e-Health & e-Epidemiology

how to improve population health
in the age of digital and Artificial Intelligence?

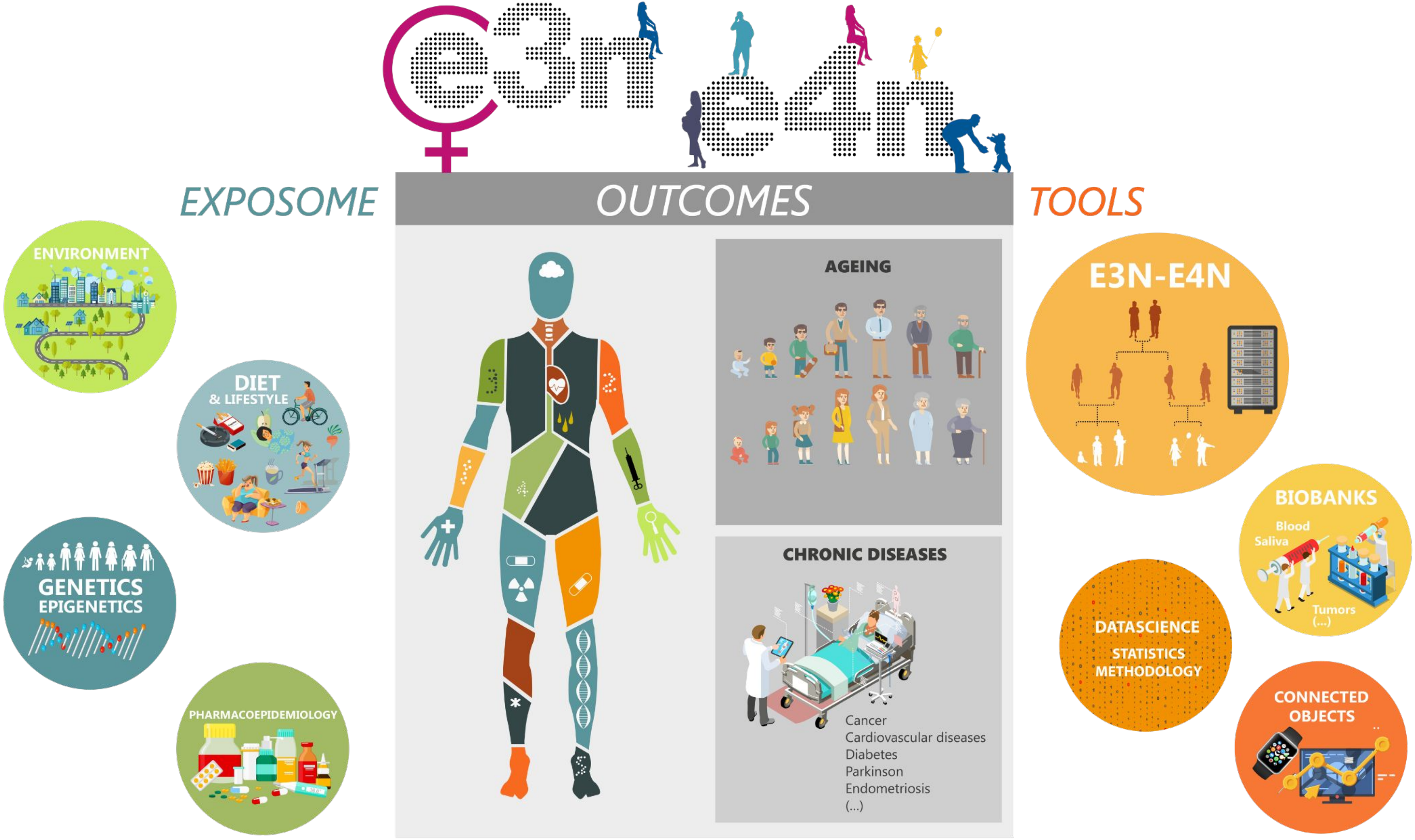
Guy Fagherazzi, MSc, PhD, HDR
Senior Research Scientist in Digital & Diabetes Epidemiology
CESP Inserm U1018
Twitter @GFaghe / Email guy.fagherazzi@gmail.com

Inserm
La science pour la santé
From science to health

UNIVERSITÉ PARIS SUD
université PARIS-SACLAY

GUSTAVE ROUSSY
CANCER CAMPUS GRAND PARIS

Center of Research in Epidemiology & Population Health (Villejuif, France)

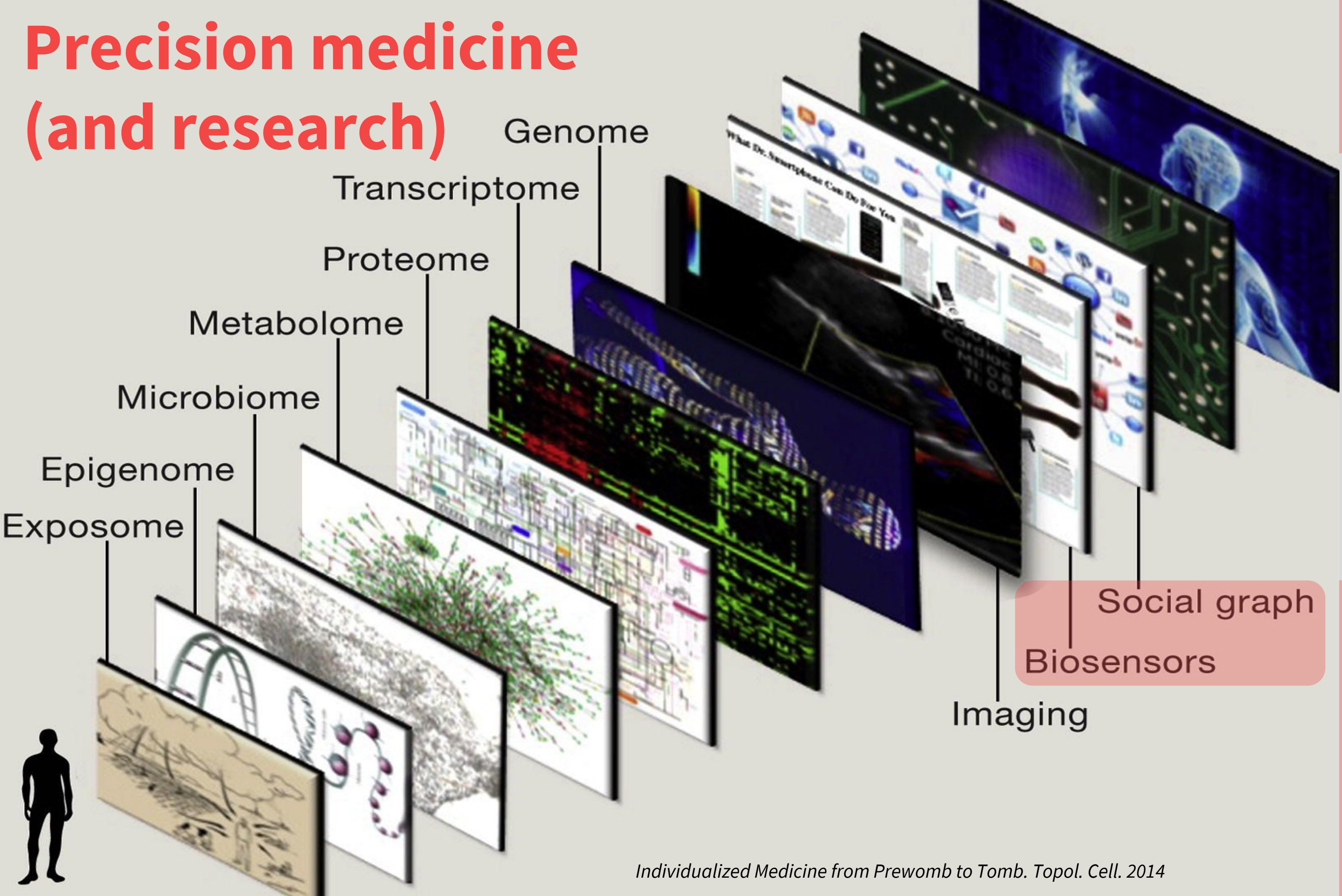

e3n e4n
EXPOSOME
ENVIRONMENT
DIET & LIFESTYLE
GENETICS EPIGENETICS
PHARMACOEPIDEMIOLOGY
OUTCOMES
AGEING
CHRONIC DISEASES
Cancer
Cardiovascular diseases
Diabetes
Parkinson
Endometriosis
(...)
TOOLS
E3N-E4N
BIOBANKS
Blood
Saliva
Tumors
(...)
DATASCIENCE
STATISTICS
METHODOLOGY
CONNECTED OBJECTS

There is a new "omic" in town.

Genome
Epigenome
Proteome
Transcriptome
Metabolome
Microbiome
….

"DIGITOSOME"
= all the data generated online by an individual (smartphones, apps, connected devices/sensors, social media, geolocation)

Precision medicine (and research)

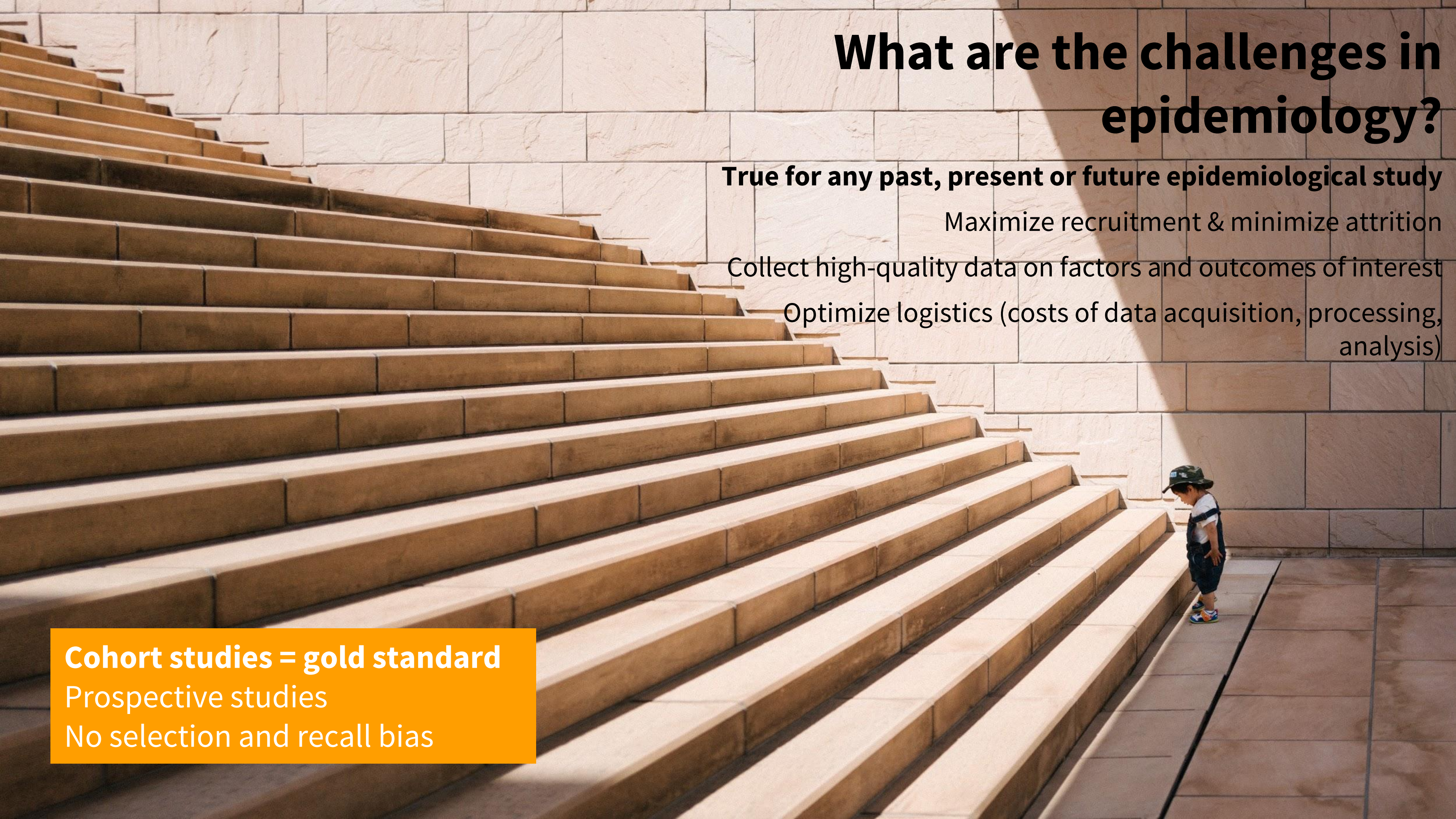


Individualized Medicine from Prewomb to Tomb. Topol. Cell. 2014

More & more data
A variety of heterogeneous "omic" data

Digitosome

Data generated online by an individual over the lifecourse
(smartphones, apps, connected devices, social media, online patients community…)



What are the challenges in epidemiology?

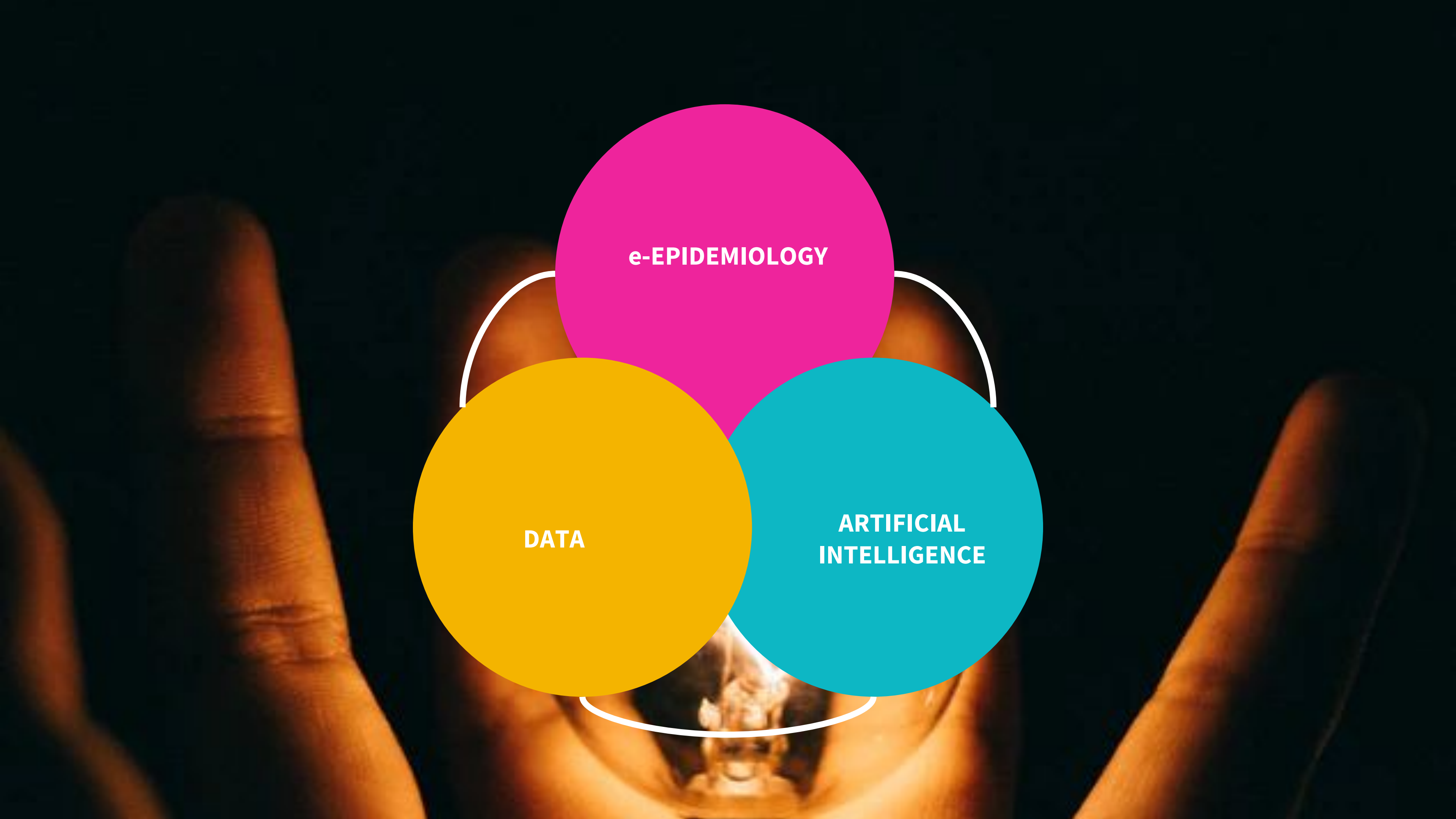
True for any past, present or future epidemiological study

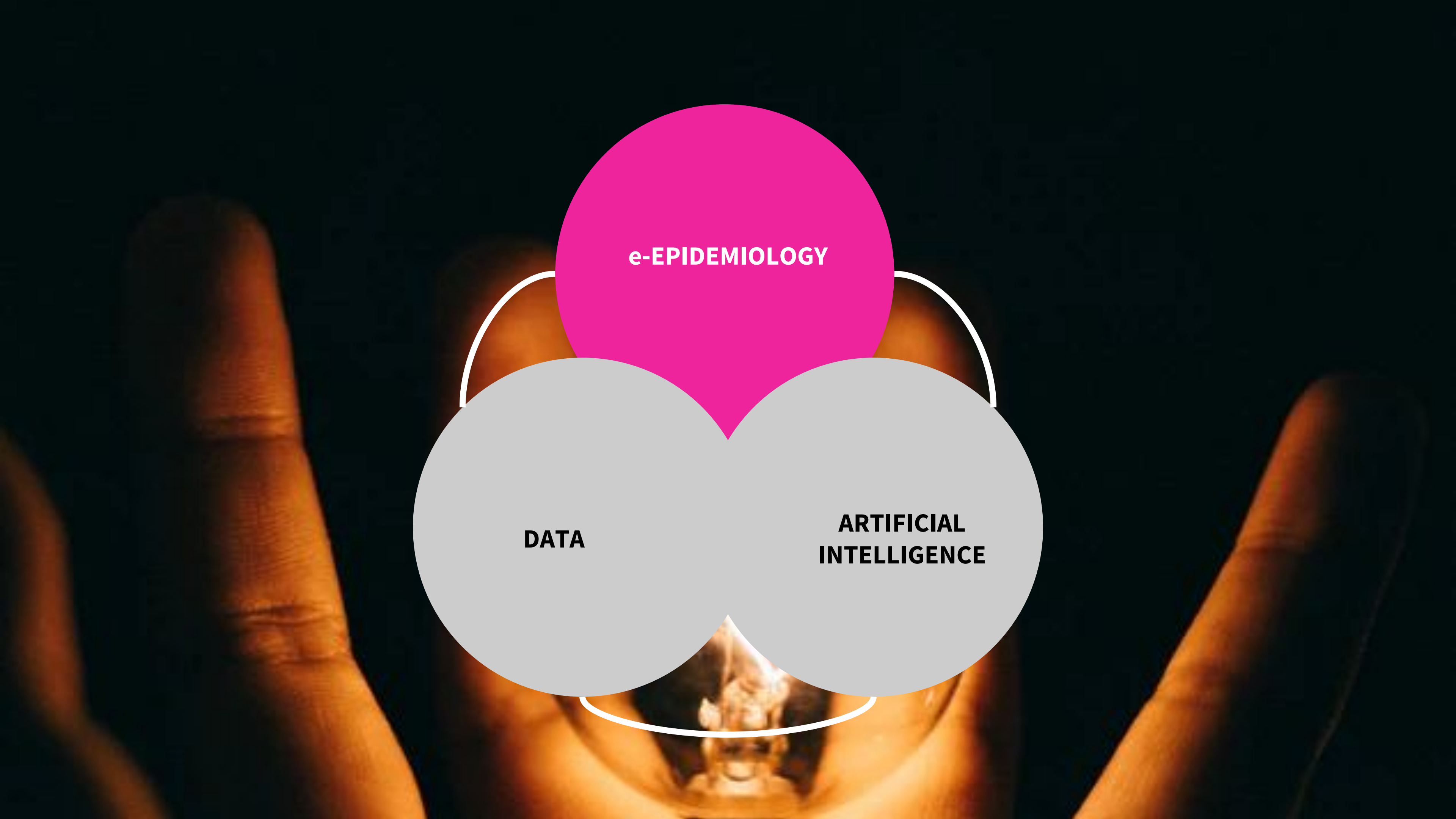
Maximize recruitment & minimize attrition

Collect high-quality data on factors and outcomes of interest

Optimize logistics (costs of data acquisition, processing, analysis)

Cohort studies = gold standard
Prospective studies
No selection and recall bias


e-EPIDEMIOLOGY
DATA
ARTIFICIAL INTELLIGENCE


e-EPIDEMIOLOGY
DATA
ARTIFICIAL INTELLIGENCE

History of cohort studies

40s-50s
(<10 000 participants)

First cohort studies

The Framingham Heart Study (1948)

The study of the Japanese atomic bomb survivors (1950)

70s
(10 000 - 100 000 participants)

First modern large cohort studies

The Nurses' Health Study (1976)

90s
(> 100 000 participants)

A new dimension in the design and the data collection

The Million Women Study UK (1996)

The E3N cohort study (1990)

The largest study on women's health in France

And one of the largest cohort studies in the world at the time

98 995 women + 28 years of follow-up

Self-reported questionnaires sent every 2-3 years since 1990

+ Rich biobank (25,000 blood & 47,000 saliva samples)

+ Medico-administrative database

Cancer **Diabetes** **Nutrition** **Reproductive factors and hormones**

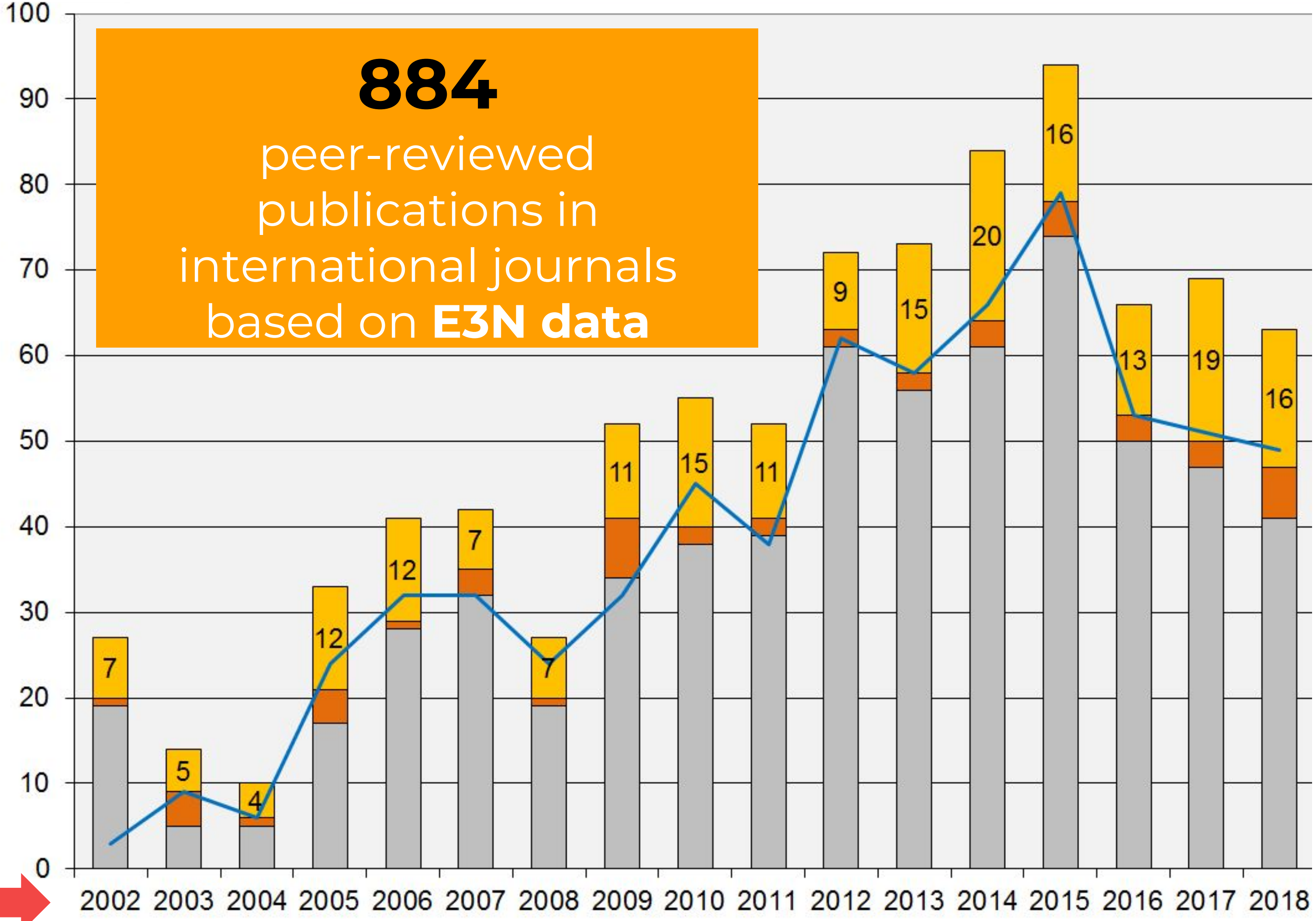
Pioneer on the data acquisition in 1990

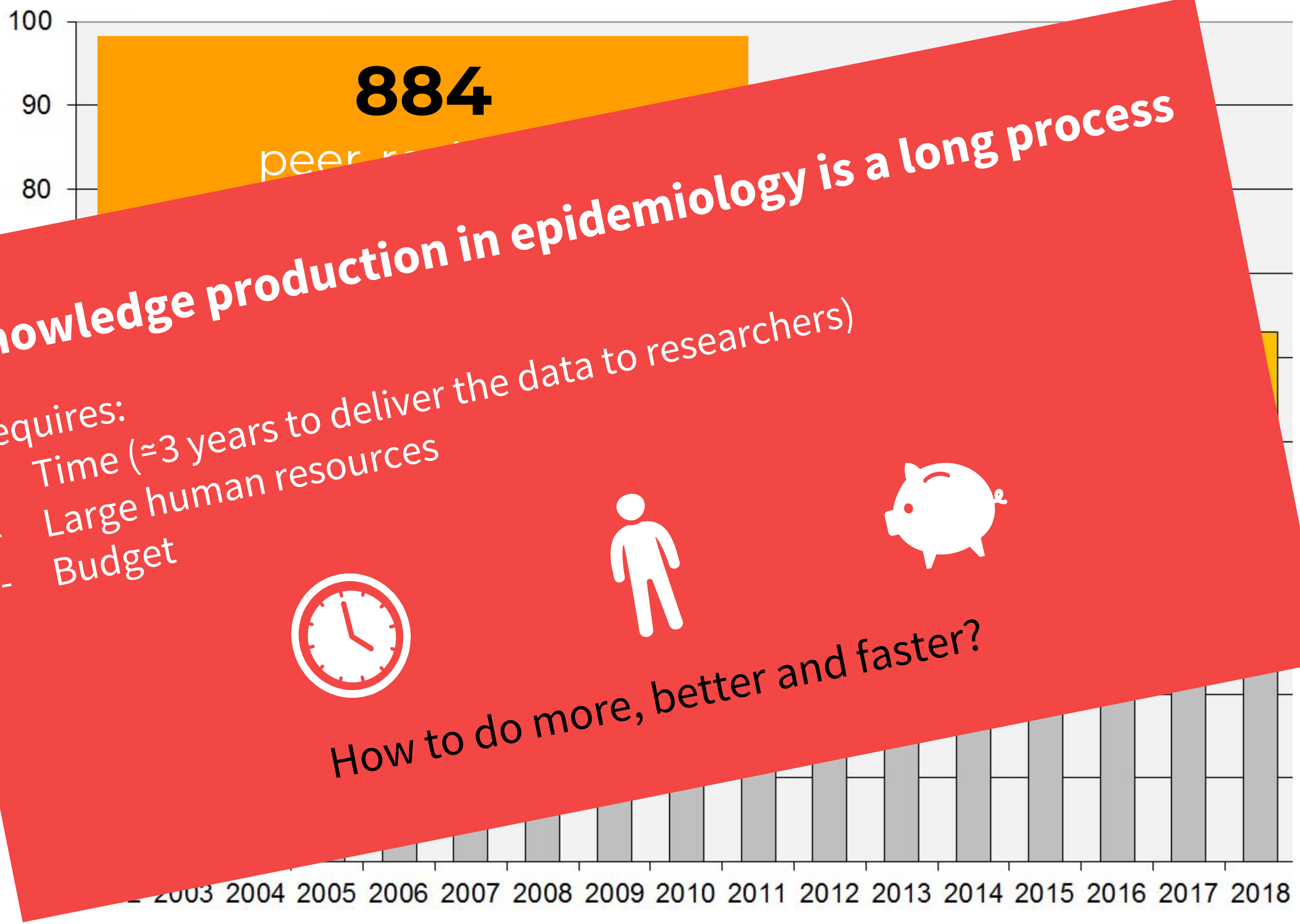
- Optical scan of paper questionnaires (Rate of 600 pages of questionnaires per hour)
- Systems of Automatic Document Reading & Character Recognition
- Video-coding and data checking

Ability to handle a large amount of data on many volunteers

884

peer-reviewed publications in international journals based on **E3N data**





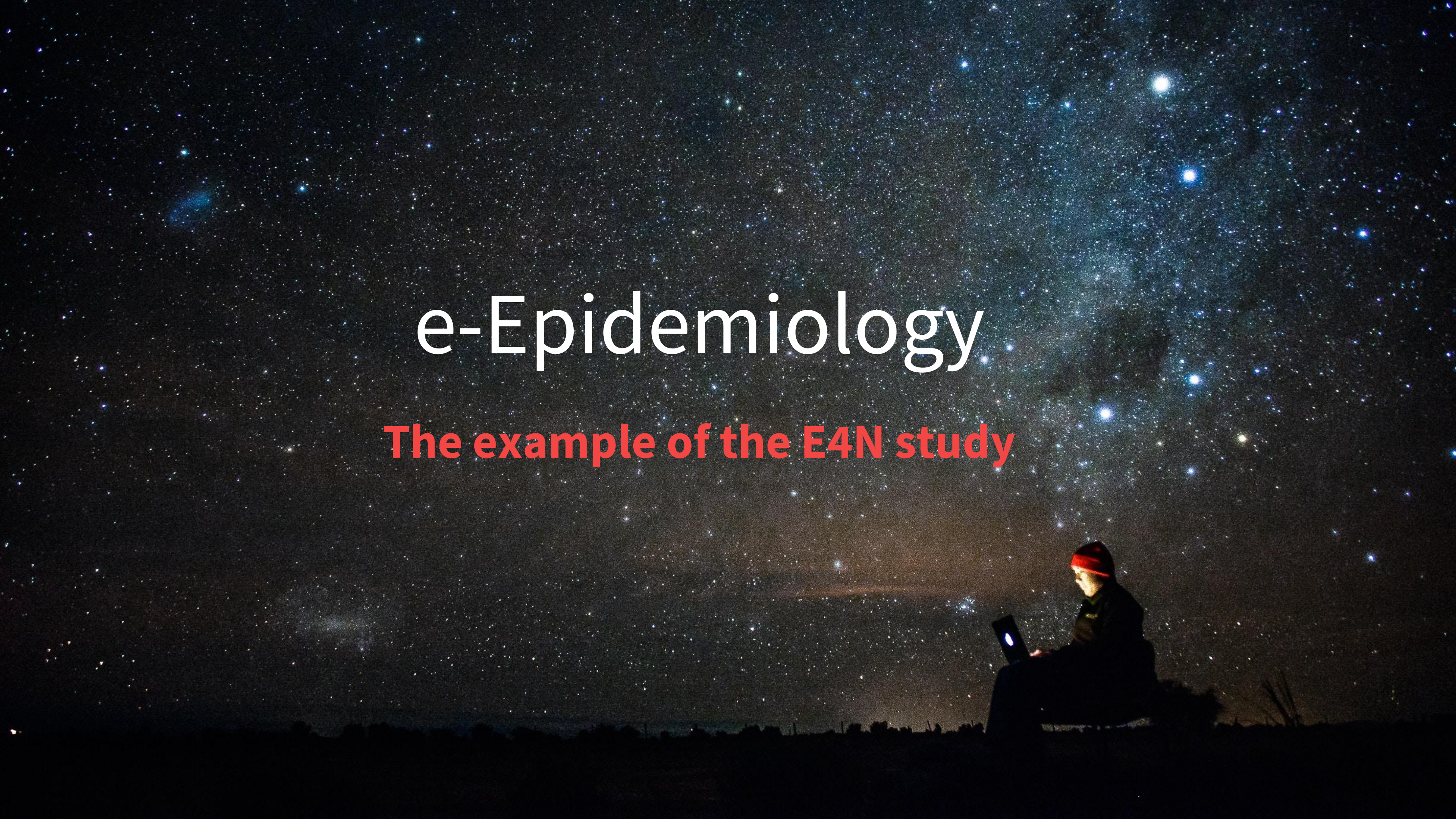
Knowledge production in epidemiology is a long process

It requires:

- Time (≃3 years to deliver the data to researchers)
- Large human resources
- Budget

How to do more, better and faster?

We need to kill the "Silo Effect"!


e-Epidemiology
The example of the E4N study

The E4N prospective study

A unique family cohort study!

Selected as an *"Investment for the Future"* by the French National Research Agency (ANR)

200 000 participants

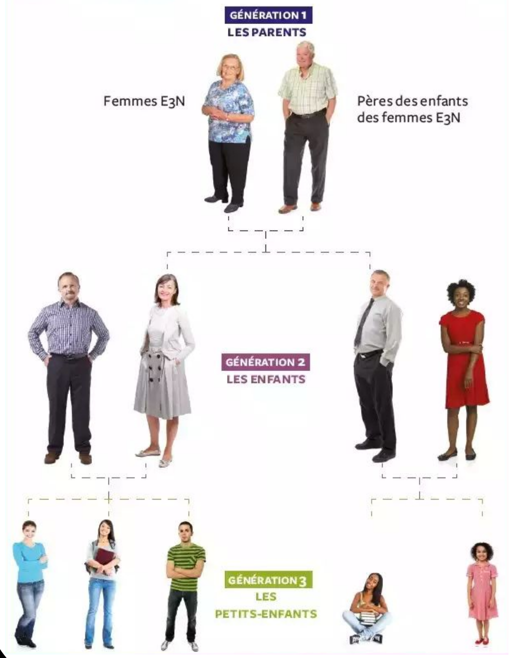
100 000 E3N women

20 000 fathers of E3N women's children

50 000 children (*ongoing*)

20 000 grand-children (*in 2019*)

3 generations



1st generation: paper questionnaires

2nd & 3rd generations: e-cohort

Saliva samples for all the participants

Trans-generational

- ➢ Heredity and transmission of health determinants
- ➢ Genetics and epigenetics of chronic diseases (cancer, diabetes,...)

Expertise on exposures

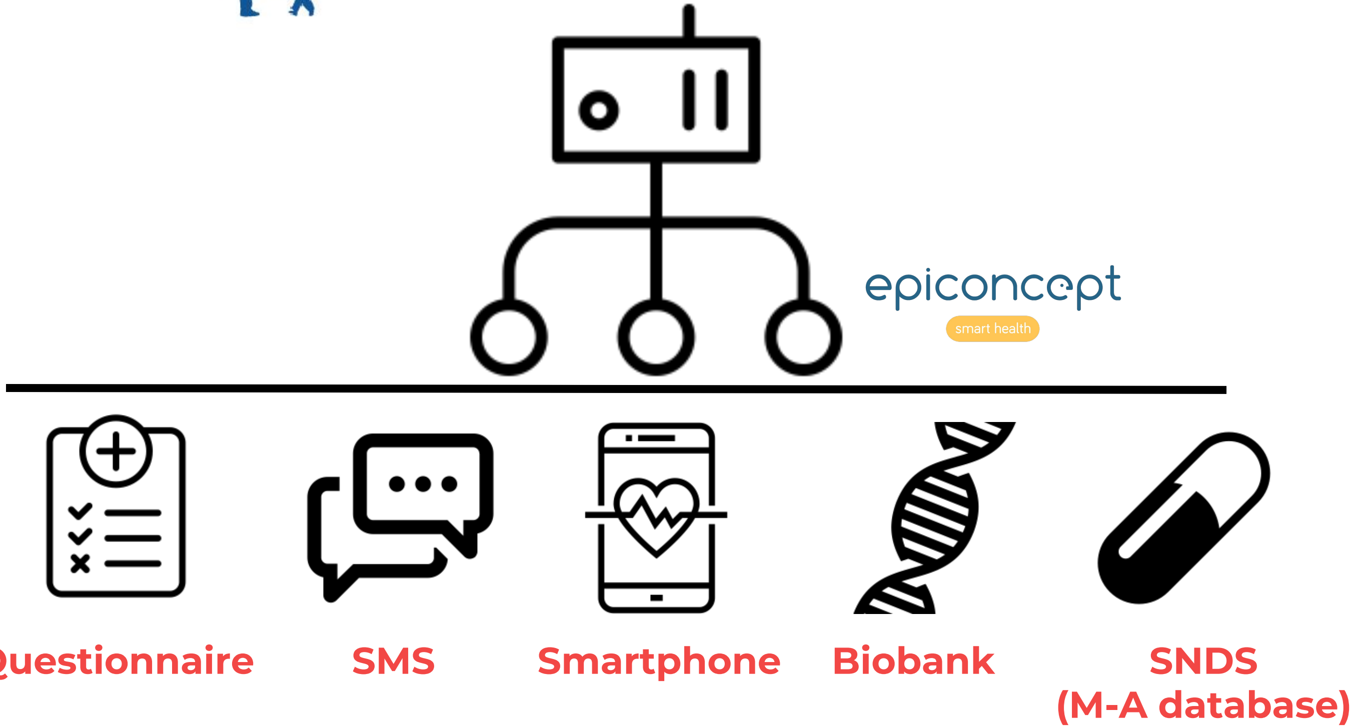
- ➢ Epigenetics on lifestyle (smoking, physical activity, diet)
- ➢ Lifestyle and microbiota
- ➢ Socioeconomic mobility through generations and its impact on lifestyle and health

e-epidemiology

- ➢ Integrate new technologies and the Internet of Things in modern epidemiology to collect high-quality data

Some research axes

Data Hub


epiconcept
smart health
Questionnaire
SMS
Smartphone
Biobank
SNDS
(M-A database)

The E4N Platform

- Short (but frequent) online questionnaires
- Answer from a laptop, smartphone or a tablet
- Questionnaires sent by SMS
- Sync with connected objects
- Automatic feedback and statistical dashboards for the participants
- Awards and badges (gamification)

A “connected” sub-cohort study

N = 700 E4N participants

Real-life study of lifestyle factors evaluated with a connected tracker

Study the associations between lifestyle factors and psychological well-being in participants with breast cancer or type 2 diabetes

Evaluation of the predictive capacity of the connected tracker data

Future extensions planned with other connected devices

Inserm
Institut national
de la santé et de la recherche médicale

Can we observe known associations with connected devices?

JMIR Publications
Journal of Medical Internet Research ISSN 1438-8871
The leading peer-reviewed journal for health and healthcare in the Internet age.
Open Access • Top Cited (Impact Factor 4.5) • Rapid Peer-Review • Medline Indexed (+20 other indices)

Determinants of a poor sleep based on data from 15 000 users of connected devices

Guy Fagherazzi, PhD[1,2]; Douae El Fatouhi, MSc[1,2]; Alice Bellicha[3,4], MSc; Amin El-Gareh[1,2], MSc; Aurélie Affret, MSc[1,2]; Courtney Dow, MPH[1,2]; PhD; Lidia Delrieu, MSc[5,6]; Matthieu Vegreville, MSc[7]; Alexis Normand, MSc[7]; Jean-Michel Oppert, MD, PhD[3]; Gianluca Severi, PhD[1,2]

Data from “hyperconnected” users of Withings/Nokia devices between 2013 and 2016

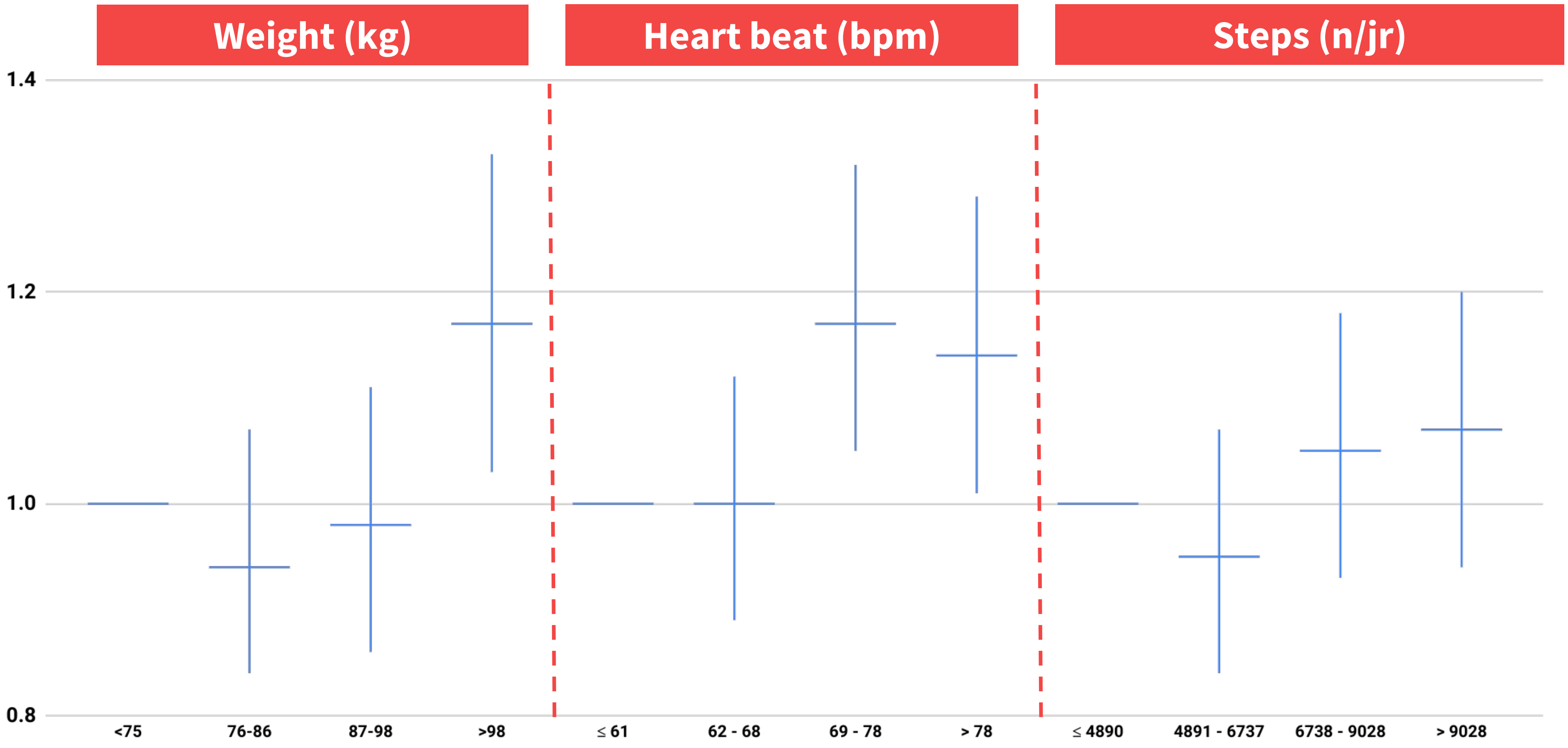
15 839 individuals (13 658 men & 2 181 women)

AI clustering methods (ongoing)

Data

Sleep (Total, Deep, Ratio Deep/total), Age, Sex, Weight, Nb of steps, Heart beat, Blood pressure

Determinants of short sleep duration (≤ 6h)



Other great e-epidemiology initiatives

In France and in the USA

ComPaRe

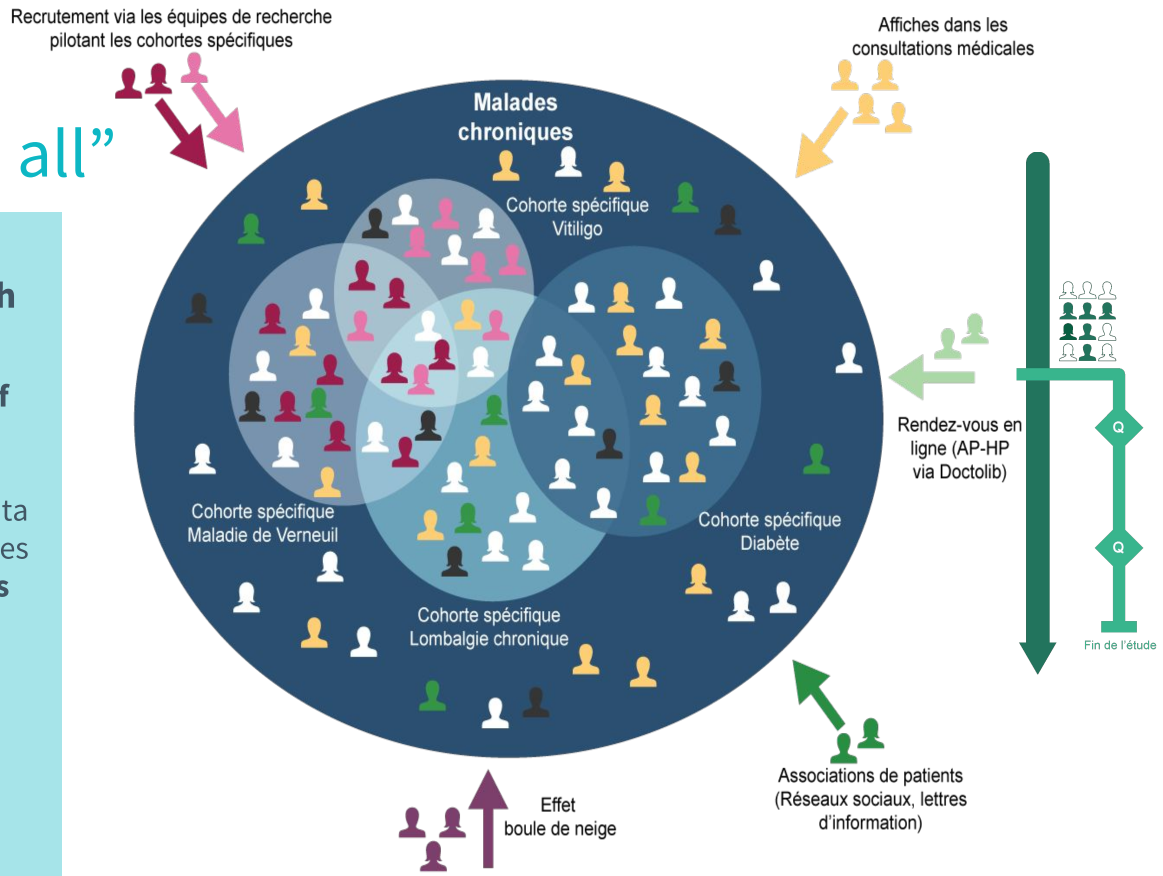
COMMUNAUTÉ DE PATIENTS POUR LA RECHERCHE

“1 study to rule them all”

An umbrella cohort for patients with any chronic disease(s)

- Study on **multimorbidity, burden of treatment/diseases, PROs**
- **Web platform** (e-questionnaires)
- Matching ongoing with the AP-HP Data Warehouse and national MA databases
- Patient recruitment for **TWICs (Trials WithIn Cohorts)**
- Patients contribute to define the research questions to investigate thanks to the Inspire platform
- New model of collaborative and mutualized research

https://compare.aphp.fr/



2010 and beyond: the mega-cohorts

NIH National Institutes of Health
All of Us Research Program

+ 1,000,000 participants

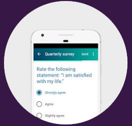
WE'VE MAPPED THE WORLD. NOW LET'S MAP HUMAN HEALTH.

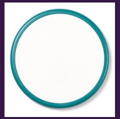
The birth of AI-oriented cohorts

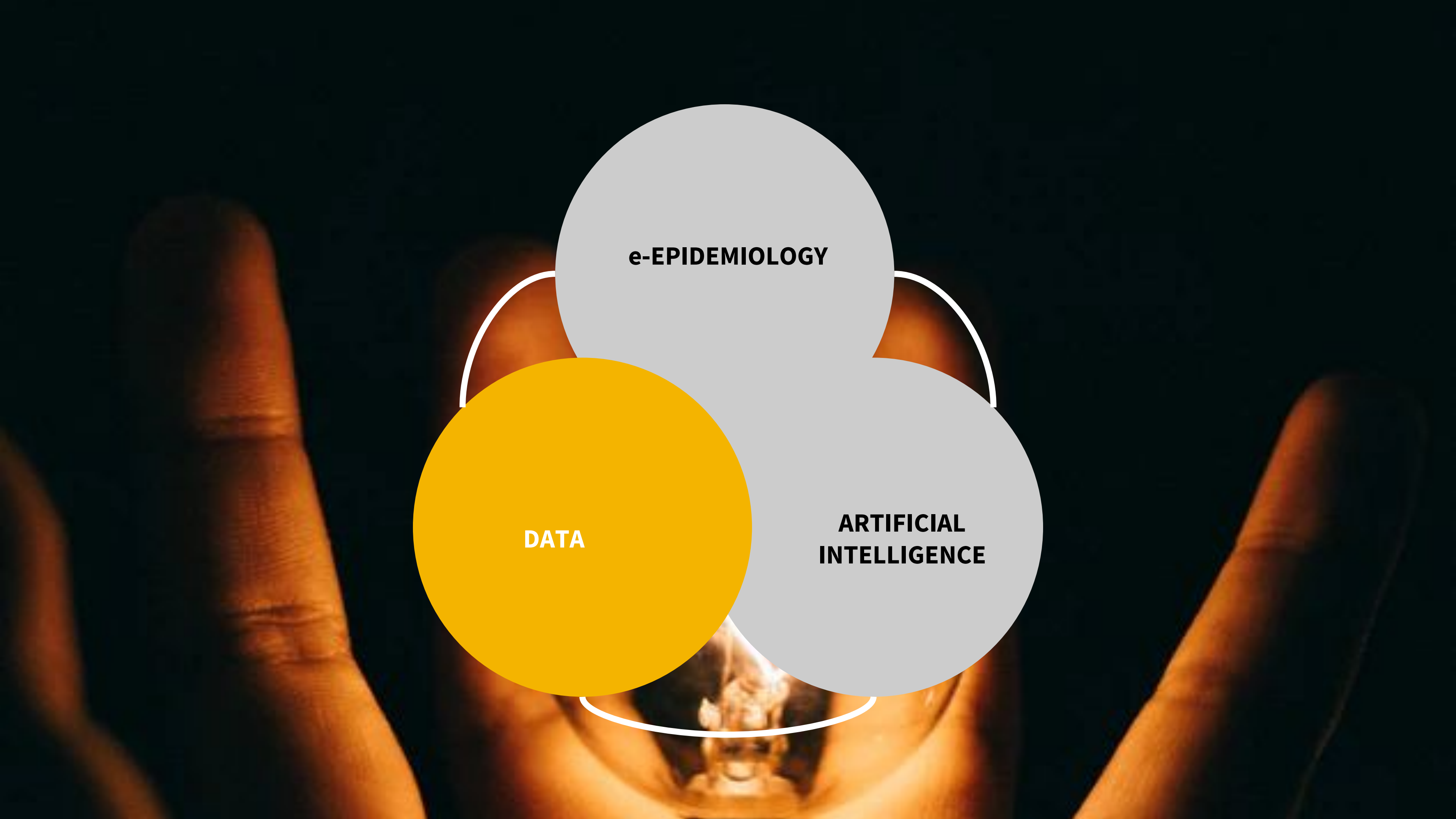
10 000 participants

Yearly clinical visits
+ biobank
+ connected devices

verily Duke University School of Medicine Stanford MEDICINE Google



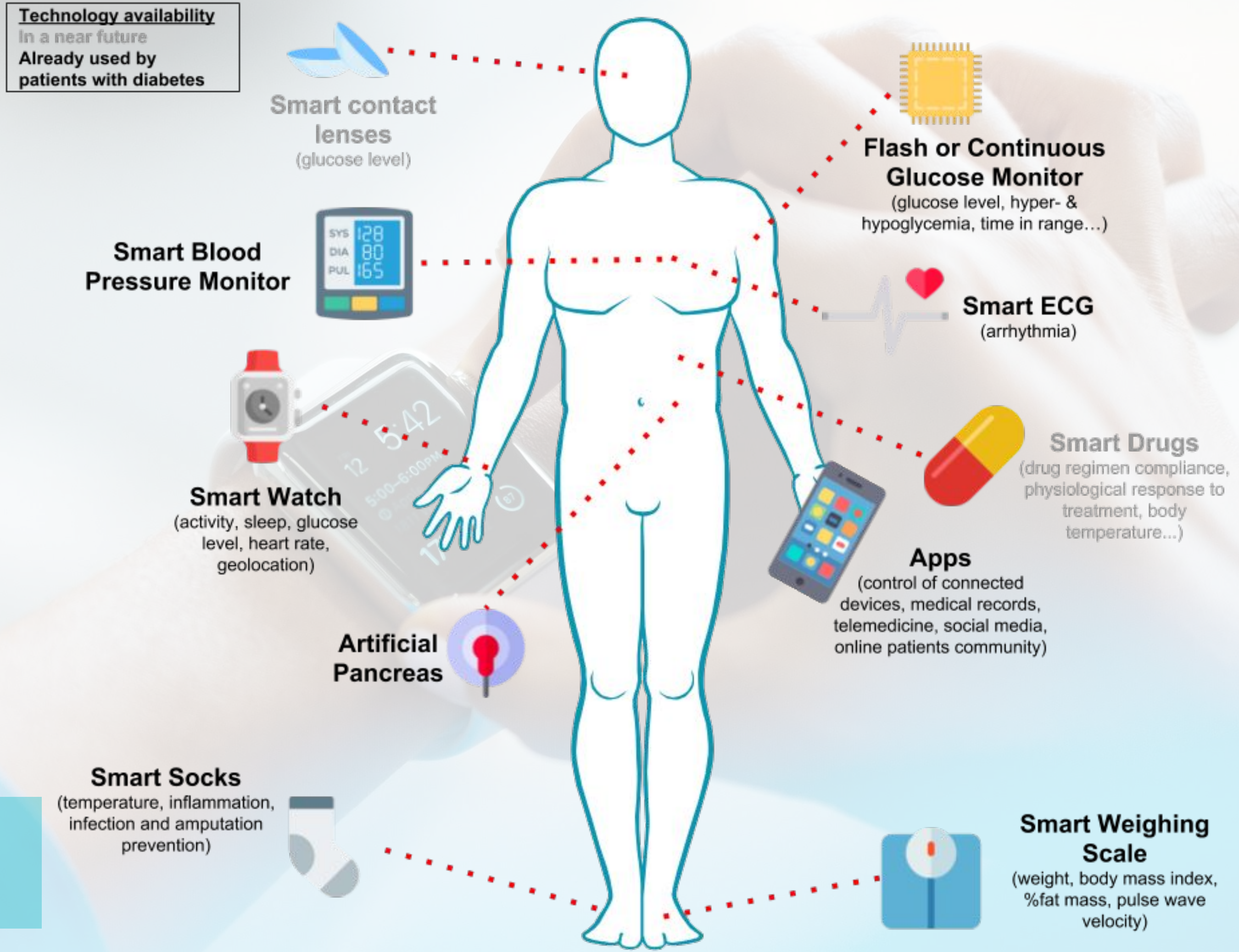



e-EPIDEMIOLOGY
DATA
ARTIFICIAL
INTELLIGENCE


The Data Rush


Technological innovation drives
modern epidemiological research

New data sources in medical research



Digital diabetes: perspectives for diabetes prevention, management and research. Fagherazzi et al. ([Diabetes & Metabolism, 2018]) - Open Access

Digital & technological innovation: from epi studies to clinical practice

FreeStyle Libre
FLASH GLUCOSE MONITORING SYSTEM

Abbott

Innovation

↓

New potential metrics (time in range, TIR)

↓

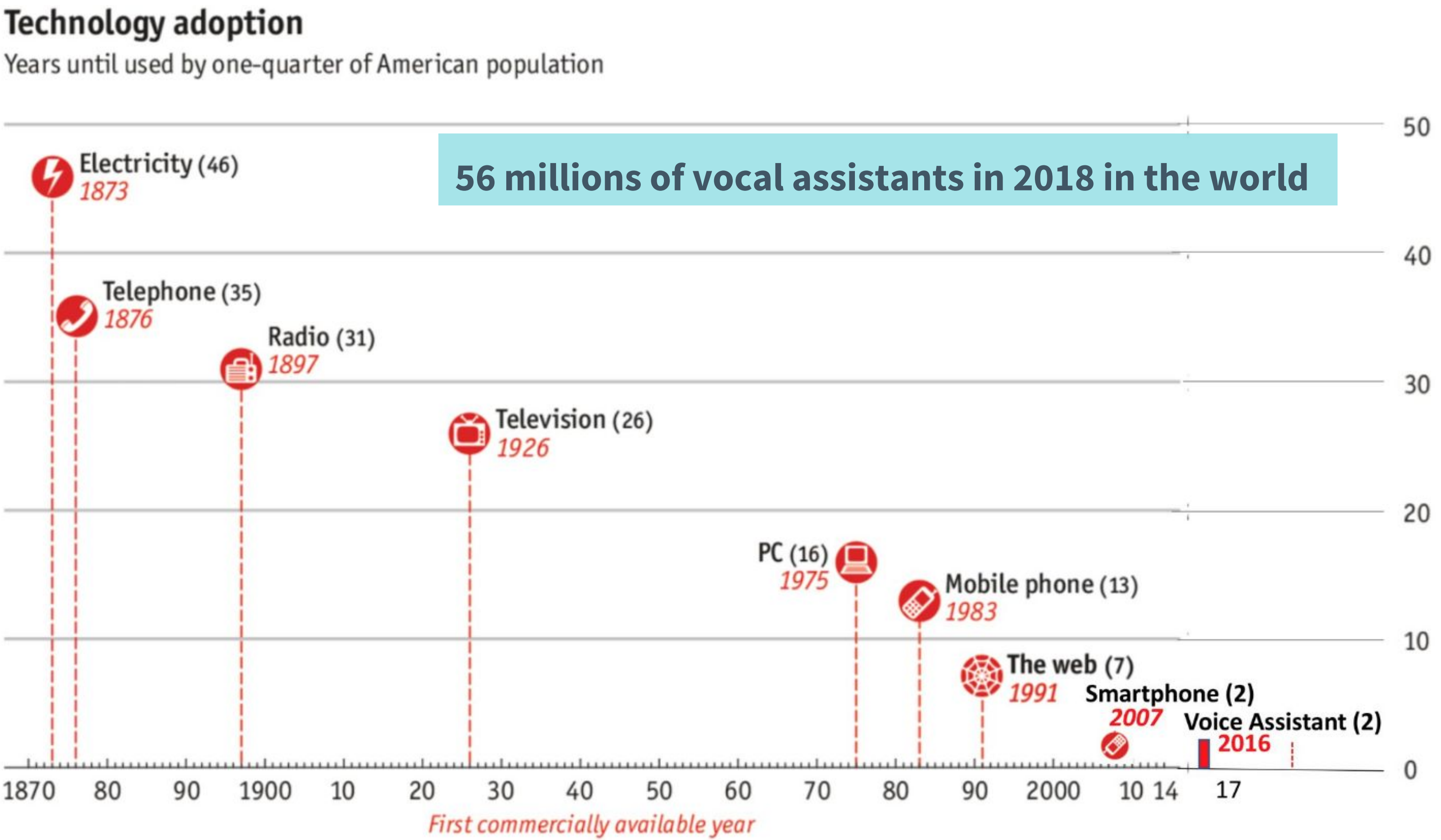
Epi studies on TIR[1]

↓

- TIR considered as a valid an outcome measure for clinical trials
- TIR used in clinical practice by physicians and diabetologists

1. Roy W. Beck et al. Diabetes Care 2018 Oct. Validation of Time in Range as an Outcome Measure for Diabetes Clinical Trials http://care.diabetesjournals.org/content/early/2018/10/17/dc18-1444

Vocal biomarkers:“the next big thing”?



Vocal biomarkers:“the next big thing”?

BEYONDVERBAL
THE EMOTIONS ANALYTICS COMPANY

Voice Signal Characteristics Are Independently Associated With Coronary Artery Disease

Elad Maor, MD, PhD; Jaskanwal D. Sara, MBChB; Diana M. Orbelo, PhD; Lilach O. Lerman, MD, PhD; Yoram Levanon, PhD; and Amir Lerman, MD

Will be useful for

- **Study of emotions and feelings of patients/study participants**
- **Early markers of risk -> prevention**
- **Diagnostic markers**

Mental health / Neurodegenerative diseases / Cardiometabolic diseases

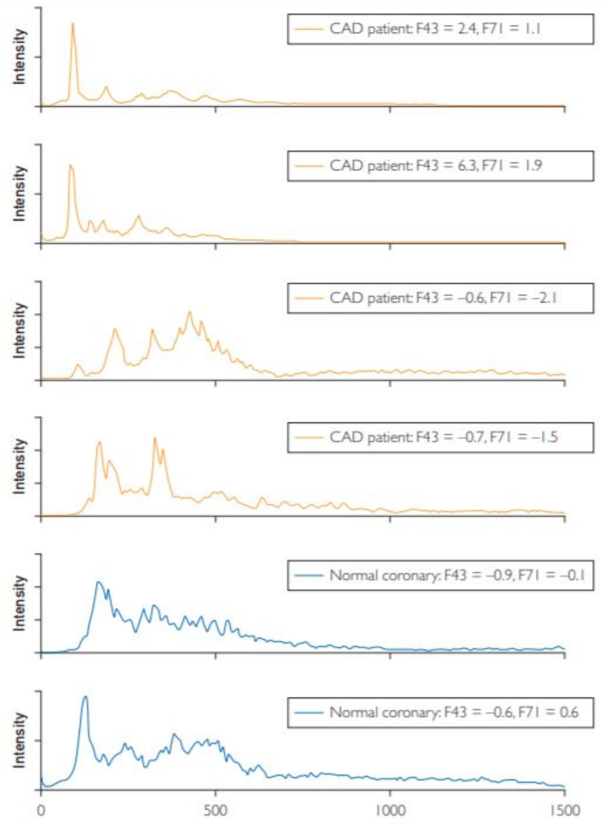


FIGURE 1. Power spectrum density plots of patients with CAD and controls. Examples of PSD of patients with and without CAD. All examples are from the third recording (negative emotional experience). The graphs show cases of patients with normal coronaries, patients with extreme decay/asymmetry of feature 43, and patients with extreme values of feature 71. Features are calculated by averaging over time the instantaneous PSD values calculated using Fourier transform on 25-ms frames with 10-ms shift. CAD = coronary artery disease; PSD = power spectrum density.

E-health records and medico-administrative databases

French administrative databases

- **SNIIRAM** (national health insurance database = claims, medical acts…)
- **PMSI** (national hospital discharge database)
- **Causes of death**

(No clinical/biological data)

Covers 96% of the French population

SNDS
Système national des données de santé

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

Health Data Warehouses (AP-HP, Paris)

- 8 millions of patients per year
- Hospital discharge database + clinical/biological/imaging results
- Used for observational studies, feasibility studies for clinical trials

Used alone = limited interest / Matched with cohorts = huge potential

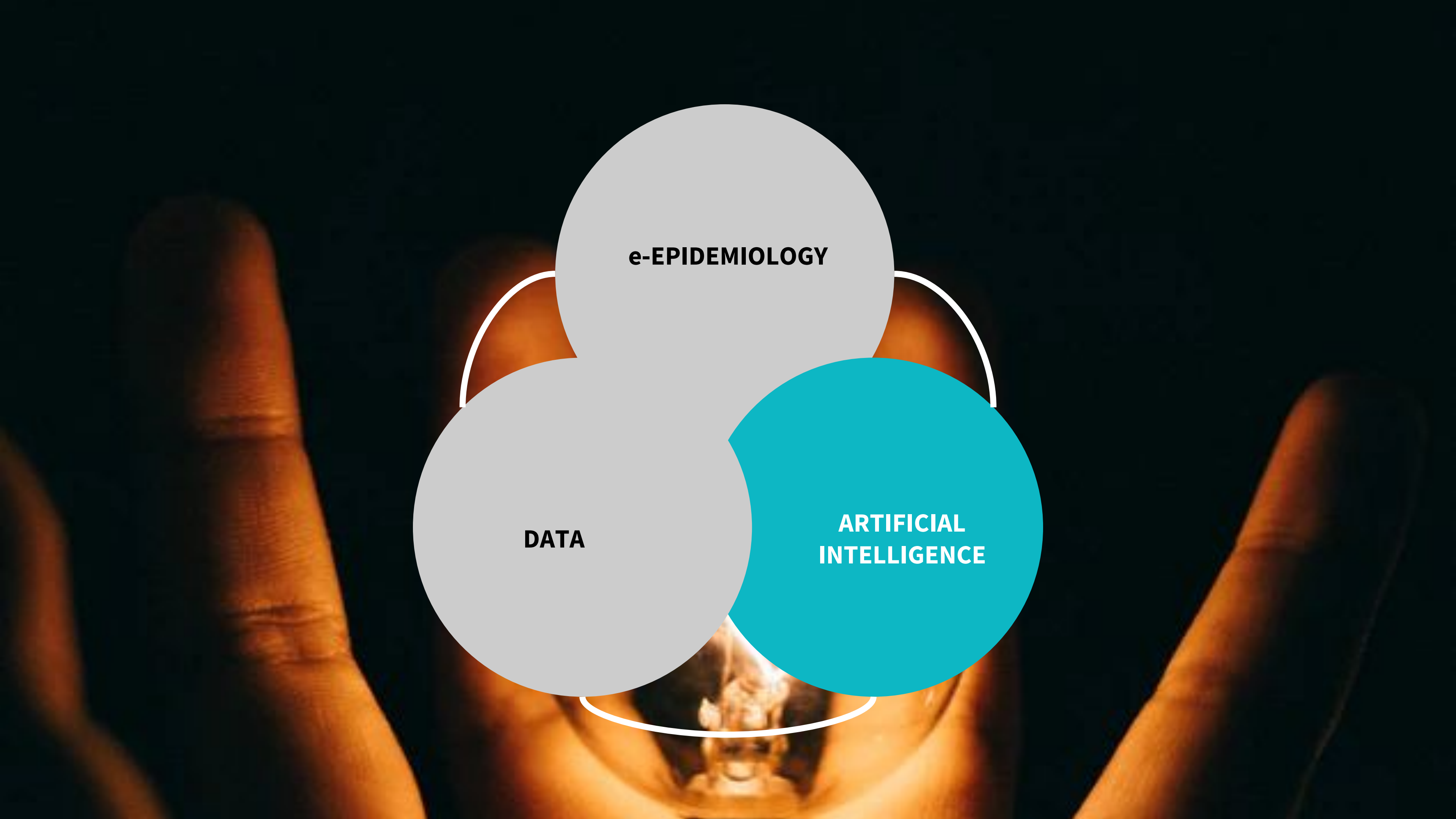
And tomorrow?

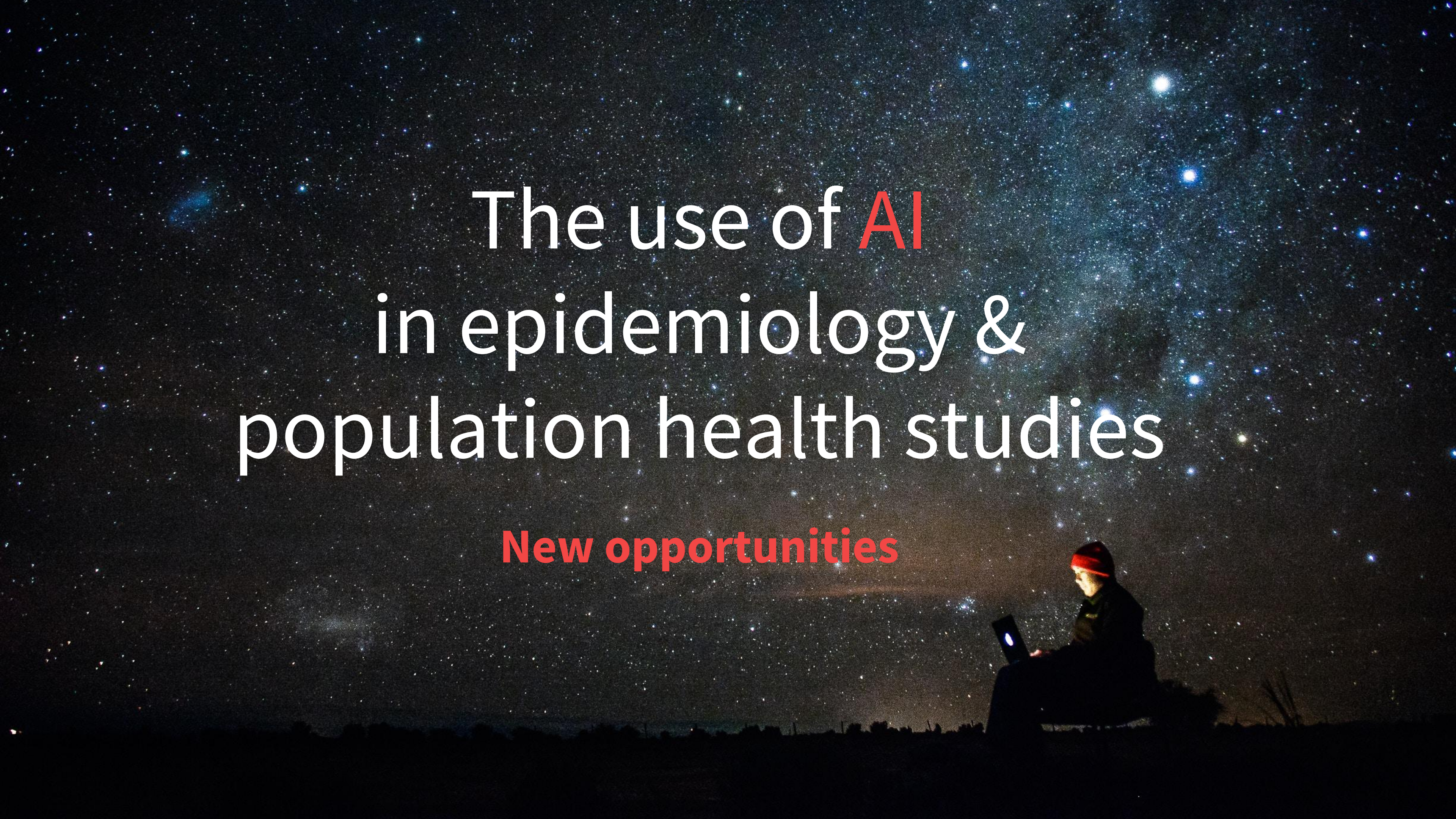
The French "Health Data Hub" will gather:

- French administrative databases
- Data from publicly funded research projects (cohorts, registries...)
- DMP (Personalized e-health records)
- Connected devices

- **Interoperability?**
- **Governance?**
- **Accessibility?**
- **Security?**
- **Sustainability?**
- **Consent from patients?**

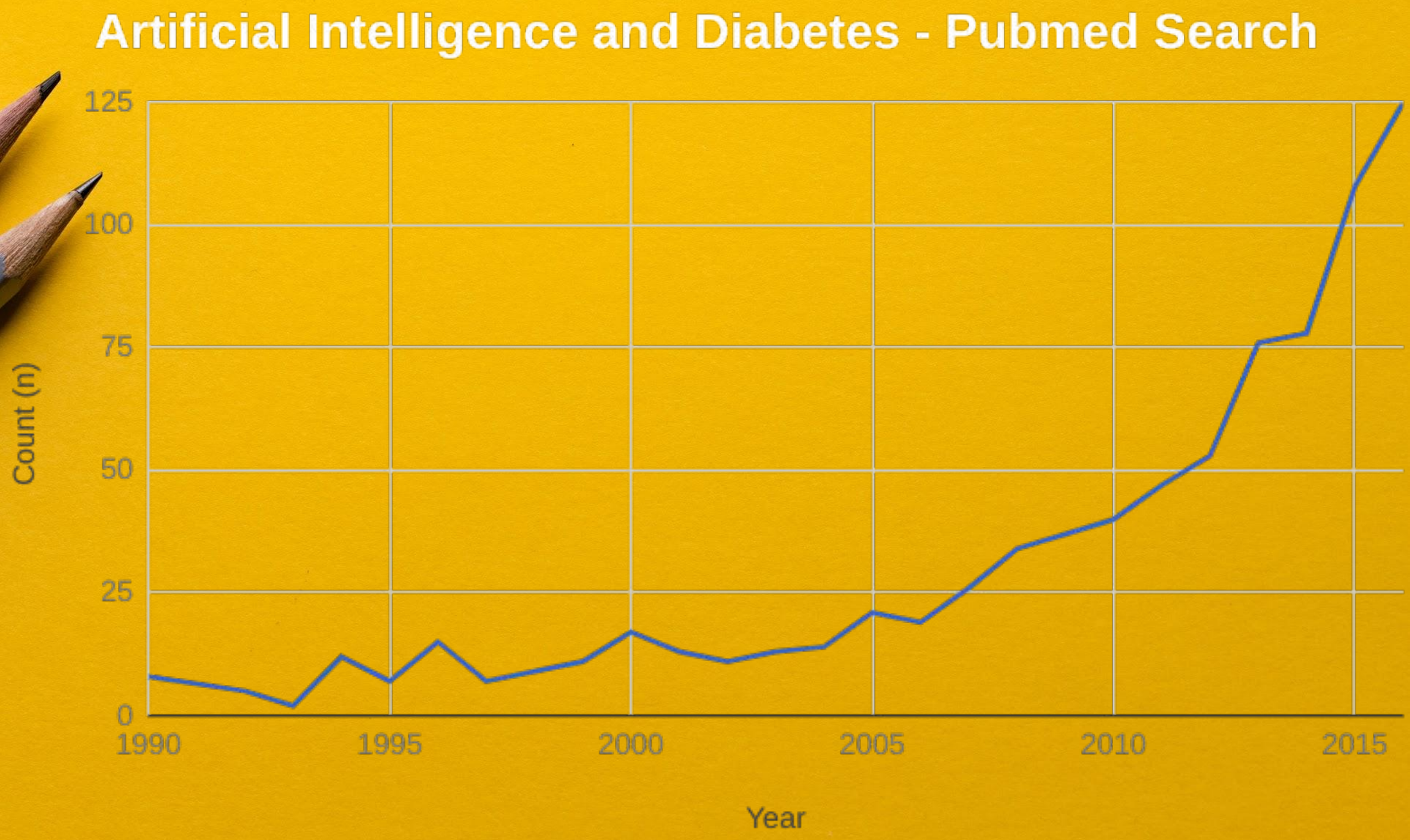
https://solidarites-sante.gouv.fr/ministere/documentation-et-publications-officielles/rapports/sante/article/rapport-health-data-hub-mission-de-prefiguration


e-EPIDEMIOLOGY
DATA
ARTIFICIAL
INTELLIGENCE

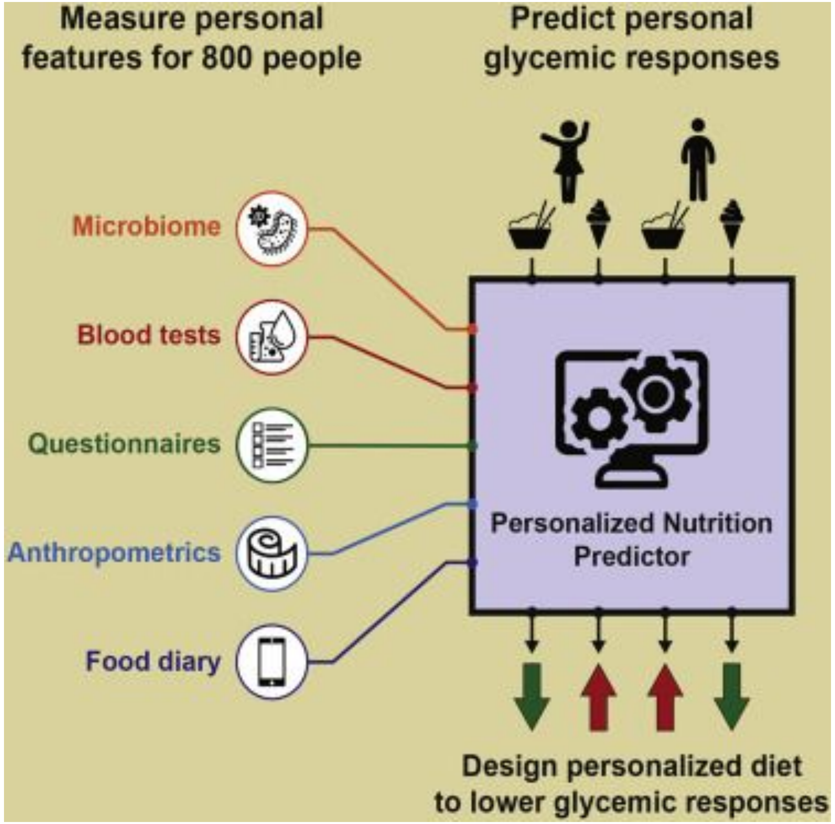

The use of AI
in epidemiology &
population health studies
New opportunities

AI

A changing landscape



Nutrition & Type 2 diabetes



"One size fits all" recommendations → Personalized recommendations

AI

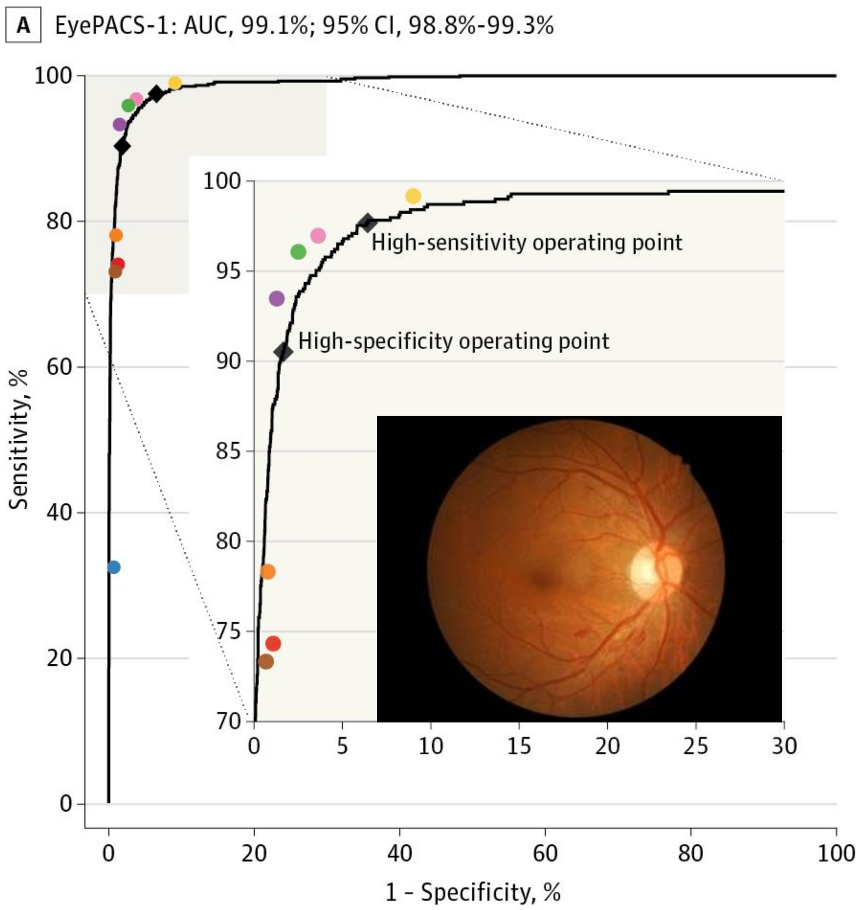
Fact: there is a large individual inter and intra-variability in the glycemic response to a given meal

Data: Continuously monitored week-long glucose levels in an 800-person cohort, measured responses to 46,898 meals

Solution: prediction of the glycemic response using **machine-learning** models based on blood parameters, dietary habits, anthropometrics, physical activity, and gut microbiota

Personalized Nutrition by Prediction of Glycemic Responses. David Zeev et al. Cell 2015

Diagnosis of diabetic retinopathy



Diagnosis by human experts → Automated diagnosis

AI

9963 retinal fundus photographs annotated by 7 experts

AUC	**Sensitivity**	**Specificity**
99,1%	**97,5%**	**93,4%**

Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. Gulshan et al. JAMA 2015

AI as a tool to do more than what the humans can do

(Almost) perfect prediction of gender from retinal fundus photographs

Predicted risk factor (evaluation metric)	UK Biobank validation dataset ($n = 12{,}026$ patients)	
	Algorithm (95% CI)	Baseline
Age: MAE, years (95% CI)	3.26 (3.22,3.31)	7.06 (6.98,7.13)
Age: R^2 (95% CI)	0.74 (0.73,0.75)	0.00
Gender: AUC (95% CI)	0.97 (0.966,0.971)	0.50

Actual: Female
Predicted: Female

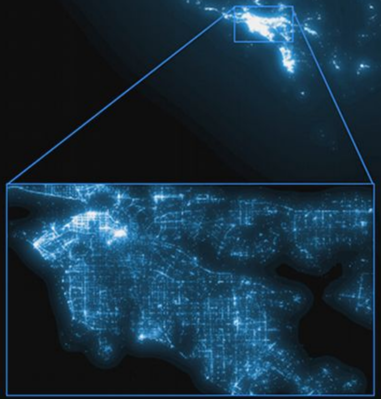
Ryan Poplin et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nature Biomedical Engineering 2018

The special case of social media data in epidemiology

AI and textual data

The digital epidemiology - You are what you Tweet

"Social media, such as Twitter, can be used to access local and timely information about **disease outbreaks** and **related events** around the world"



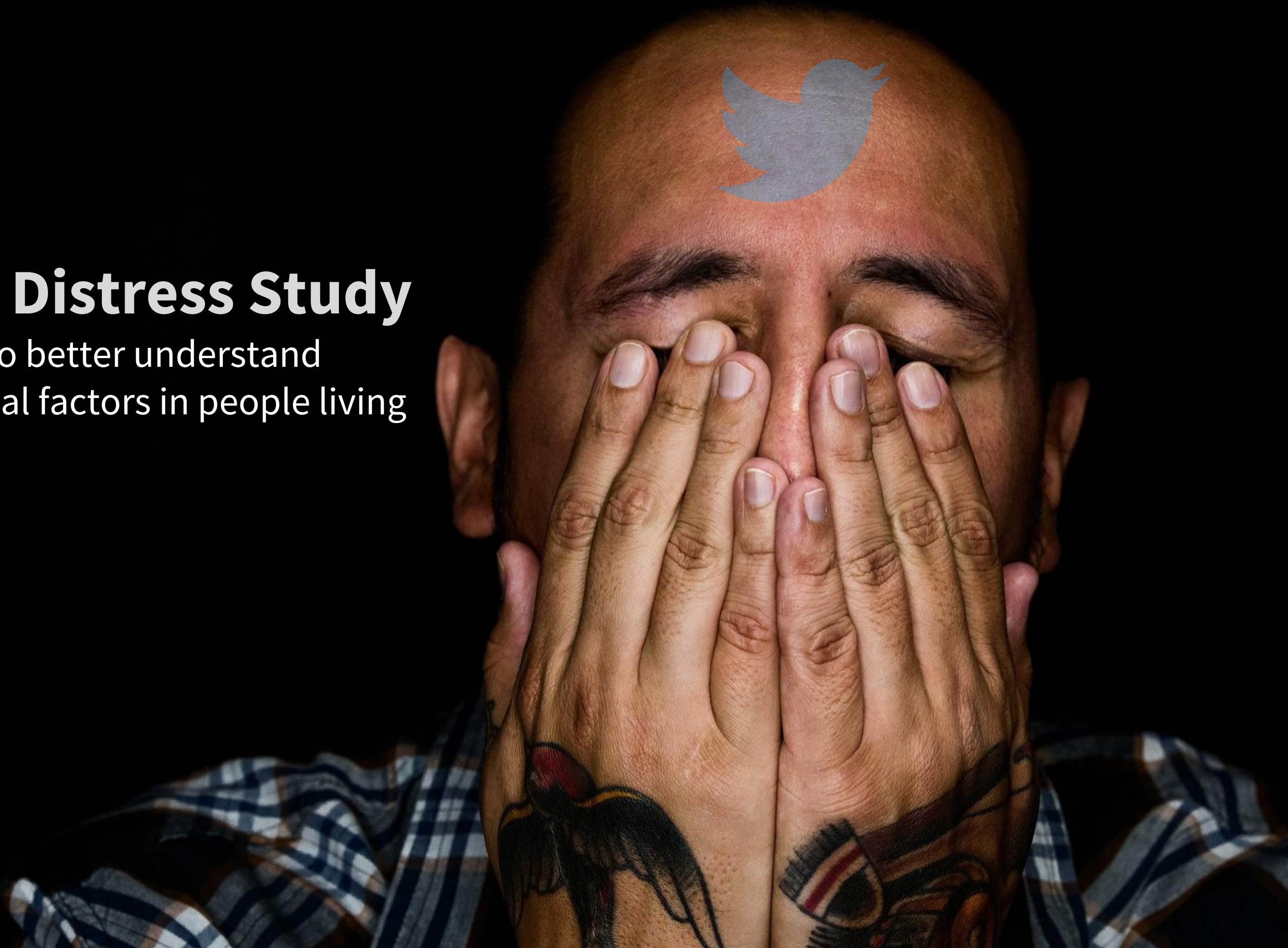
Marcel Salathé, EPFL Lausanne

One of the founding fathers

Marcel Salathé et al. Digital Epidemiology. Plos Computational Biology 2012

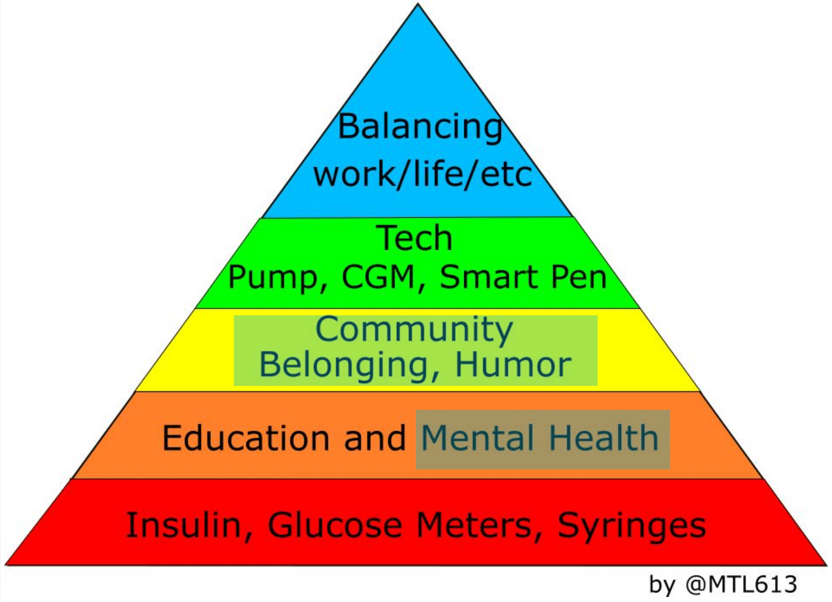
World Diabetes Distress Study

The use of Twitter and AI to better understand emotions and psychological factors in people living with diabetes



Why working on this topic?

Type 1 diabetes



Type 2 diabetes

Priority 4

How do stress and anxiety influence the management of Type 2 diabetes and does a positive mental wellbeing have an effect?

Priority 9

How can psychological or social support be best used to help people with, or at risk of Type 2 diabetes, and how should this be delivered to account for individual needs?

DIABETES UK
KNOW DIABETES. FIGHT DIABETES.

Diabetes distress

Emotional burden

Stress

Fatigue

Anxiety

Associated with day-to-day disease management

Diet
Physical activity
Fear of hypoglycemia
Fear of complications
Work
Family support
HCP relationships...

Up to 25% of adults with diabetes experienced elevated or severe diabetes distress at any given time

	Not a problem	Minor problem	Moderate problem	Somewhat serious problem	Serious problem
1. Not having clear and concrete goals for your diabetes care?	0	1	2	3	4
2. Feeling discouraged with your diabetes treatment plan?	0	1	2	3	4
3. Feeling scared when you think about living with diabetes?	0	1	2	3	4
4. Uncomfortable social situations related to your diabetes care (e.g., people telling you what to eat)?	0	1	2	3	4
5. Feelings of deprivation regarding food and meals?	0	1	2	3	4
6. Feeling depressed when you think about living with diabetes?	0	1	2	3	4
7. Not knowing if your mood or feelings are related to your diabetes?	0	1	2	3	4
8. Feeling overwhelmed by your diabetes?	0	1	2	3	4
9. Worrying about low blood sugar reactions?	0	1	2	3	4
10. Feeling angry when you think about living with diabetes?	0	1	2	3	4
11. Feeling constantly concerned about food and eating?	0	1	2	3	4
12. Worrying about the future and the possibility of serious complications?	0	1	2	3	4
13. Feelings of guilt or anxiety when you get off track with your diabetes management?	0	1	2	3	4
14. Not "accepting" your diabetes?	0	1	2	3	4
15. Feeling unsatisfied with your diabetes physician?	0	1	2	3	4
16. Feeling that diabetes is taking up too much of your mental and physical energy every day?	0	1	2	3	4
17. Feeling alone with your diabetes?	0	1	2	3	4
18. Feeling that your friends and family are not supportive of your diabetes management efforts?	0	1	2	3	4
19. Coping with complications of diabetes?	0	1	2	3	4
20. Feeling "burned out" by the constant effort needed to manage diabetes?	0	1	2	3	4

PAID Problem Areas In Diabetes)

Limitations

- Self-reported
- Non evolutive
- Some components of DD are missing (work-related issues, cost of treatment, HCP relationships...)
- Interpreted with a HCP
- “Make my doctor happy” effect
- Risk of denial and bias

Change of paradigm

Let’s analyze what people with diabetes are saying online.

All over the world.

- “Let’s use Twitter!”
- “But… why Twitter?”
- “Well...because Twitter!”

Key figures

- 313 millions/months of active users
- > 4,7 millions of people are actually tweeting
- <u>Data coming from the entire world</u>
- <u>Public data by default</u>

Languages on Twitter

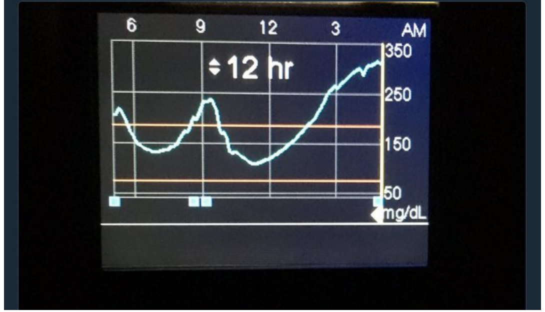
- English (34%)
- Japanese (16%)
- Spanish (12%)
- Malay (8%)
- Portuguese(6%)
- Arab (6%)
- French(2%)

Diabetes Online Community on Twitter

Data unavailable in a traditional clinical setting

Kate
@aDoseOfDiabetes
26 years old. type 1 diabetic for 25 years. I like pretty things and nice people.
California, USA

Kate
@aDoseOfDiabetes
Went to bed at 10....not sure why this keeps happening but I hate waking up at 4 feeling so sick
À l'origine en anglais



MeaT @TimeaTheTranny
Il y a 13 mins
Everytime i feel a minor ache in my foot i assume it's because of diabetes.

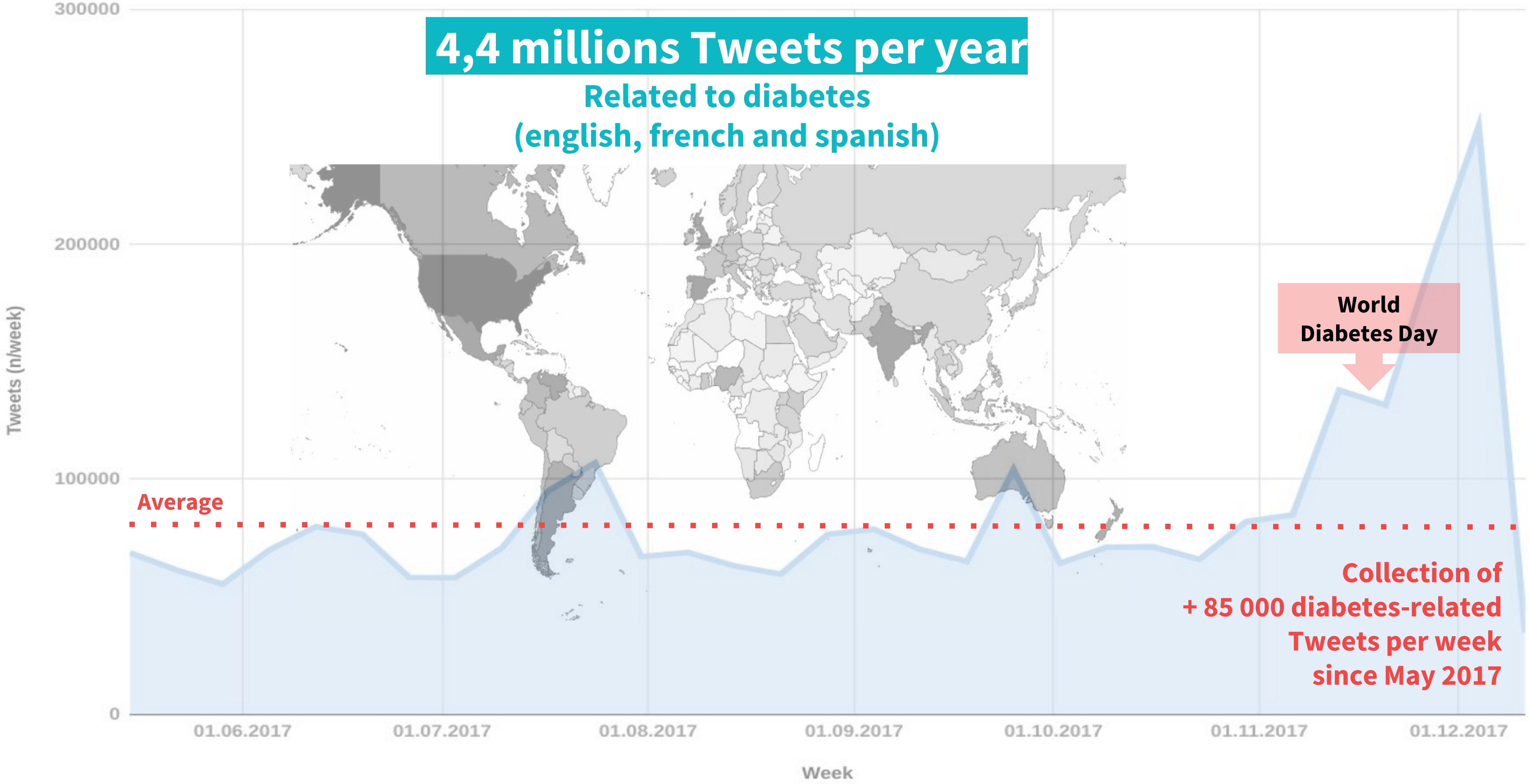
Type1bri @type1bri · 18 févr.
My latest floury with the libre to check basal and ratios has been a success, no changes needed atm until next time #gbdoc #doc #ourd
À l'origine en anglais

Tasha @tasha__louise
Il y a 15 mins
Just because you can't SEE my conditions, does not mean they're not there #diabetes #mentalhealth #longstandingillness

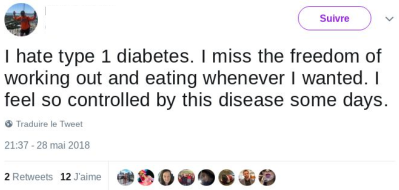
JllyJllyFish Retweeté
EnvyDaTropic™ @envydatropic
il y a 3 jours
If it doesn't clog arteries or cause diabetes I'm not eating it
152 203

Pat Atarte @UnSdfEnCorse · 2 min
Tant d'années de combat contre les glycémies à stabiliser, un sentiment de fatigue apparaît souvent, de la lassitude à être tout le temps aux aguets, à être différent des autres…Même si cela ne se voit pas ! Je suis malade.. Et sérieusement.. #diabete

Data collection



Methodology



Tweets

Personnal

~~Institutional~~

Text analysis

Emojis / emoticons 😄 :-) 😭

Phonology ([ʀ], [d], [t]...)

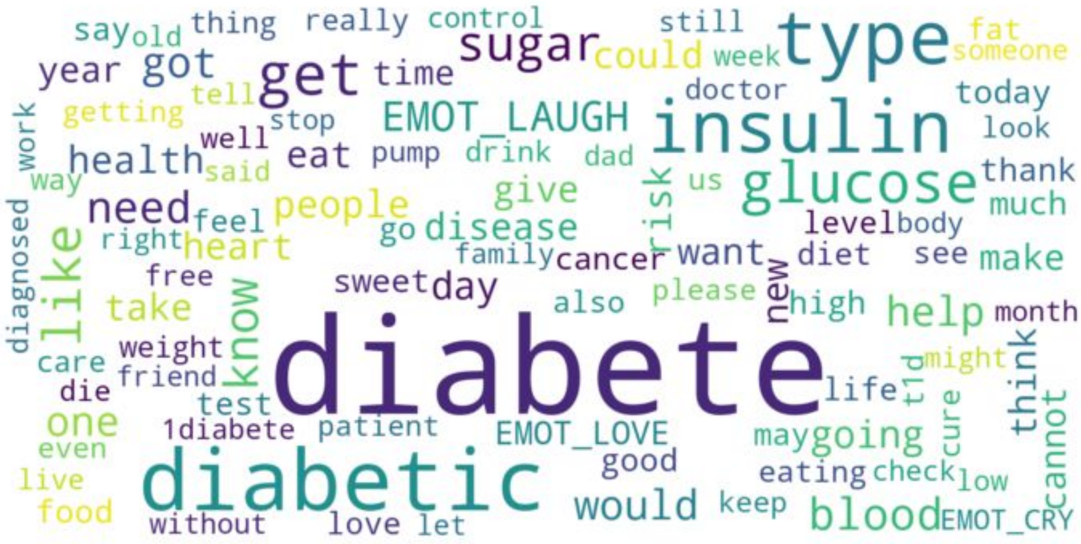
Topic modeling (Latent Dirichlet Allocation)

Sentiment analysis

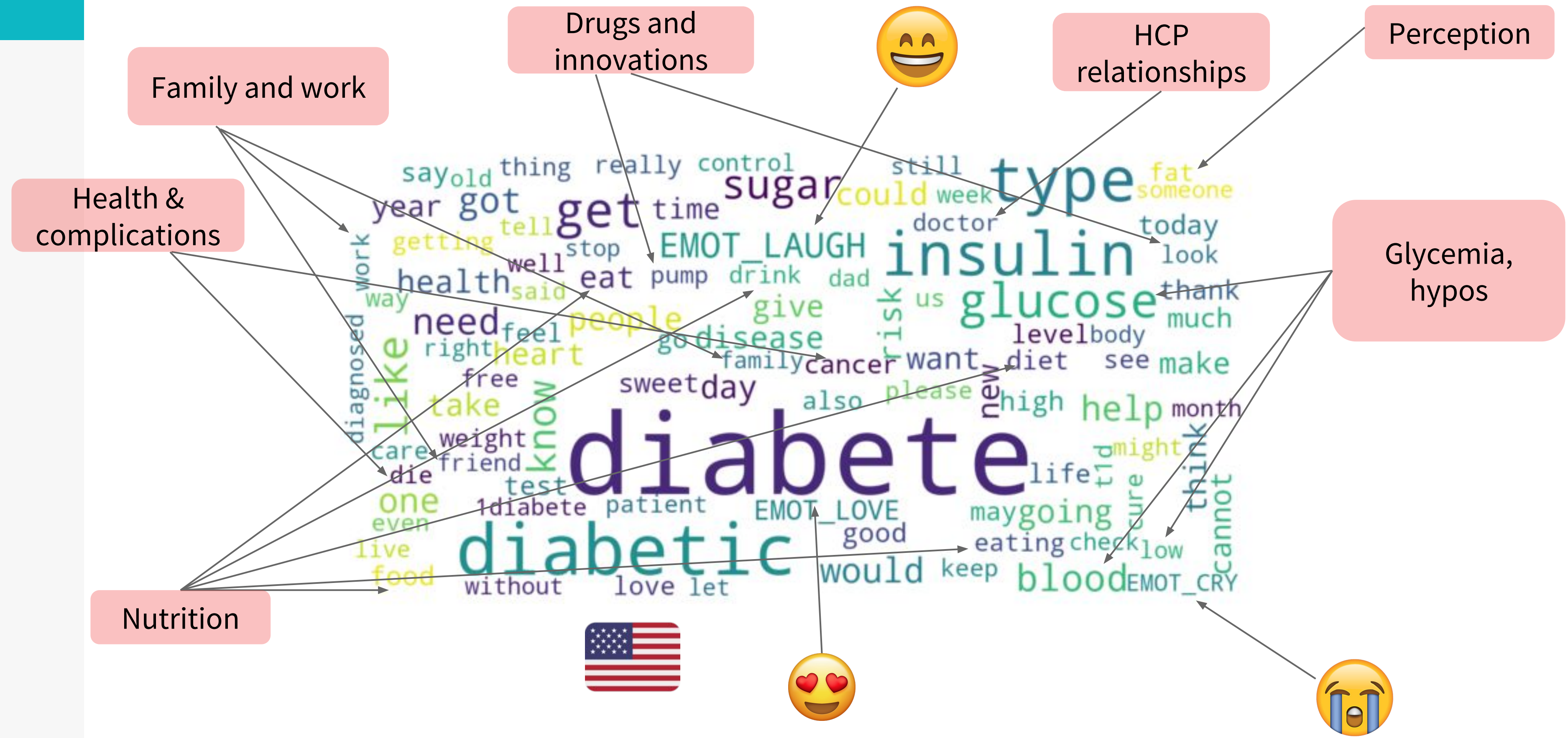
Clustering of diabetes distress patterns (AI, NLP)

Spatial analysis

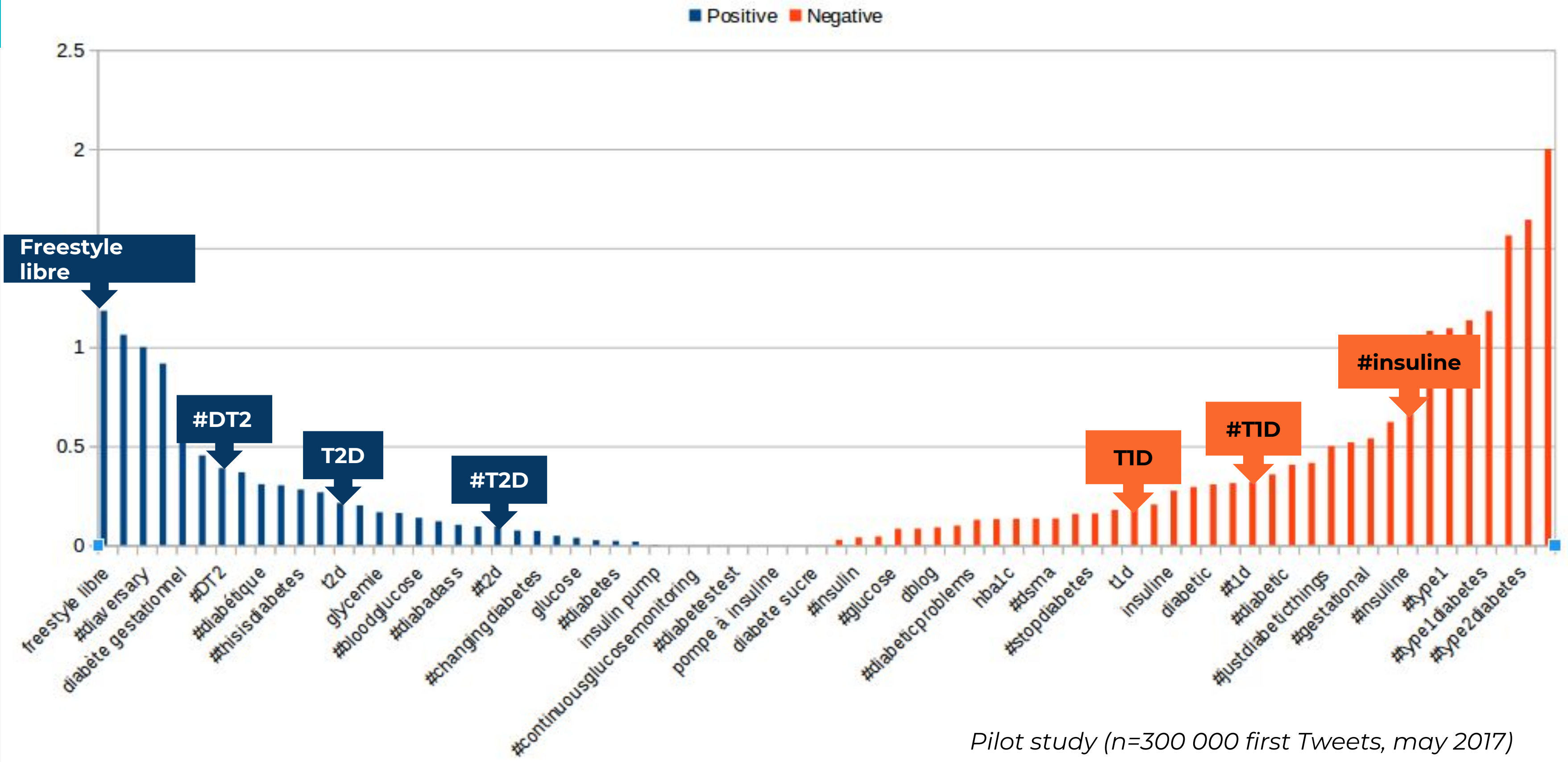
Feelings and hot topics about diabetes all around the world



Feelings and hot topics about diabetes all around the world



Sentiment analysis on frequent hashtags



Pilot study (n=300 000 first Tweets, may 2017)

Diabetes Distress Patterns and related emotions

- **Classification of emotions** by Parrott
- Used diabetes-related Tweets with emotions (N=129 313)
- Gather all the synonyms of these emtions
- Clustering method (**LDA**) -> **10 clusters**

Primary emotion	Secondary emotion	Tertiary emotions
Love	Affection	Adoration, affection, love, fondness, liking, attraction, caring, tenderness, compassion, sentimentality
	Lust	Arousal, desire, lust, passion, infatuation
	Longing	Longing
Joy	Cheerfulness	Amusement, bliss, cheerfulness, gaiety, glee, jolliness, joviality, joy, delight, enjoyment, gladness, happiness, jubilation, elation, satisfaction, ecstasy, euphoria
	Zest	Enthusiasm, zeal, zest, excitement, thrill, exhilaration
	Contentment	Contentment, pleasure
	Pride	Pride, triumph
	Optimism	Eagerness, hope, optimism
	Enthrallment	Enthrallment, rapture
	Relief	Relief
Surprise	Surprise	Amazement, surprise, astonishment
Anger	Irritation	Aggravation, irritation, agitation, annoyance, grouchiness, grumpiness
	Exasperation	Exasperation, frustration
	Rage	Anger, rage, outrage, fury, wrath, hostility, ferocity, bitterness, hate, loathing, scorn, spite, vengefulness, dislike, resentment
	Disgust	Disgust, revulsion, contempt
	Envy	Envy, jealousy
	Torment	Torment
Sadness	Suffering	Agony, suffering, hurt, anguish
	Sadness	Depression, despair, hopelessness, gloom, glumness, sadness, unhappiness, grief, sorrow, woe, misery, melancholy
	Disappointment	Dismay, disappointment, displeasure
	Shame	Guilt, shame, regret, remorse
	Neglect	Alienation, isolation, neglect, loneliness, rejection, homesickness, defeat, dejection, insecurity, embarrassment, humiliation, insult
	Sympathy	Pity, sympathy
Fear	Horror	Alarm, shock, fear, fright, horror, terror, panic, hysteria, mortification
	Nervousness	Anxiety, nervousness, tenseness, uneasiness, apprehension, worry, distress, dread

Diabetes Distress Patterns and related emotions

		Frequent words in each cluster
5 clusters related to Diabetes Distress	Frustration with glucose tests in particular during pregnancy	EMOT_JOY, EMOT_SADNESS, glucose, test, drink, guardian, EMOT_FEAR, hour, three, wait
	Burden of diabetes and diabetes treatment, including insulin and blood glucose control	insulin, sugar, day, blood, time, today, long, work, use, pump
	Disappointment about common misunderstanding and confusion around the types of diabetes and miraculous tips to cure diabetes	get, go, love, would, eat, one, make, like, take, think
	Nutrition-related feeling of guiltiness	sweet, give, might, tea, chocol, cake, candi, tast, buy, coma
	Fears related to poor health coverage	care, heart, high, cancer, diseas, suffer, health, pain, caus, medic
1 cluster related to WDD	World diabetes day	EMOT_LOVE, happi, thank, dsma, world, birthday, awar, fight, share, celebr
4 clusters non related to diabetes	Excluded	

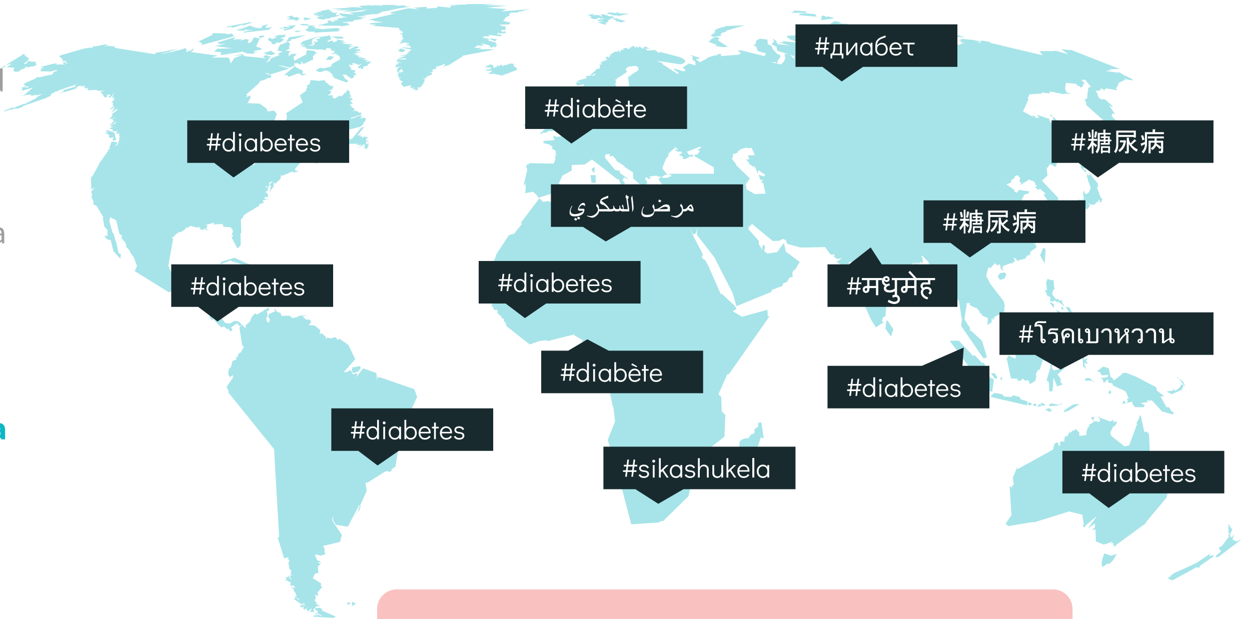
Worldwide e-cohort study

e-cohort

- Targeted online recruitment
- Creation of an independent and secured web platform
- Prospective data collection on lifestyle and psychological factors and health events with a **chatbot**

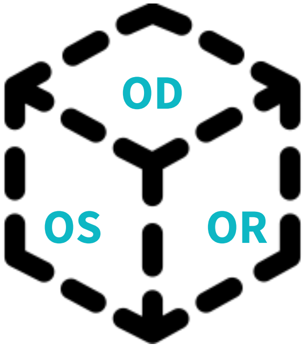
Sync connected devices

Match with third-party open databases based on geolocation data



Open Source/Open Data project

Encourage open medical research



Open Data

Datasets

Open Source

Tweets-extractor engine

Chatbot

Algorithms (NLP, text mining, deep learning)

Methodology

Open Research

Creation of a virtual community to work on diabetes

Increase transparency & reproducible research

Open access publications

Better understand diabetes in the "real world"

Improve diabetes awareness

Detect weak signals in big data

Generate new research questions

In short

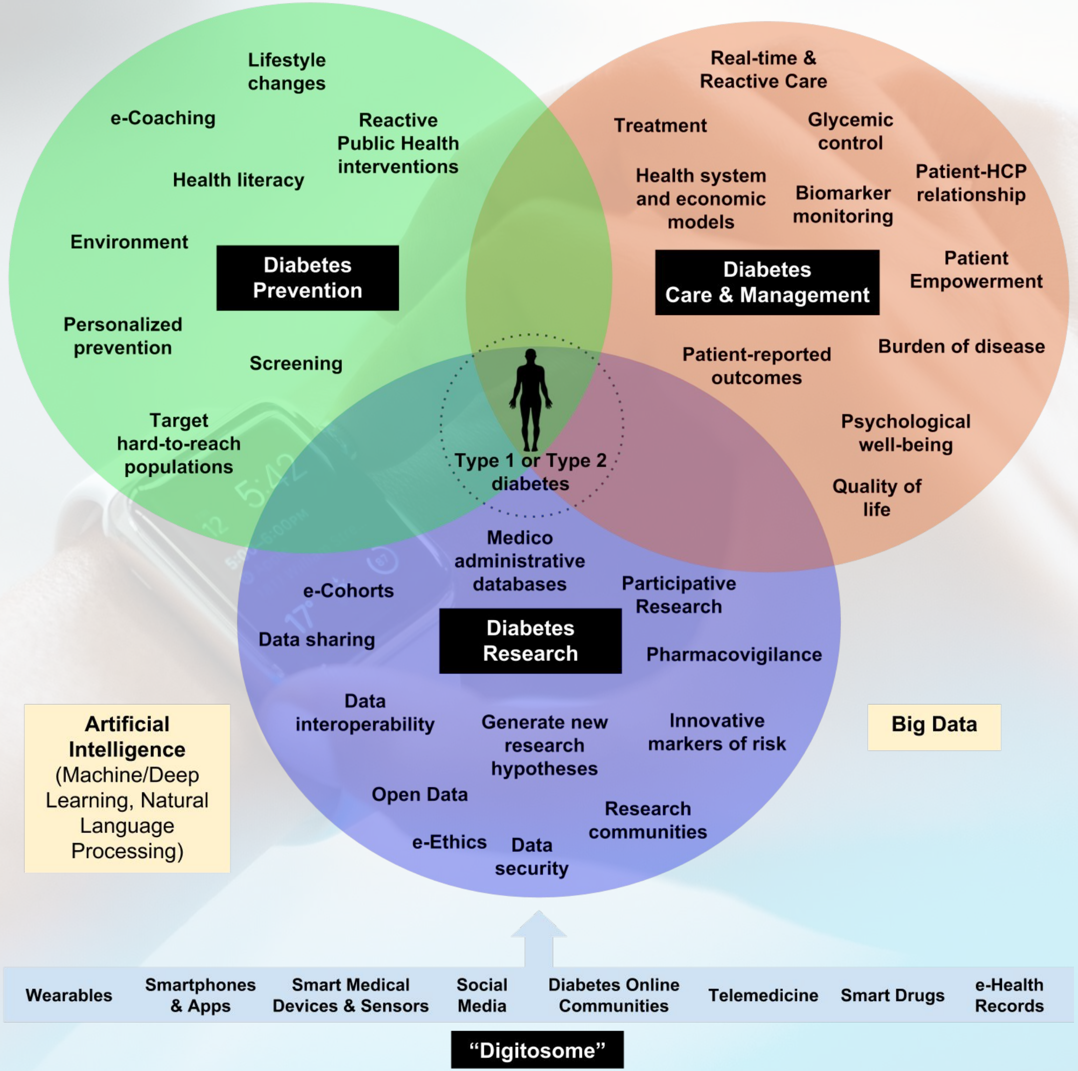
A worldwide project on diabetes

An unprecedented resource

At the cutting edge of technology

And now what ?

Impact of AI and digitosome on chronic diseases



Digital diabetes: perspectives for diabetes prevention, management and research.
Fagherazzi et al. (Diabetes & Metabolism, 2018) - Open Access

Let’s do better!

Need of dataviz tools and user-friendly solutions

Key for patients and HCP to use innovative solutions

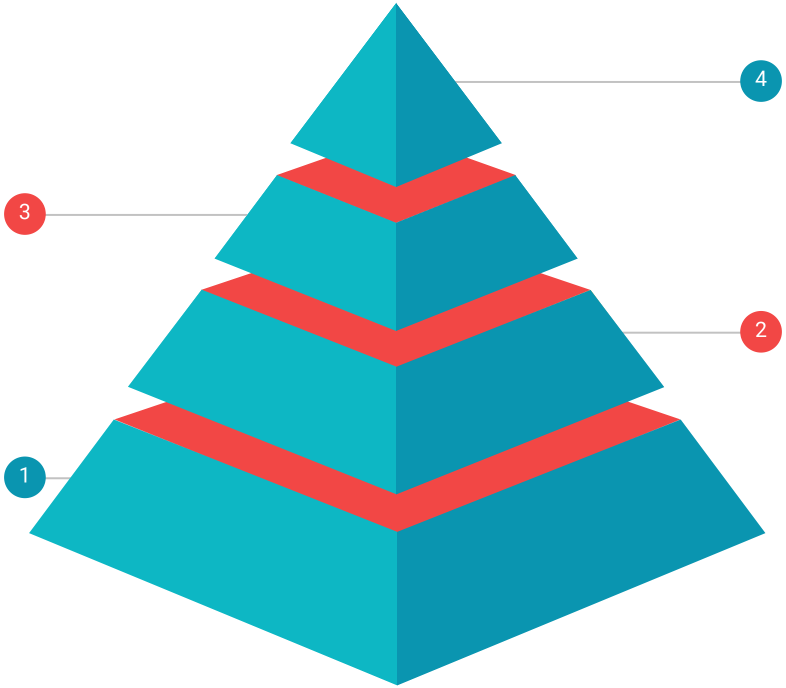
Do not increase the digital barrier and socioeconomic inequalities

Improve **UX/UI** to optimize long-term use

Patients/Participants available for research are a scarce resource

Unlikely that a given individual participates to several long term initiatives simultaneously

But we need large population sizes + long follow-up + variability



Promote Open Data/Open Source

Share algorithms

Definition of **new business models**, for both academics and private sectors (from based on **proprietary data** to based on **services and know-how**)

Ensure **reproducibility**

Be careful with isolated projects

Do not contribute to the creation of **data in silo with poor interoperability**

Include participants/patients from scratch

My message for future epi or PH experts

Be at the interface!

Actors

- Governments / Agencies
- Academics
- Hospitals
- Pharmas
- GAFAM / BATX
- Startups
- Patients / Citizens

Disciplines

- Medicine
- Public Health
- Technology
- Data Science / AI
- Medical informatics
- Psychology
- Social Science

Modern epidemiology

Alternate & combine

Data driven approaches with **Hypothesis driven** research

Soft skills

+++

Communication
Teamwork
Adaptability
Problem-Solving
Creativity
Work Ethic
Interpersonal Skills
Time Management
Leadership
Attention to Detail

MERCI ;-)

Guy Fagherazzi

guy.fagherazzi@gmail.com

Twitter : @GFaghe