Prognosis of infectious diseases

A plea for meta-analysis of individual participant data

Thomas Debray

UNIVERSITAS KATOLIK INDONESIA
ATMA JAYA
Tepercaya Kualitas Lulusannya

About me

Background in biostatistics, epidemiology & machine learning

- Assistant Professor
- Founder
- Affiliated Researcher
- Honorary Senior Research Associate
- Honorary Departmental Senior Research Fellow

UMC Utrecht

SMART DATA ANALYSIS AND STATISTICS

Cochrane

UCL

UNIVERSITY OF OXFORD

UNIVERSITAS KATOLIK INDONESIA ATMA JAYA
Tepercaya Kualitas Lulusannya

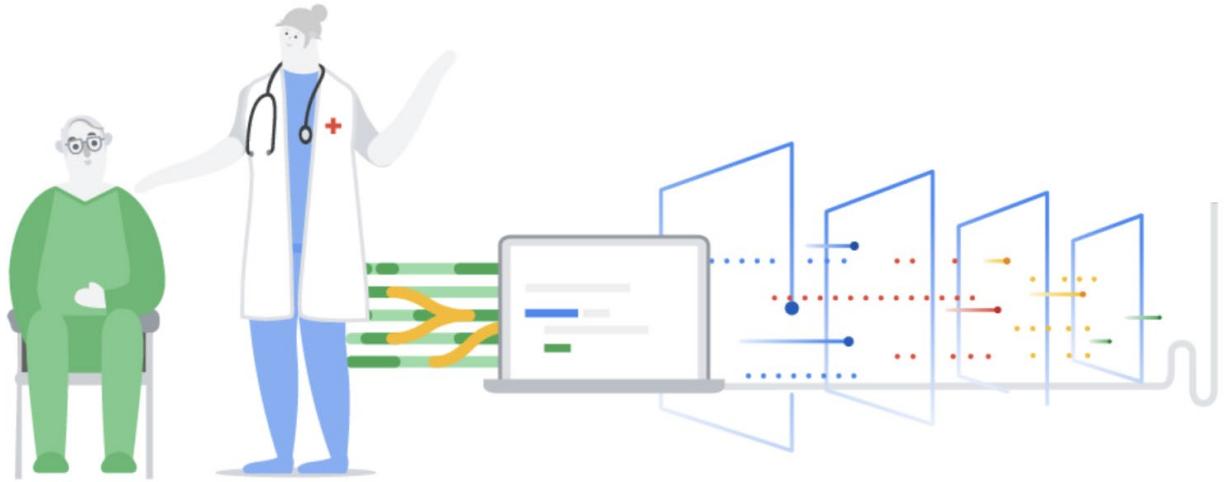
@TPA_Debray

www.thomasdebray.be

Prediction

Estimate something that is yet unknown

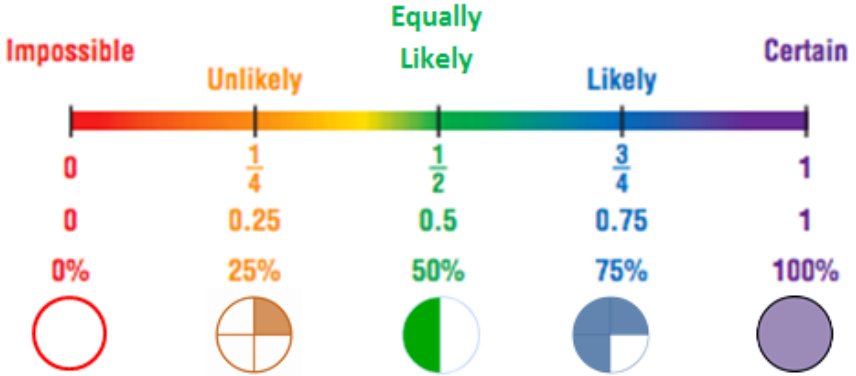
- Presence of a certain disease (diagnosis)
- Future occurrence of a particular event (prognosis)



Prediction

Calculate the absolute risk (probability) for distinct individuals

Example: What is my risk of having a coronavirus-19 infection?



Prediction

What is my risk of being diseased with coronavirus-19?

COVID-19 Early Warning Score (COVID-19 EWS)

Parameters	Assessment	Score
Signs of pneumonia on CT	Yes	5
History of close contact with COVID-19 confirmed patient	Yes	5
Fever	Yes	3
Age	≥ 44 years old	1
Sex	Male	1
Tmax[a]	≥ 37.8 ℃ (100 ℉)	1
Meaningful respiratory symptoms (including cough, expectoration, and dyspnea)	≥ 1 sympotom	1
NLR[b]	≥ 5.8	1
	Highly suspected patient	≥ 10

*SARS-CoV-2 nucleicacid detection positive is the independent diagnostic indicator.
[a]Tmax: the highest body temperature from illness onset to first hospital admission
[b]NLR: neutrophil-to-lymphocyte ratio

Prediction

What is my risk of being diseased with coronavirus-19? Quite low!

COVID-19 Early Warning Score (COVID-19 EWS)

Parameters	Assessment	Score
Signs of pneumonia on CT	Yes	
History of close contact with COVID-19 confirmed patient	Yes	
Fever	Yes	
Age	≥ 44 years old	
Sex	Male	1
Tmax[a]	≥ 37.8 ℃ (100 ℉)	
Meaningful respiratory symptoms (including cough, expectoration, and dyspnea)	≥ 1 sympotom	
NLR[b]	≥ 5.8	
	Highly suspected patient	≥ 10

*SARS-CoV-2 nucleicacid detection positive is the independent diagnostic indicator.
[a]Tmax: the highest body temperature from illness onset to first hospital admission
[b]NLR: neutrophil-to-lymphocyte ratio

← Only one risk factor present

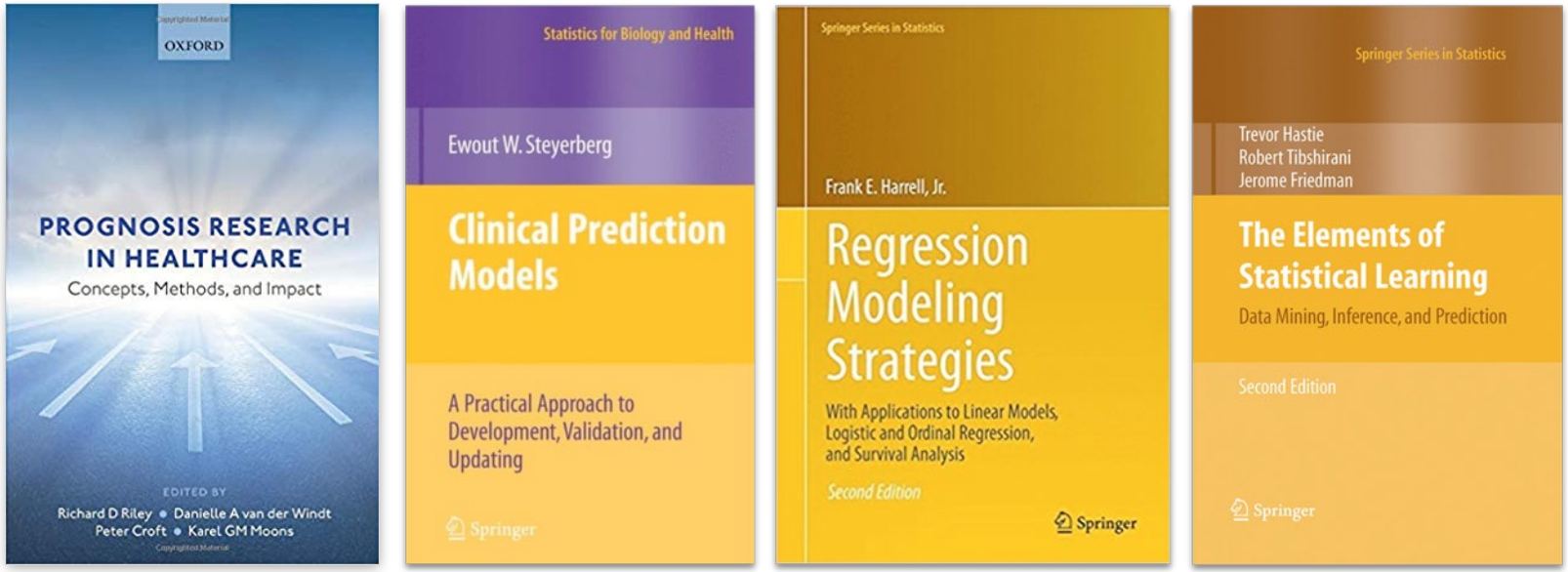
Why do we predict?

To support clinical decision-making for individual patients

- Inform patients and their families
- Decide upon further testing
  (e.g. magnetic resonance imaging)
- Decide upon patient referral
  (e.g. to secondary care)
- Targeting prevention strategies
  (e.g. vaccination)
- Guide treatment decisions
  (e.g. chemotherapy)

How to develop a prediction model?

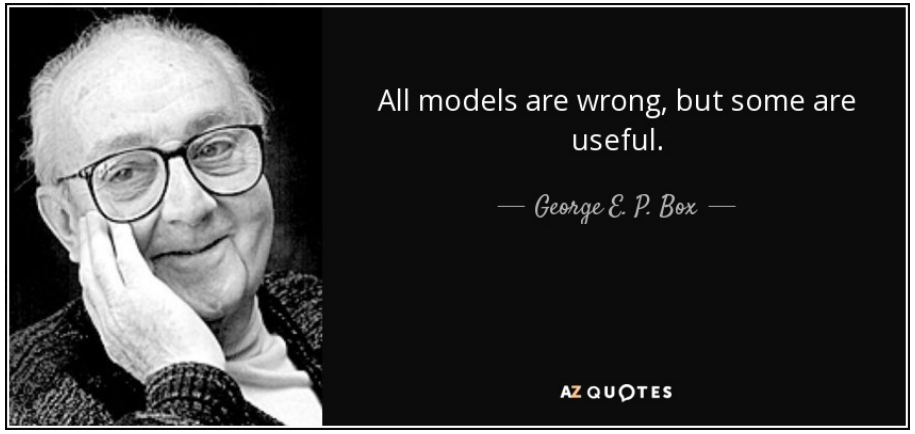
Adopt regression modeling and/or machine learning methods



The reality of (most) prediction models

Many prediction models perform more poorly than anticipated, do not affect clinical practice, or are implemented for the wrong reasons

- Small & poor quality studies
- Limited variation in studied patients, settings or populations
- Lack of validity and effectiveness assessments



Prediction models for COVID-19

RESEARCH

thebmj

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] Johanna A A Damen,[8,9] Thomas P A Debray,[8,9] Maarten De Vos,[2,11] Paula Dhiman,[4,5] Maria C Haller,[7,12] Michael O Harhay,[13,14] Liesbet Henckaerts,[15,16] Nina Kreuzberger,[17] Anna Lohmann,[18] Kim Luijken,[18] Jie Ma,[5] Constanza L Andaur Navarro,[8,9] Johannes B Reitsma,[8,9] Jamie C Sergeant,[19,20] Chunhu Shi,[21] Nicole Skoetz,[17] Luc J M Smits,[1] Kym I E Snell,[6] Matthew Sperrin,[22] René Spijker,[8,9] Ewout W Steyerberg,[3] Toshihiko Takada,[4] Sander M J van Kuijk,[23] Florien S van Royen,[8] Christine Wallisch,[7,24,25] Lotty Hooft,[8,9] Karel G M Moons,[8,9] Maarten van Smeden[8]

Results: 4909 titles were screened, and 51 studies describing 66 prediction models were included (31 march 2020)

Prediction models for COVID-19

Status quo: 66 prediction models

- **3 models for predicting hospital admission** from pneumonia
- **47 diagnosis models** for COVID-19 or COVID-19 pneumonia
  - 34 based on medical images (deep learning)
- **16 prognosis models** for predicting mortality risk, progression to severe disease, or length of stay

Prediction models for COVID-19

Critical appraisal using PROBAST

Assess risk of bias on **four domains** using "signaling questions"

- Participants (2 questions)
- Predictors (3 questions)
- Outcome (6 questions)
- Analysis (9 questions)

If risk of bias was high in at least one domain,
overall risk of bias was judged to be high

Prediction models for COVID-19: critical appraisal

- Participants domain: **24/51 at high risk of bias**
  - Non-representative of the target population (e.g., non-consecutive patients)
- Predictors domain: **6/51 at high risk of bias**
  - Predictors not available at time of intended model use
- Outcome domain: **18/51 at high risk of bias**
  - Subjective or proxy outcomes
- Analysis domain: **50/51 at high risk of bias**
  - Small sample size (->overfitting & no adjustment), incomplete reporting of model performance (e.g., no calibration)

All studies at high risk of bias

Prediction models for COVID-19: what is the problem?

- Available data sources
  - Are often small
  - Entail a particular setting or population (e.g. single hospital)

- Prediction model studies
  - Are hastily conducted
  - Adopt inappropriate statistical methods
  - Do not adequately report methods & results

The majority of developed prediction models are unreliable and unsuitable for use in routine care.

How to move forward?

Collaborative research is urgently needed

- To defragment ongoing research activities
- To improve the overall quality and validity of COVID-19 related prediction models

3 strategies:

- Formation of international consortia
- Development of data sharing platforms
- Meta-analysis of individual participant data

Improving collaborative research

Reconciliation of Cohort Data for Infectious Diseases

Intellectual Property		Technology
Public platform where cohort meta-data can be uploaded. Potential users can contact the *owners*, to get access to the data (or the data is freely available).	**Harmonization & Analysis**	"**Cohort Cloud„** Hosting the data Supported by the *Danish Computerome* and *EMBL*

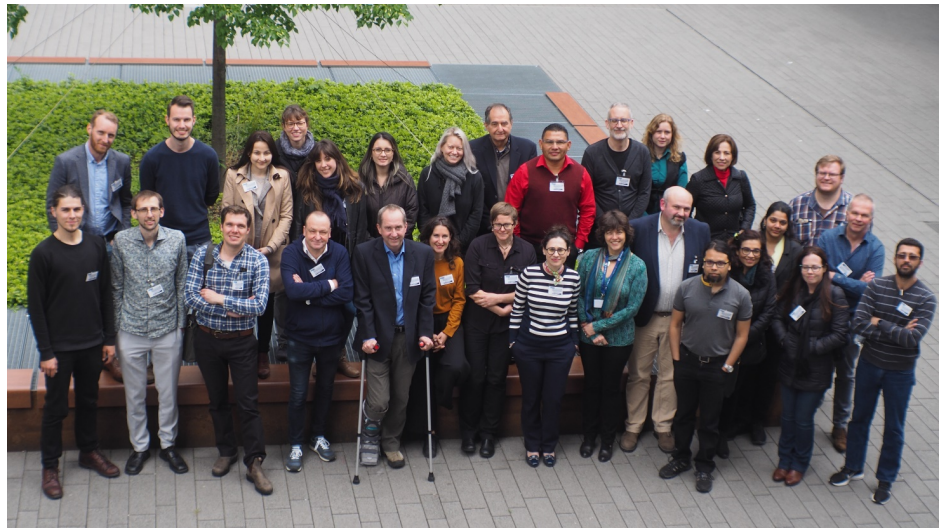
"PEARL„ solutions – politial, ethical, administrative, regulatory, legal

www.recodid.eu

Improving collaborative research

- "In this project, we combine ground-breaking work on **data sharing in public health emergencies** [...] to implement a new model for intensified sharing, reuse, knowledge discovery and collaborative research in epidemic response."

- "... we will [...] bridge infectious disease cohorts and the open science community to ensure that populations in **the global south** are not left behind by the **personalized medicine revolution**".



Improving collaborative research: Zika Virus

Zika Virus Individual Participant Data Consortium

- Global collaboration to streamline an international response to ZIKV
  - Studies and surveillance systems
  - Data from Brazil, Colombia, Ecuador, Cuba , St Martin, Martinique, Mexico, Guadaloupe, French Guyana, Honduras, Haiti, Jamaica, Panama, El Salvador, Spain, Suriname, and Venezuela

- Sharing of deidentified participant level data
  - To perform a pooled cohort analysis
  - To investigate the relation between Zika virus infection during pregnancy and adverse fetal, infant and child outcomes

Improving collaborative research: Zika Virus

Zika Virus Individual Participant Data Consortium

Epidemiology
Protocol

PDF

Understanding the relation between Zika virus infection during pregnancy and adverse fetal, infant and child outcomes: a protocol for a systematic review and individual participant data meta-analysis of longitudinal studies of pregnant women and their infants and children

https://bmjopen.bmj.com/content/9/6/e026092

Improving collaborative research: COVID-19

COVID-19 PRECISE

Large consortium to validate existing models for COVID-19

- Lead by UMC Utrecht (the Netherlands)
- Summarize performance and explore sources of heterogeneity

Partners from

- Europe (the Netherlands, Switzerland, Sweden, UK)
- America (USA)
- Asia (China, Singapore)

We are looking for more partners

UMC Utrecht

UNIVERSITAS KATOLIK INDONESIA ATMA JAYA
Tepercaya Kualitas Lulusannya

Acknowledgements

Slides

Prediction models for COVID-19 – Maarten van Smeden

Images

Deep Learning - https://ai.googleblog.com/2018/05/deep-learning-for-electronic-health.html
Probabilities - http://mathfor7thgrade.weebly.com/probability.html
COVID-19 score - https://www.medrxiv.org/content/10.1101/2020.03.05.20031906v1
Collaboration - https://www.nature.com/articles/d41586-018-06037-5