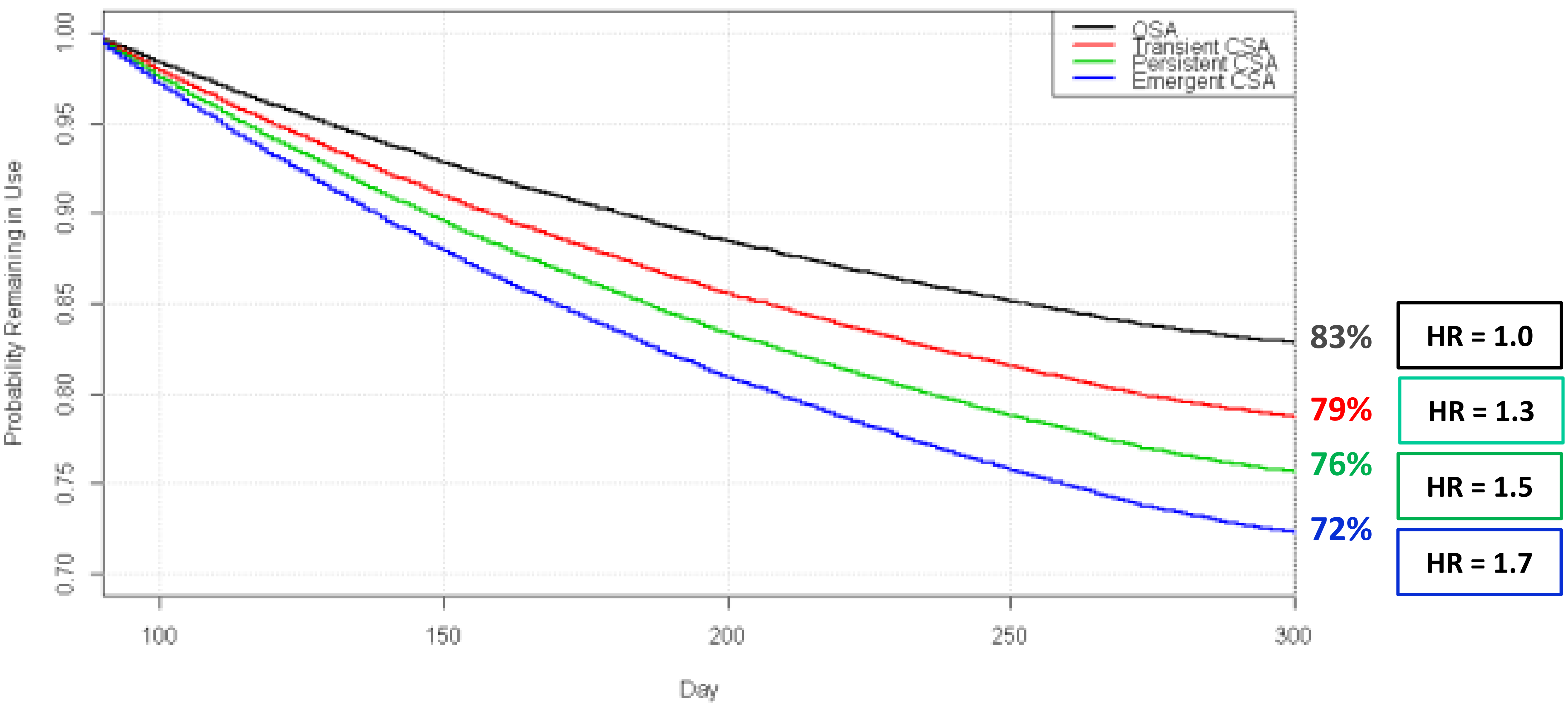
Transient CSA
(CAI ≥5/h week 1, <5/h week 12)

Liu et al., Chest 2017

"CSA Trajectories"

CSA = CAI > 5

Therapy Termination



Liu et al., Chest 2017

AirView – ASV Big Data: Switchers

pii: jc-17-00284

http://dx.doi.org/10.5664/jcsm.6880

JCSM
Journal of Clinical Sleep Medicine

SCIENTIFIC INVESTIGATIONS

Adherence to Positive Airway Therapy After Switching From CPAP to ASV: A Big Data Analysis

Jean-Louis D. Pépin, MD[1]; Holger Woehrle, MD[2,3]; Dongquan Liu, PhD[4]; Shiyun Shao, PhD[4]; Jeff P. Armitstead, PhD[3]; Peter A. Cistulli, MD[5]; Adam V. Benjafield, PhD[6]; Atul Malhotra, MD[7]

[1]Institut National de la Santé et de la Recherche Médicale (INSERM), HP2 Laboratory (Hypoxia: Pathophysiology), Grenoble Alpes University, Grenoble, France; [2]Sleep and Ventilation Center Blaubeuren, Respiratory Center Ulm, Ulm, Germany; [3]ResMed Science Center, Sydney, Australia; [4]ResMed Asia Ltd, Singapore; [5]Charles Perkins Centre, University of Sydney, and Royal North Shore Hospital, Sydney, Australia; [6]ResMed Science Center, ResMed Corp., San Diego, California; [7]University of California San Diego, La Jolla, California

pii: jc-17-00284 http://dx.doi.org/10.5664/jcsm.6880

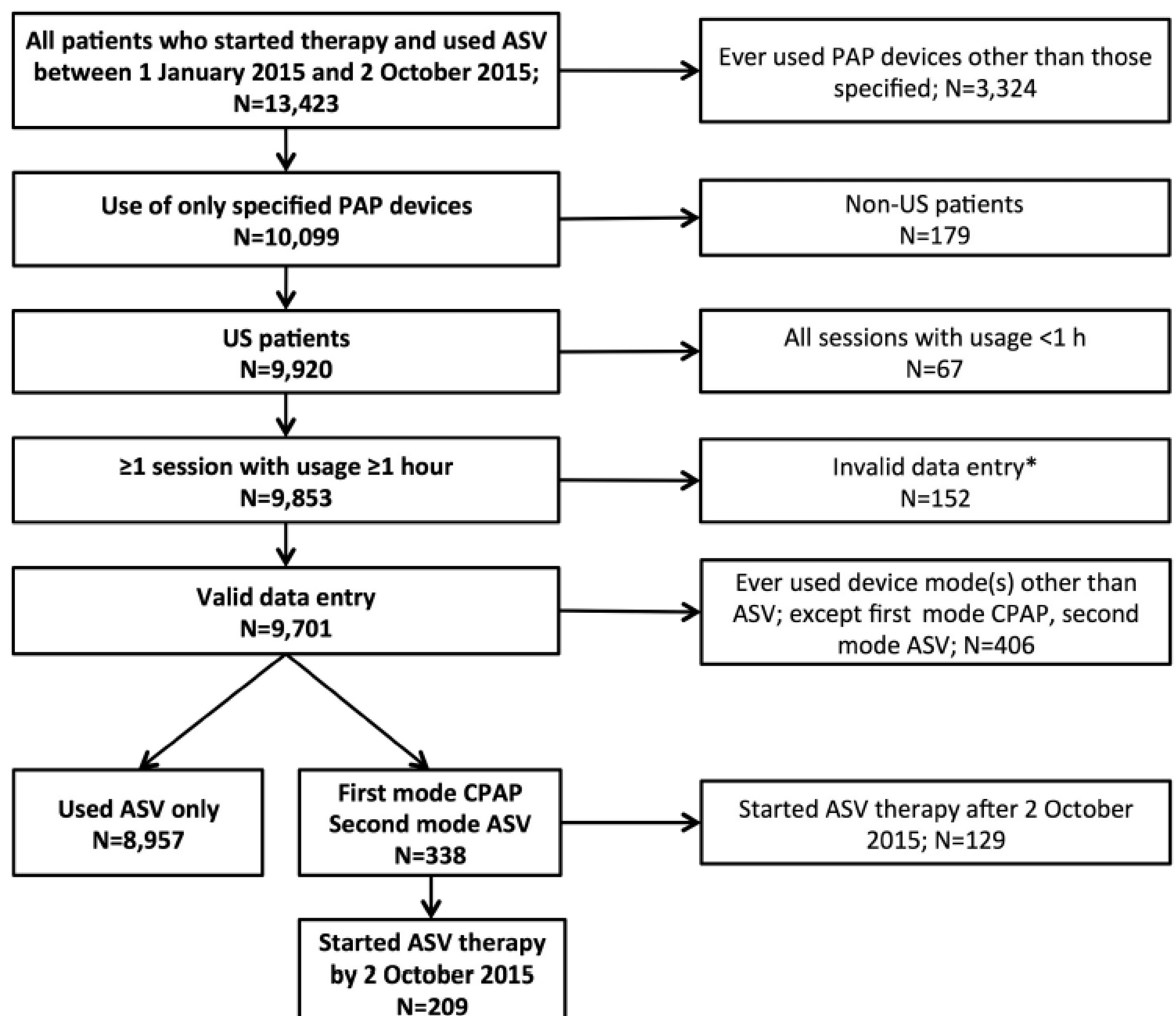
JCSM
Journal of Clinical Sleep Medicine

SCIENTIFIC INVESTIGATIONS

Adherence to Positive Airway Therapy After Switching From CPAP to ASV: A Big Data Analysis

Jean-Louis D. Pépin, MD[1]; Holger Woehrle, MD[2,3]; Dongquan Liu, PhD[4]; Shiyun Shao, PhD[4]; Jeff P. Armitstead, PhD[3]; Peter A. Cistulli, MD[5]; Adam V. Benjafield, PhD[6]; Atul Malhotra, MD[7]

B



pii: jc-17-00284

http://dx.doi.org/10.5664/jcsm.6880

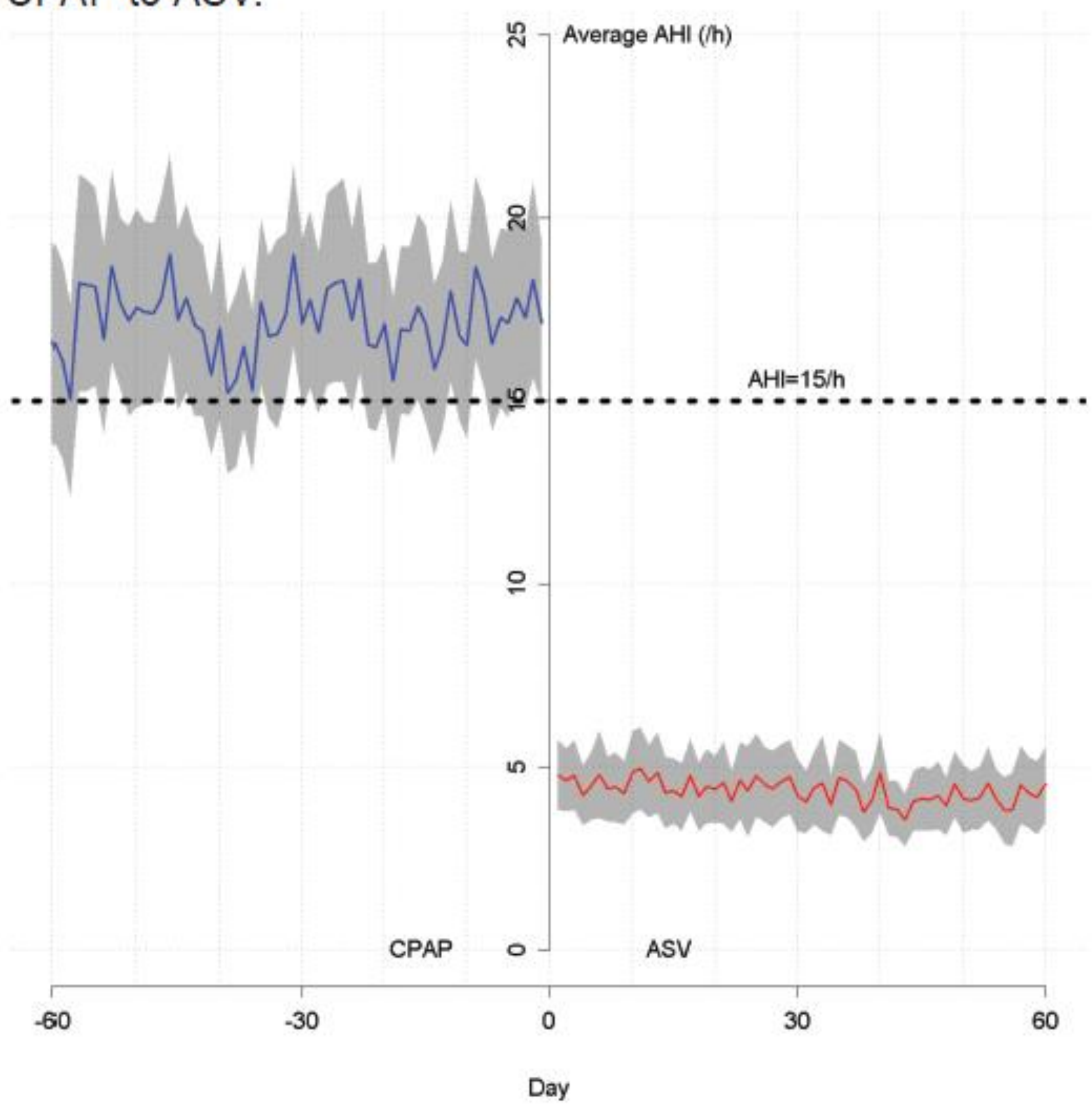
JCSM
Journal of Clinical Sleep Medicine

SCIENTIFIC INVESTIGATIONS

Adherence to Positive Airway Therapy After Switching From CPAP to ASV: A Big Data Analysis

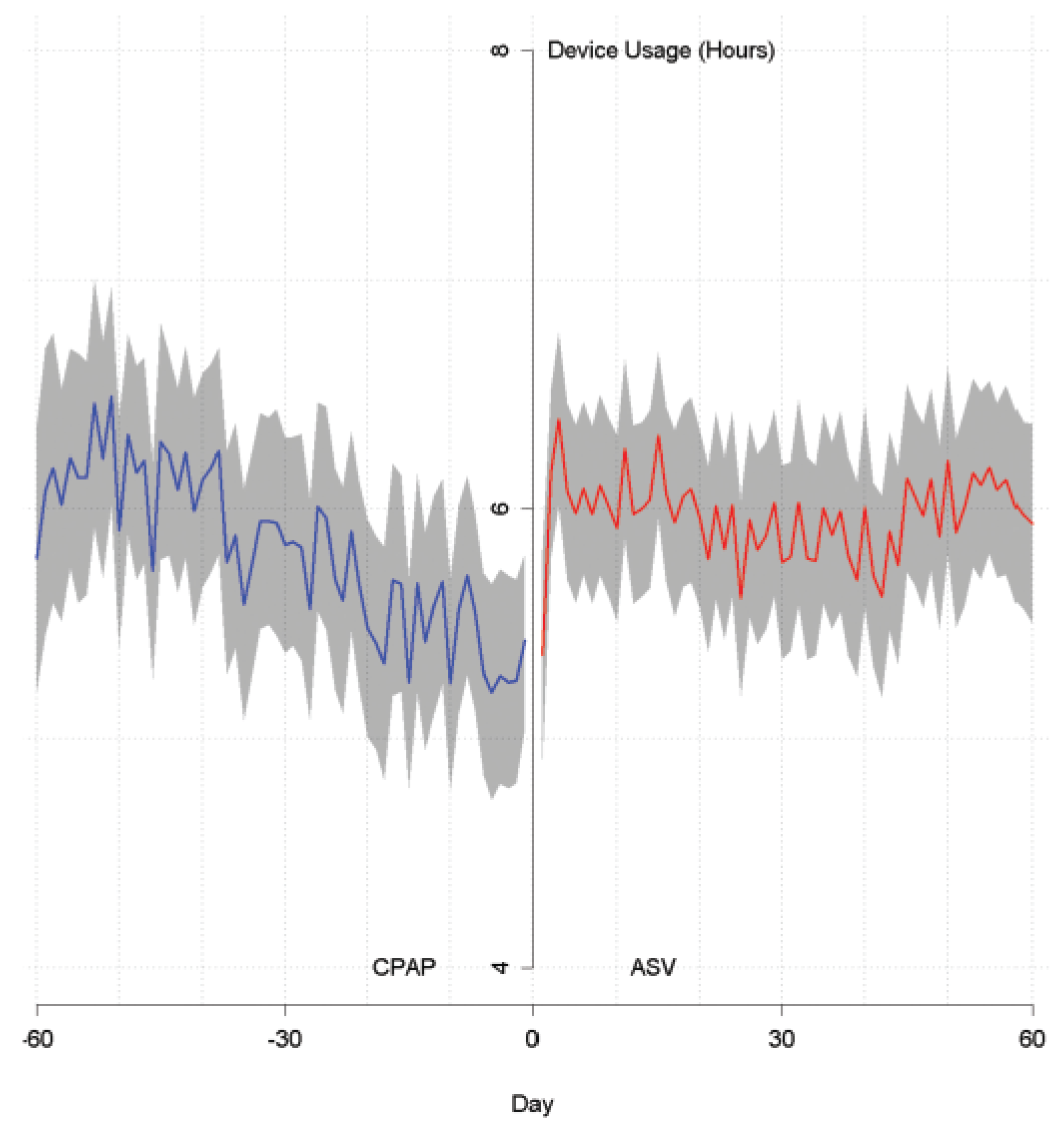
Jean-Louis D. Pépin, MD[1]; Holger Woehrle, MD[2,3]; Dongquan L

Figure 3—Average AHI before versus after the switch from CPAP to ASV.



AHI = apnea-hypopnea index, ASV = adaptive servoventilation, CPAP = continuous positive airway pressure.

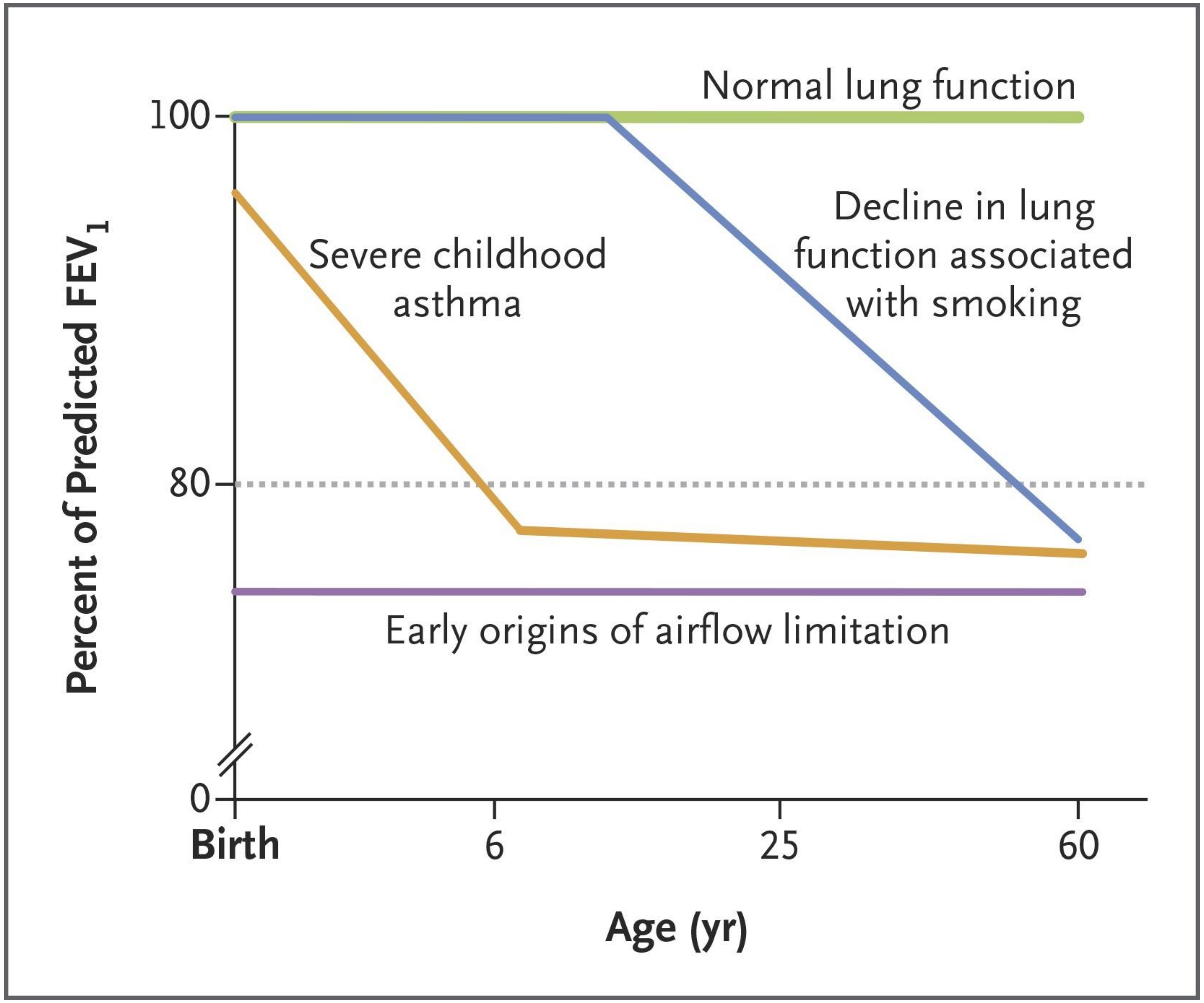
Figure 2—Trajectories of average PAP usage before versus after the switch from CPAP to ASV.



ASV = adaptive servoventilation, CPAP = continuous positive airway pressure.

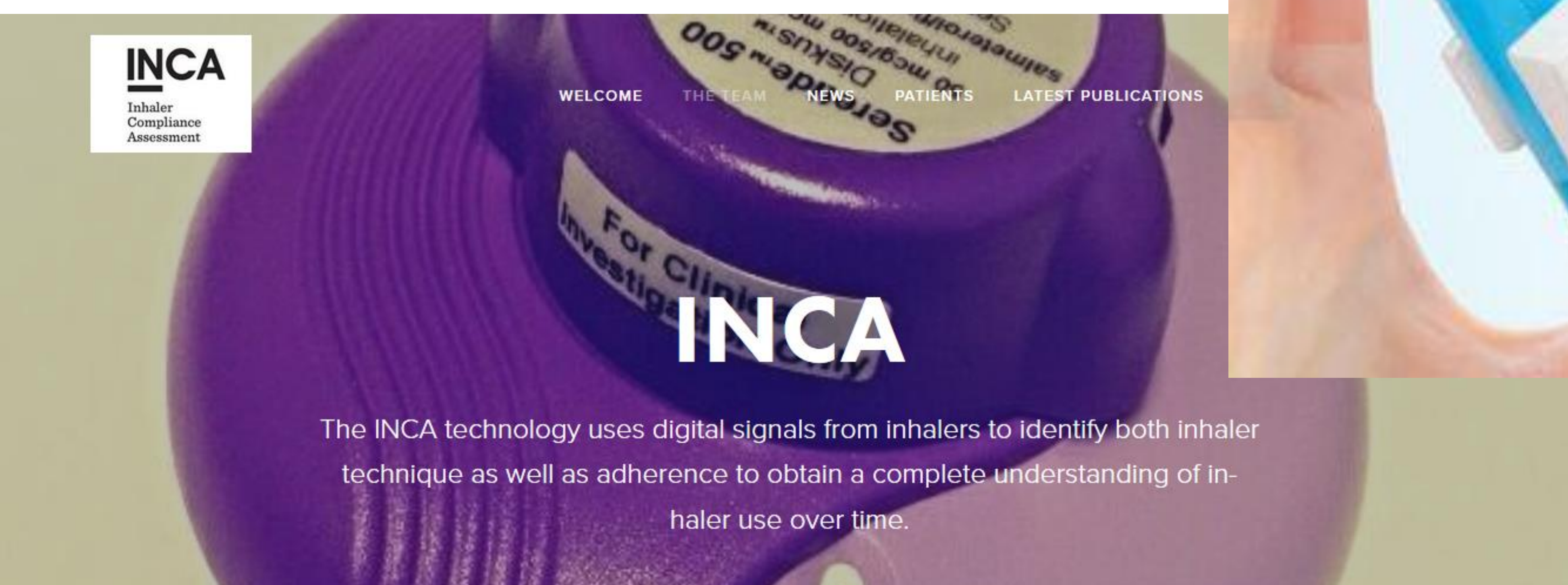
Digital Spirometry

Three Major Lung-Function Pathways That Lead to Stage 2 Chronic Obstructive Pulmonary Disease



Martinez FD. N Engl J Med 2016;375:871-878

Smarthaler/E-haler



http://www.incadevice.com;/https://de.pinterest.com/pin/178947785173369028/

Data is the new oil

Big data: a new era in sleep medicine?

H. Woehrle
Sleep and Ventilation Center Blaubeuren/
Lung Center Ulm
hwoehrle@lungenzentrum-ulm.de