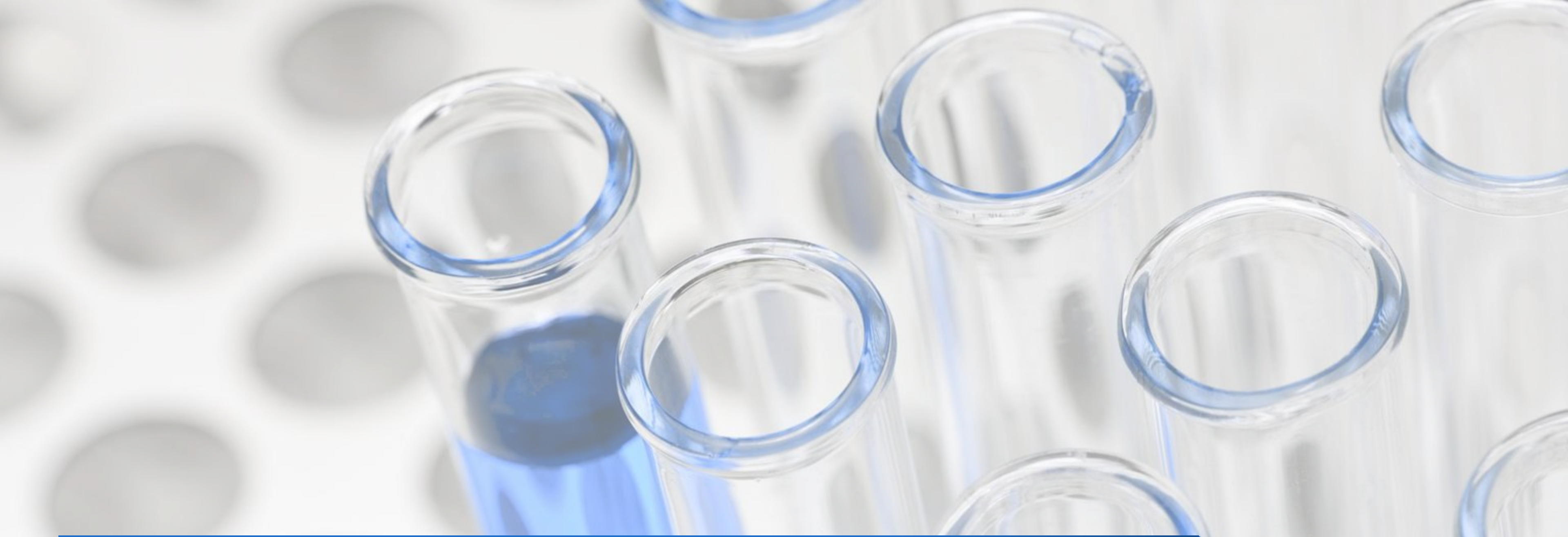


Machine learning for disease modelling

P. Bauvin, PhD
U1190
17/12/2021

Introduction

“Modelling”, “machine learning”, “big data”

Machine learning in health : What for ?

1. Epidemiology in the Big Data Era

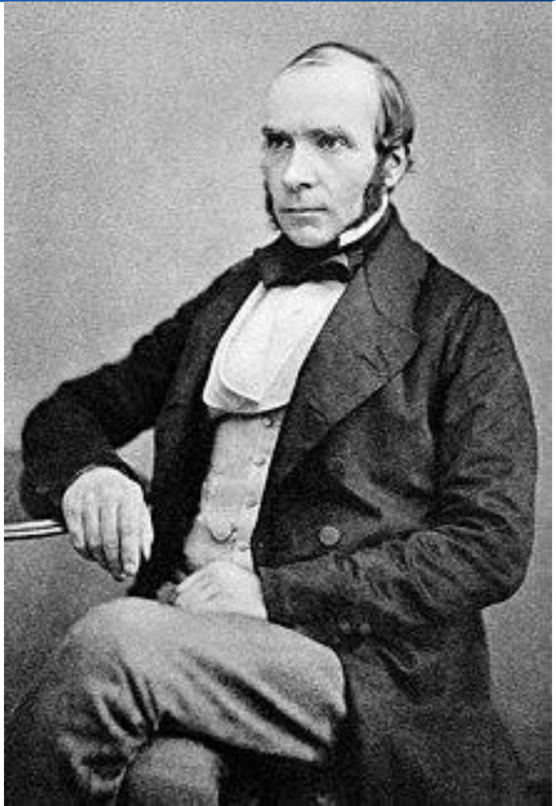
2. Diagnostic and Prognostic models

Machine learning for epidemiology in the Big Data Era

1

Who is your John Snow?

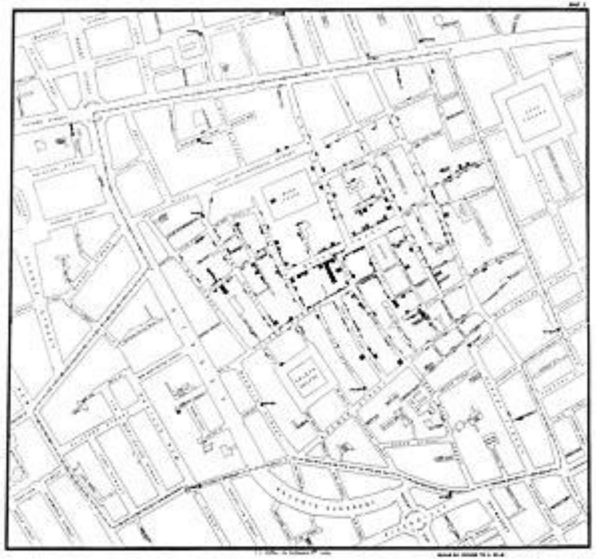
Kit Harrington, playing Jon Snow in Game of Thrones



John Snow, English physician (1813-1858)

John Snow and the Broad Street Pump

- Cholera outbreak in London, 1854
- John Snow investigated by reporting cholera cases
- Identified contaminated water from a water pump: the Broad Street pump



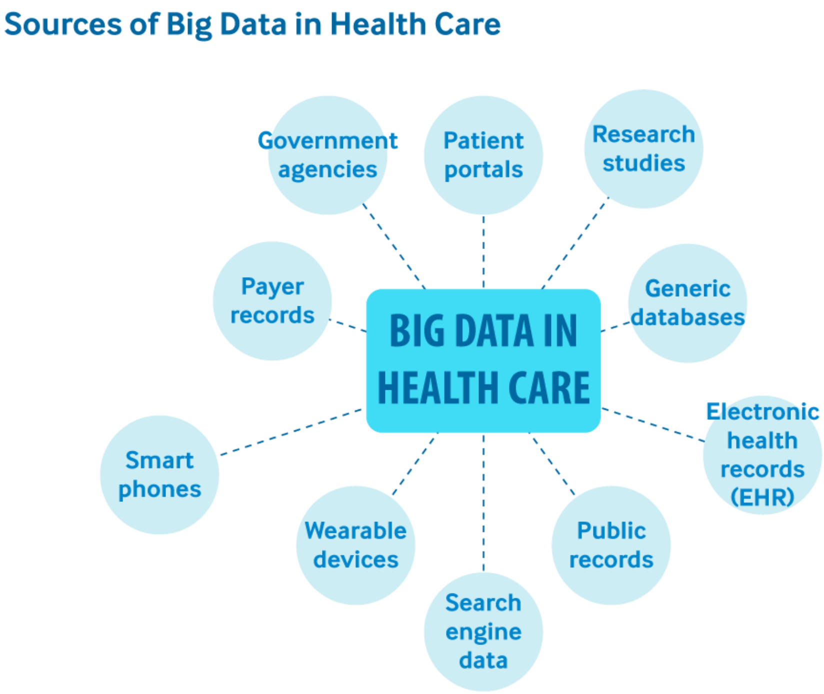
Original map by John Snow showing the clusters of cholera cases in the London epidemic of 1854

John Snow and the Broad Street Pump

- Cholera outbreak in London, 1854
- John Snow investigated by reporting cholera cases
- Identified contaminated water from a water pump: the Broad Street pump
- The removal of the handle led to cessation of the epidemic

The Broad Street pump, Soho , London

From John Snow to modern reporting



NEJM Catalyst (catalyst.nejm.org) © Massachusetts Medical Society

A first example: obesity in France

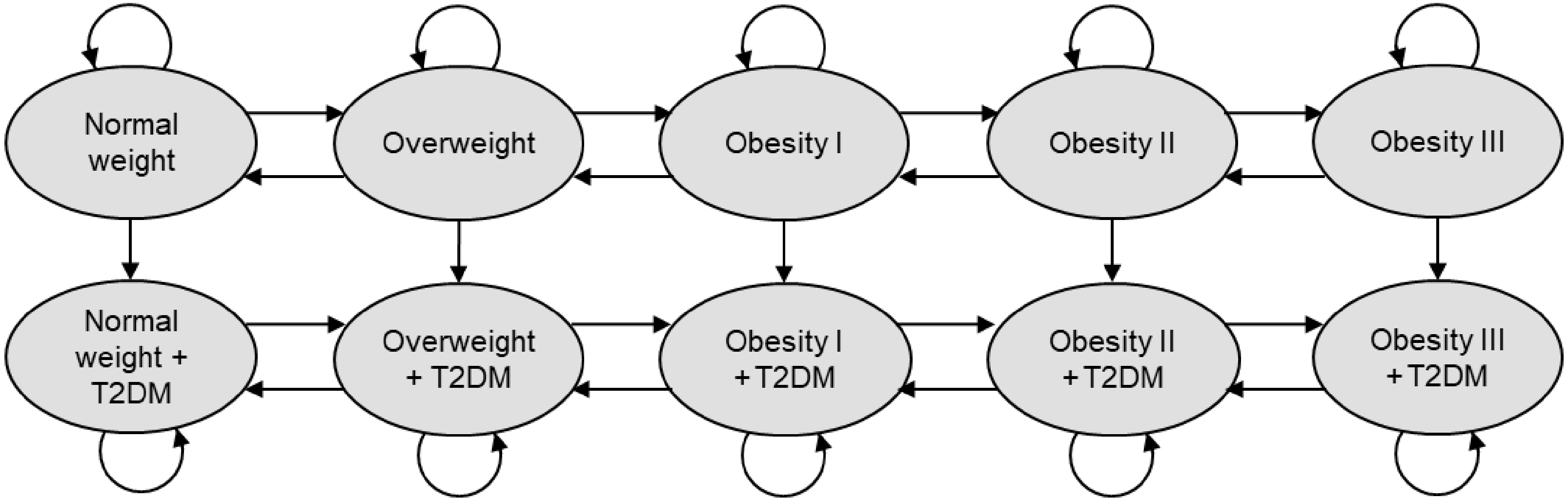
- What are the dynamics of overweight, obesity and diabetes in France?
- Previous statistical studies estimating **prevalence** evolution
- We want to investigate the evolution of **incidence**

Machine learning approach

Why does incidence matters?

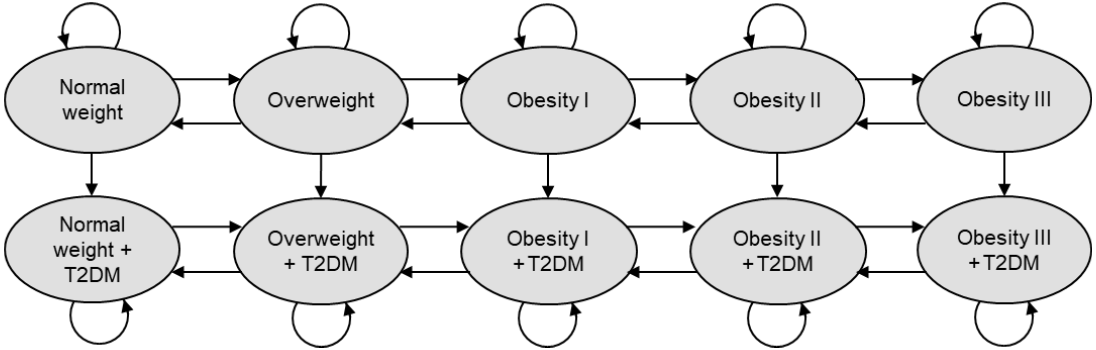
- PNNS: Programme National de Nutrition Santé 2019-2023
- Goal: Prevalence of obesity reduction by 15% in 5 years
- How does it translates into new cases allowed per year, ie incidence?

Model of overweight, obesity and type 2 diabetes

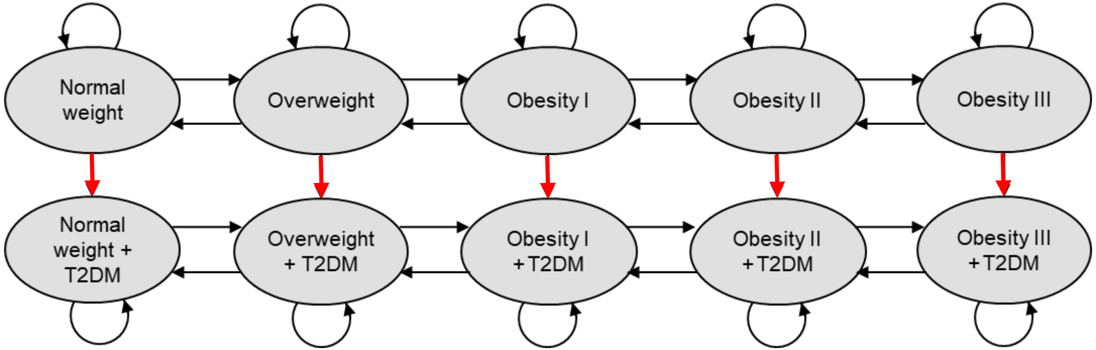


Parameter estimation

- Using 1998 to 2014 surveys (ESPS)
- 97,948 respondents
- Age, sex and calendar year impacting transition rates



Some brief results

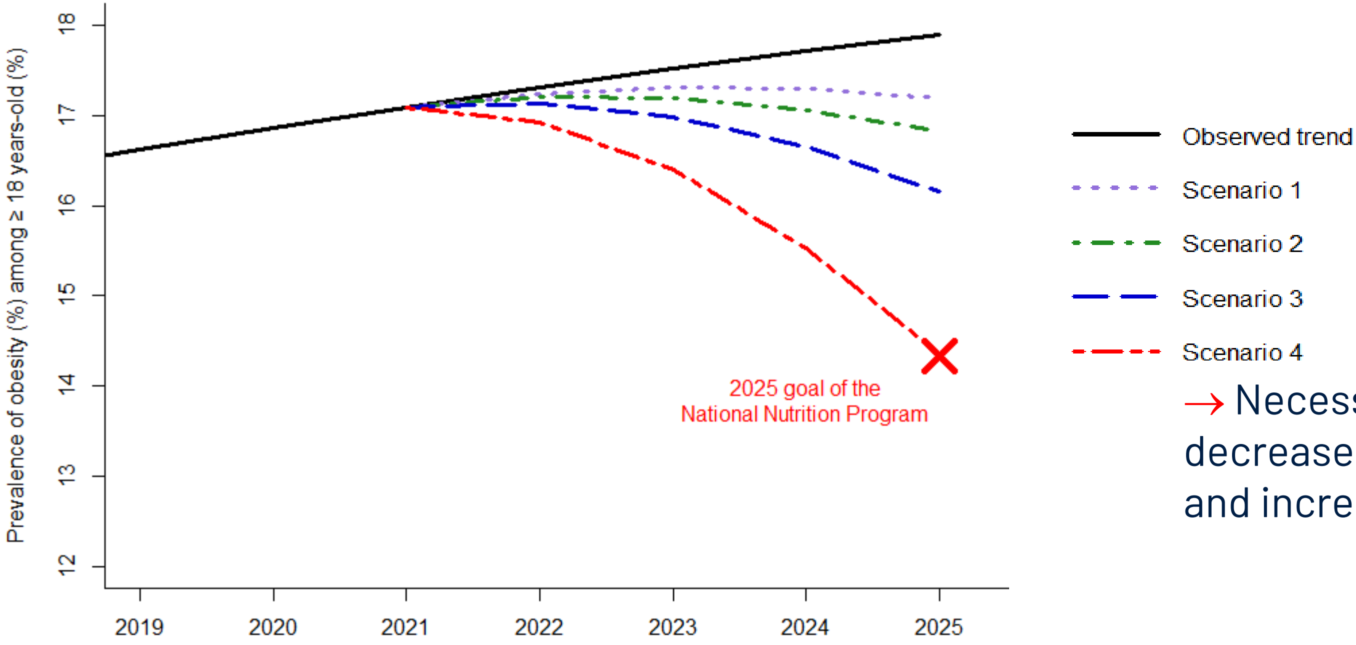


Some brief results

- Yearly risk of type 2 diabetes, 30 y-o man in 2020s:
  - 0,035% with normal weight
  - 0,26% with overweight
  - 0,56% with class I obesity
  - 1,46% with class II or III obesity
- 48% higher risk to develop diabetes in the 2010s than in the 1980s within the same BMI class

What about the PNNS goals?

What about the PNNS goals?



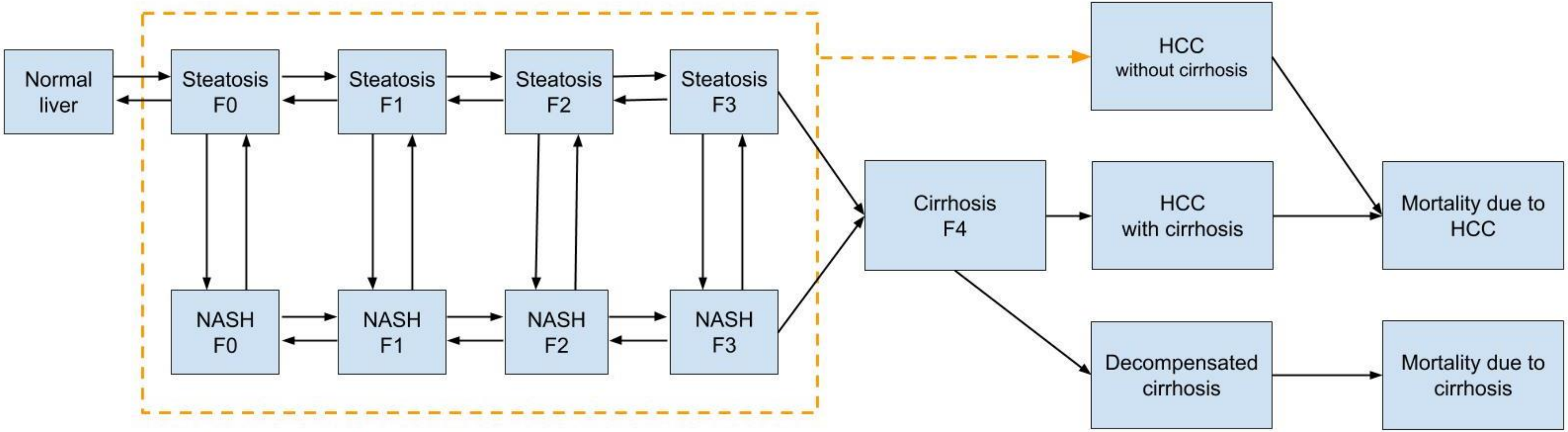
→ Necessitates a **63%** decrease in weight gain, and increase in weight loss

A second example: NAFLD (with bigger data!)

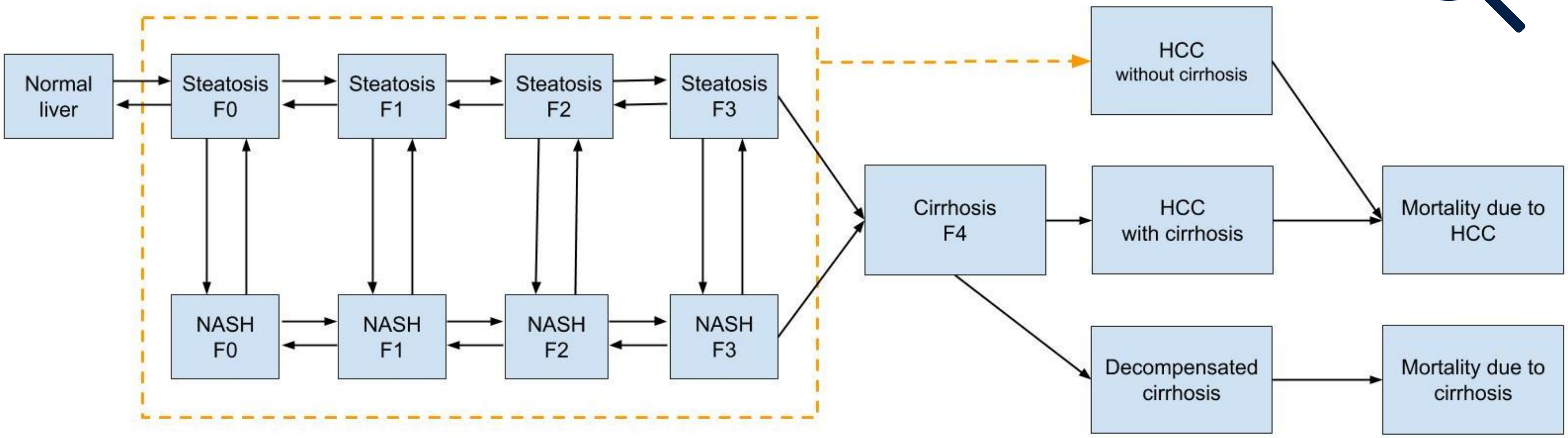
- NALFD: Non-alcoholic Fatty Liver Disease
- Spectrum of NAFLD: isolated steatosis and NASH
- Silent disease related to obesity and metabolic syndrome, asymptomatic until serious events (HCC, cirrhosis)

Le Parisien, 13/04/2018

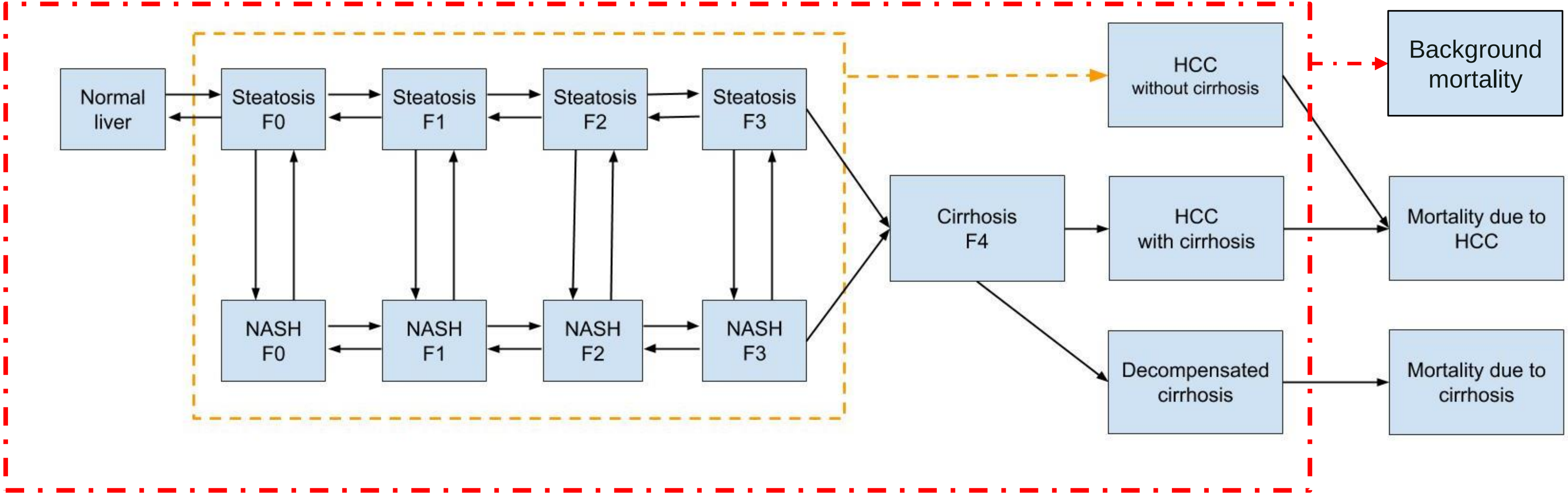
Model of NAFLD progression



Model of NAFLD progression



Model of NAFLD progression



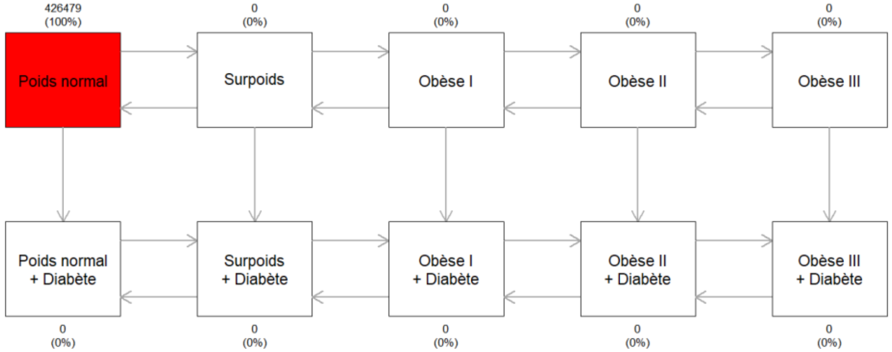
A more complex model

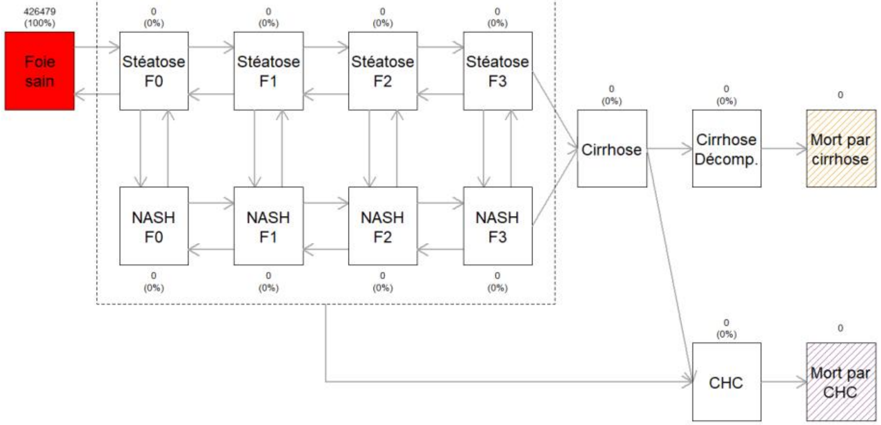
- A huge number of parameters
- A lot of disease states, some never observed

→ **Use of the PMSI database:**

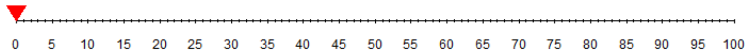
- All patients with liver disease identified
- 389 453 patients
- 27 306 341 hospital stays

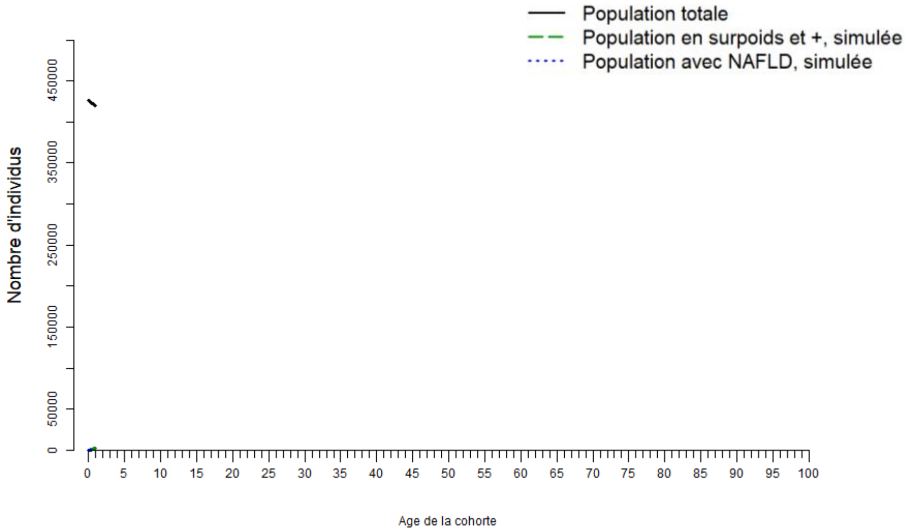
Programme de Médicalisation des Systèmes d'Information

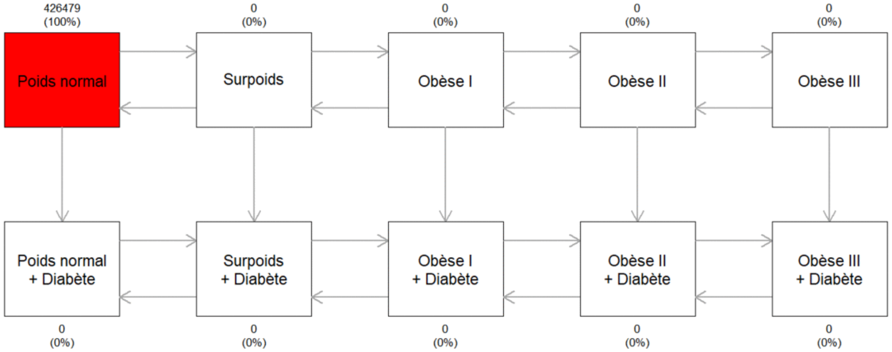


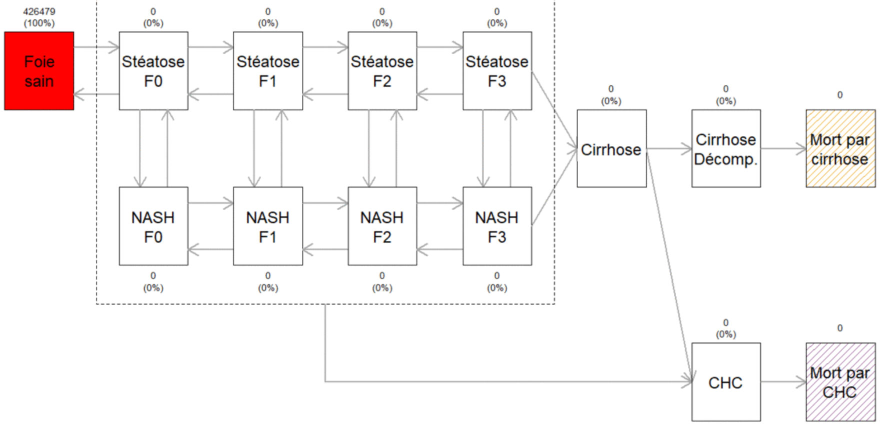


Cohorte née en 1950
Âge actuel :

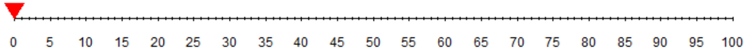


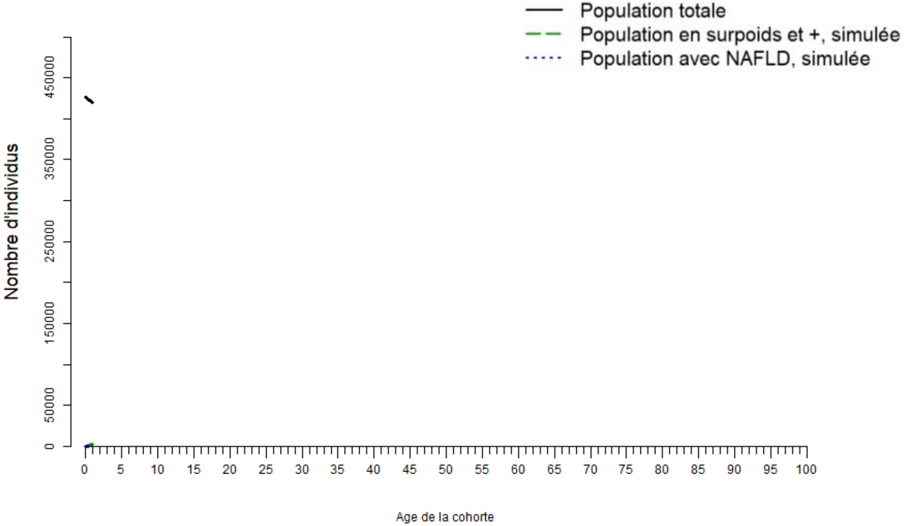


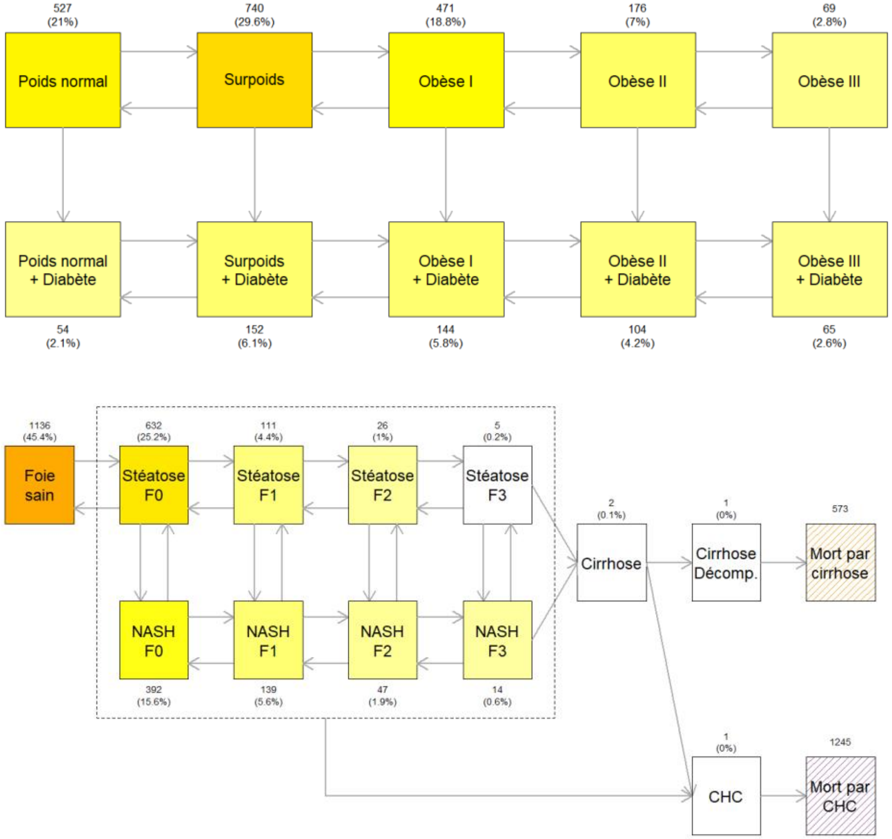




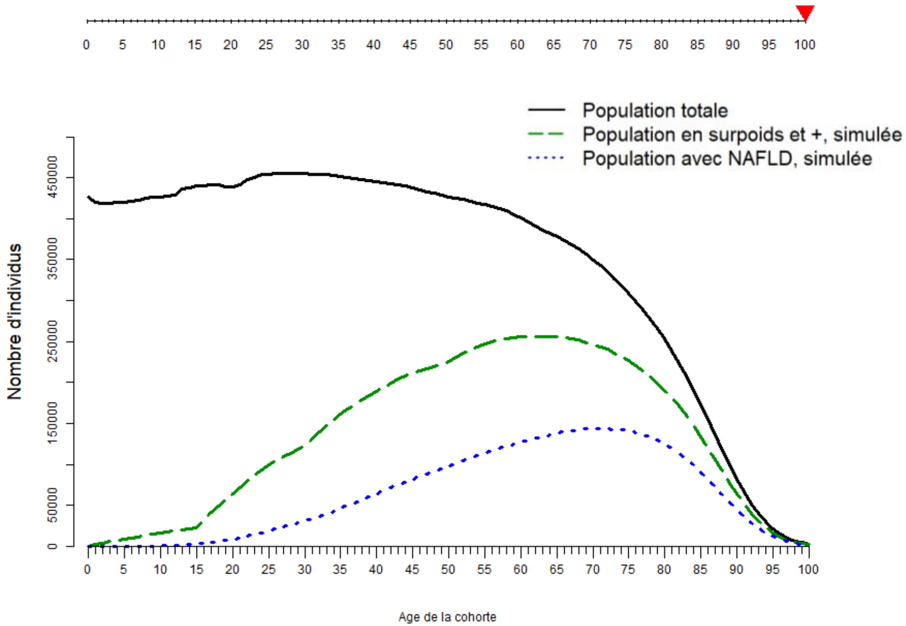
Cohorte née en 1950
Âge actuel :







Cohorte née en 1950
Âge actuel :



Model of NAFLD progression

- Predict NAFLD progression nationwide
- Forecast actual and future burden related to NAFLD
- Evaluate the impact of
  - a medicine targetting fibrosis, NASH,...
  - a public policy targetting overweight, obesity,...

Machine learning for diagnostic and prognostic models

2

Machine learning is everywhere

Development and validation of the automated imaging differentiation in parkinsonism (AID-P): a multicentre machine learning study

CrossMark

oa

Derek B Archer, Justin T Bricker, Winston T Chu, Roxana G Burciu, Johanna L McCracken, Song Lai, Stephen A Coombes, Ruogu Fang, Angelos Barmpoutis, Daniel M Corcos, Ajay S Kurani, Trina Mitchell, Mieniecia L Black, Ellen Herschel, Tanya Simuni, Todd B Parrish, Cynthia Comella, Tao Xie, Klaus Seppi, Nicolaas I Bohnen, Martijn LTM Müller, Roger L Albin, Florian Krismer, Guangwei Du, Mechelle M Lewis, Xuemei Huang, Hong Li, Ofer Pasternak, Nikolaus R McFarland, Michael S Okun, David E Vaillancourt

Summary
Background Development of valid, non-invasive biomarkers for parkinsonian syndromes is crucially needed. We

Lancet Digital Health 2019

JACC: CARDIOVASCULAR INTERVENTIONS
© 2019 PUBLISHED BY ELSEVIER ON BEHALF OF THE
AMERICAN COLLEGE OF CARDIOLOGY FOUNDATION

VOL. 12, NO. 14, 2019

FOCUS ON ARTIFICIAL INTELLIGENCE

Leveraging Machine Learning Techniques to Forecast Patient Prognosis After Percutaneous Coronary Intervention

Chad J. Zack, MD, MS,[a,*] Conor Senecal, MD,[b,*] Yaron Kinar, PhD,[c] Yaakov Metzger, MD, PhD,[c] Yoav Bar-Sinai, MS,[c] R. Jay Widmer, MD, PhD,[b] Ryan Lennon, MS,[d] Mandeep Singh, MD, MPH,[b] Malcolm R. Bell, MD,[b] Amir Lerman, MD,[b] Rajiv Gulati, MD, PhD[b]

JAMA Network | Open

Original Investigation | Substance Use and Addiction

Identifying Smoking Environments From Images of Daily Life With Deep Learning

Matthew M. Engelhard, MD, PhD; Jason A. Oliver, PhD; Ricardo Henao, PhD; Matt Hallyburton, BA; Lawrence E. Carin, PhD; Cynthia Conklin, PhD; F. Joseph McClernon, PhD

npj | Digital Medicine

www.nature.com/npjdigitalmed

ARTICLE OPEN

Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices

Michael D. Abràmoff[1,2,3,4], Philip T. Lavin[5], Michele Birch[6], Nilay Shah[7] and James C. Folk[1,2,3]

Artificial Intelligence (AI) has long promised to increase healthcare affordability, quality and accessibility but FDA, until recently, had never authorized an autonomous AI diagnostic system. This pivotal trial of an AI system to detect diabetic retinopathy (DR) in people with diabetes enrolled 900 subjects, with no history of DR at primary care clinics, by comparing to Wisconsin Fundus Photograph Reading Center (FPRC) widefield stereoscopic photography and macular Optical Coherence Tomography (OCT), by FPRC certified photographers, and FPRC grading of Early Treatment Diabetic Retinopathy Study Severity Scale (ETDRS) and Diabetic

Promises of machine learning

Doctors about to be replaced by hospital AI systems offering better diagnosis and less arrogance

Tuesday, September 12, 2017 by: Jhoanna Robinson
Tags: artificial intelligence, doctors, healthcare, Hospitals, patient treatment, physicians, robots

17K
VIEWS

Promises of machine learning

Time to reality check the promises of machine learning-powered precision medicine

Jack Wilkinson, Kellyn F Arnold, Eleanor J Murray, Maarten van Smeden, Kareem Carr, Rachel Sippy, Marc de Kamps, Andrew Beam, Stefan Konigorski, Christoph Lippert, Mark S Gilthorpe, Peter W G Tennant

Machine learning methods, combined with large electronic health databases, could enable a personalised approach to medicine through improved diagnosis and prediction of individual responses to therapies. If successful, this strategy would represent a revolution in clinical research and practice. However, although the vision of individually tailored medicine is alluring, there is a need to distinguish genuine potential from hype. We argue that the goal of personalised medical care faces serious challenges, many of which cannot be addressed through algorithmic complexity, and call for collaboration between traditional methodologists and experts in medical machine learning to avoid extensive research waste.

Lancet Digital Health 2020

Published Online
September 16, 2020
https://doi.org/10.1016/S2589-7500(20)30200-4

Centre for Biostatistics, Manchester Academic Health Science Centre, Division of

Promises of machine learning

- Several success with machine learning applied in health, eg diabetic retinopathy
- A systematic review of 82 studies makes for sobering reading:
  - Only 24% with external validation
  - Only 17% compared to health professionals
  - Unrealistic conditions regarding clinical practice
- Poor performance of diabetic retinopathy algorithm in poorly-lit eye clinics in Thailand

Liu X, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digital Health 2019

Beede E, et al. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. April, 2020

Promises of machine learning

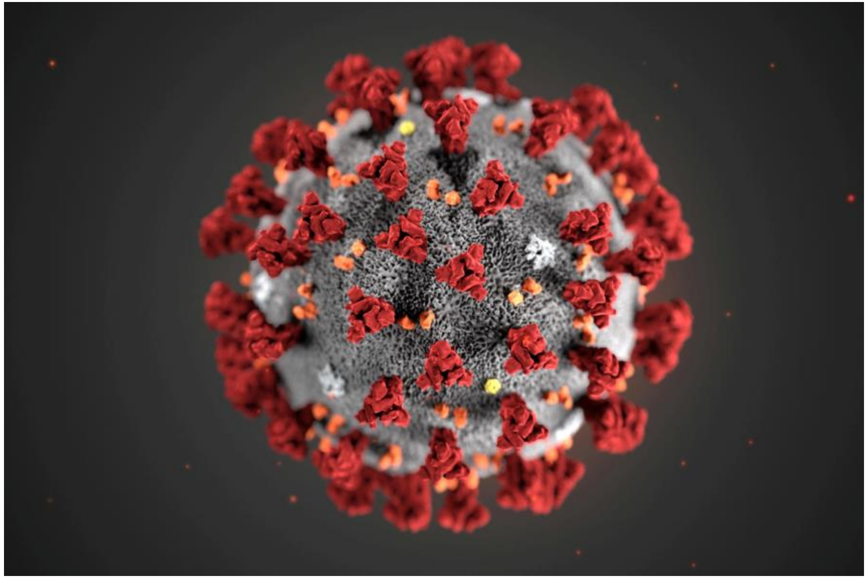
Attack text label iPod

Granny Smith	85.6%
iPod	0.4%
library	0.0%
pizza	0.0%
toaster	0.0%
dough	0.1%

Granny Smith	0.1%
iPod	99.7%
library	0.0%
pizza	0.0%
toaster	0.0%
dough	0.0%

When we put a label saying “iPod” on this Granny Smith apple, the model erroneously classifies it as an iPod in the zero-shot setting.

COVID and machine learning



COVID and machine learning

RESEARCH

Check for updates

FAST TRACK

Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal

Laure Wynants,[1,2] Ben Van Calster,[2,3] Gary S Collins,[4,5] Richard D Riley,[6] Georg Heinze,[7] Ewoud Schuit,[8,9] Marc M J Bonten,[8,10] Darren L Dahly,[11,12] Johanna A Damen,[8,9] Thomas P A Debray,[8,9] Valentijn M T de Jong,[8,9] Maarten De Vos,[2,13] Paula Dhiman,[4,5] Maria C Haller,[7,14] Michael O Harhay,[15,16] Liesbet Henckaerts,[17,18] Pauline Heus,[8,9] Michael Kammer,[7,19] Nina Kreuzberger,[20] Anna Lohmann,[21] Kim Luijken,[21] Jie Ma,[5] Glen P Martin,[22] David J McLernon,[23] Constanza L Andaur Navarro,[8,9] Johannes B Reitsma,[8,9] Jamie C Sergeant,[24,25] Chunhu Shi,[26] Nicole Skoetz,[19] Luc J M Smits,[1] Kym I E Snell,[6] Matthew Sperrin,[27] René Spijker,[8,9,28] Ewout W Steyerberg,[3] Toshihiko Takada,[8] Ioanna Tzoulaki,[29,30] Sander M J van Kuijk,[31] Bas C T van Bussel,[1,32] Iwan C C van der Horst,[32] Florien S van Royen,[8] Jan Y Verbakel,[33,34] Christine Wallisch,[7,35,36] Jack Wilkinson,[22] Robert Wolff,[37] Lotty Hooft,[8,9] Karel G M Moons,[8,9] Maarten van Smeden[8]

BMJ: first published as 10.1136/bmj.m1328 o

- 232 diagnostic or prognostic prediction models

COVID and machine learning

Conclusion Prediction models for covid-19 are quickly entering the academic literature to support medical decision making at a time when they are urgently needed. This review indicates that almost all pubished prediction models are poorly reported, and at high risk of bias such that their reported predictive performance is probably optimistic. However, we have identified two (one diagnostic and one prognostic) promising models that should soon be validated in multiple cohorts, preferably through collaborative efforts and data sharing to also allow an investigation of the stability and heterogeneity in their performance across populations and settings. Details on all reviewed models are publicly available at https://www.covprecise.org/. Methodological guidance as provided in this paper should be followed because unreliable predictions could cause more harm than benefit in guiding clinical decisions. Finally, prediction model authors should adhere to the TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) reporting guideline.

Systematic review registration Protocol https://osf.io/ehc47/, registration https://osf.io/wy245.

COVID and machine learning

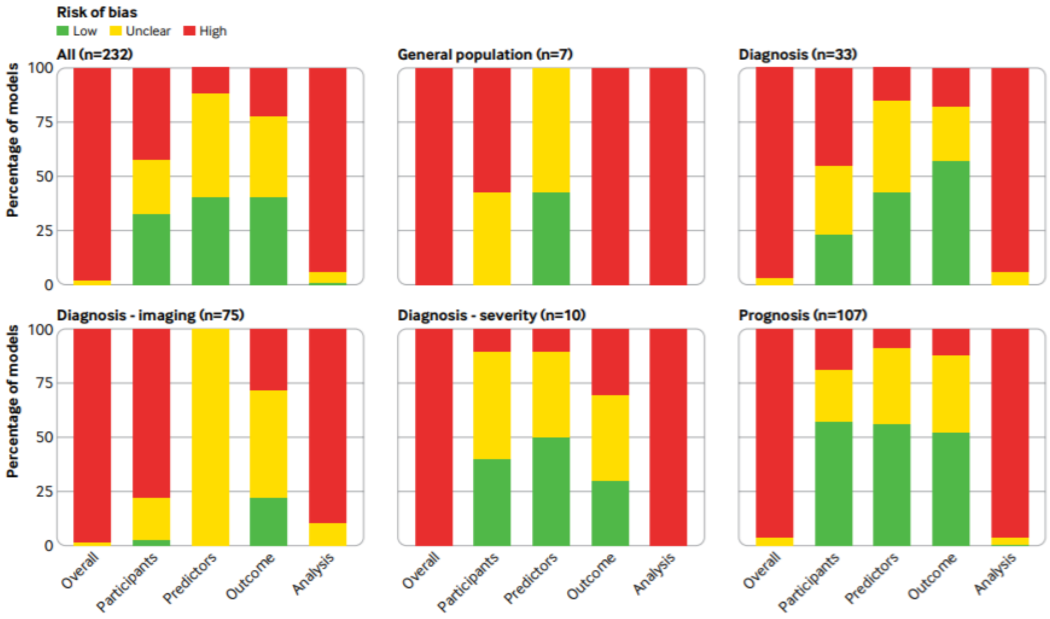


Fig 2 | PROBAST (prediction model risk of bias assessment tool) risk of bias for all included models combined (n=232) and broken down per type of model

COVID and machine learning

RESEARCH

Check for updates

FAST TRACK

Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal

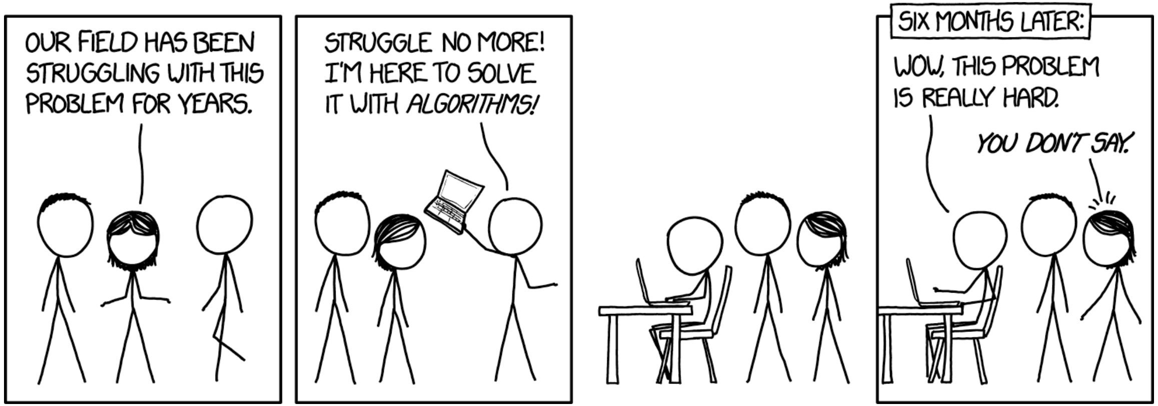
Laure Wynants,[1,2] Ben Van Calster,[2,3] Gary S Collins,[4,5] Richard D Riley,[6] Georg Heinze,[7] Ewoud Schuit,[8,9] Marc M J Bonten,[8,10] Darren L Dahly,[11,12] Johanna A Damen,[8,9] Thomas P A Debray,[8,9] Valentijn M T de Jong,[8,9] Maarten De Vos,[2,13] Paula Dhiman,[4,5] Maria C Haller,[7,14] Michael O Harhay,[15,16] Liesbet Henckaerts,[17,18] Pauline Heus,[8,9] Michael Kammer,[7,19] Nina Kreuzberger,[20] Anna Lohmann,[21] Kim Luijken,[21] Jie Ma,[5] Glen P Martin,[22] David J McLernon,[23] Constanza L Andaur Navarro,[8,9] Johannes B Reitsma,[8,9] Jamie C Sergeant,[24,25] Chunhu Shi,[26] Nicole Skoetz,[19] Luc J M Smits,[1] Kym I E Snell,[6] Matthew Sperrin,[27] René Spijker,[8,9,28] Ewout W Steyerberg,[3] Toshihiko Takada,[8] Ioanna Tzoulaki,[29,30] Sander M J van Kuijk,[31] Bas C T van Bussel,[1,32] Iwan C C van der Horst,[32] Florien S van Royen,[8] Jan Y Verbakel,[33,34] Christine Wallisch,[7,35,36] Jack Wilkinson,[22] Robert Wolff,[37] Lotty Hooft,[8,9] Karel G M Moons,[8,9] Maarten van Smeden[8]

BMJ: first published as 10.1136/bmj.m1328 c

Discussion

Discussion

- Machine learning as a high-value tool to manage Big Data
- Machine learning for prediction model: ups and downs
- Never a magic answer



Source: xkdc

Acknowledgements

- U1190: Translational Research Laboratory for Diabetes
- WP3 of RHU PreciNASH
- François PATTOU, Violeta RAVERDY, Guillaume LASSAILLY, Sylvie DEUFFIC-BURBAN, Philippe MATHURIN
- Martin VAN SMEDEN (@MaartenvSmeden) for slides ideas

pierre.bauvin@univ-lille.fr

"Prediction is very difficult, especially if it's about the future

Niels Bohr, physicist

Questions

?