www.ispor.org

ISPOR

Improving healthcare decisions

ISPOR Machine Learning Methods in HEOR Emerging Good Practices Task Force Forum:

Is ML Ready for Prime Time?

ISPOR Europe 2021

ISPOR

Questions for the speakers?

Submit at anytime
ia the Q & A or chat box.

e will address comments in
e discussion session *after*
the presentations.

SECTION

1

Overview

William Crown, PhD, BA, MA

Distinguished Research Scientist
Brandeis University
Boston, MA, USA

Moderator:

William Crown, PhD, BA, MA, Distinguished Research Scientist, Brandeis University Waltham, MA, USA and Task Force Co-Chair

Speakers:

- **Federico Felizzi, MS, PhD, BSc**, Global HEOR Director, Novartis, Basel, Switzerland
- **Noemi Kreif, MSc, PhD**, Senior Research Fellow, Centre of Health Economics, University of York, York, England, UK
- **Pall Jonsson, MSc, BS, PhD,** Programme Director - Data, National Institute for Health and Care Excellence (NICE), Manchester, England, UK

On behalf of the Machine Learning in HEOR Emerging Good Practices Task Force

Machine Learning in HEOR Task Force Members

- **William Padula, MSc, MS, PhD,** Assistant Professor, University of Southern California, Los Angeles, CA, USA and Task Force Co-Chair
- **Blythe Adamson, PhD, MPH,** Director, Quantitative Sciences, Flatiron Health, New York, NY, United States
- **Atul Butte, PhD,** Professor, Pediatrics and Priscilla Chan and Mark Zuckerberg Distinguished Professor, School of Medicine, University of California, San Francisco San Francisco, CA, US
- **Maarten IJzerman, MSc, PhD**, Chair of Cancer Health Services Research, University of Melbourne and University of Twente, Melbourne, Australia
- **Juan-David Rueda, MSc, MD**, Manager, Health Economics & Payer Evidence (Oncology), AstraZeneca, Gaithersburg, MD, USA
- **David Vanness, PhD,** Professor, Pennsylvania State University, University Park, PA, USA

Next Steps: Final Formal Review & Submission

Our report was reviewed and revised this summer.

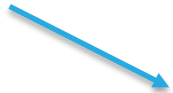
We will be sending the final draft out for review next week.

If you would like to review the final draft, please JOIN our task force review group.

Join Our Task Force Review Group!

1. Visit ISPOR home page www.ispor.org
2. Select “Member Groups”
3. Select “Task Forces”
4. Scroll down to Join a Task Force Review Group
5. Click button to “Join a Review Group”

You must be an ISPOR member to join a Task Force Review Group.



Task Forces

Task forces develop ISPOR’s Good Practices Reports, which are highly cited expert consensus guidance recommendations that set international standards for outcomes research and its use in healthcare decision making.

- Consolidated Health Economic Evaluation Reporting Standards (CHEERS) II
- Joint HTAi - ISPOR Deliberative Processes for HTA NEW
- Machine Learning Methods in HEOR
- Measurement Comparability Between Modes of Administration of PROMs
- Measuring Patient Preferences for Decision Making
- Performance Outcome (PerfO) Assessments
- Systematic Reviews with Cost and Cost-Effectiveness Outcomes

Join a Task Force Review Group

All ISPOR members who are knowledgeable and interested in a task force’s topic may participate in a task force review group. To join a task force review group:

JOIN A REVIEW GROUP

SECTION

2

Machine Learning for Causal Inference

Noemi Kreif, MSc, PhD
Senior Research Fellow, Centre for Health Economics, University of York

Prediction versus causal inference

Prediction

What is the *expected outcome* of a patient with a given covariate profile?

Source: Habets JG, et al.. PeerJ. 2020 Nov 18;8:e10317.

- "Ground truth" can be observed (by holding out some of the data for validation)
- Can't learn about "what if" we change some variables

Prediction versus causal inference

Prediction

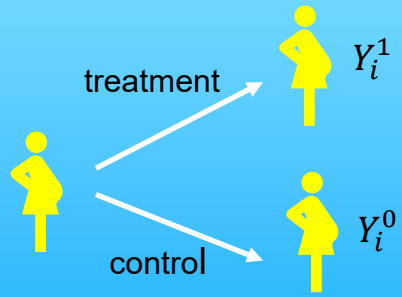
What is the *expected outcome* of a patient with a given covariate profile?

Source: Habets JG, et al.. PeerJ. 2020 Nov 18;8:e10317.

- "Ground truth" can be observed (by holding out some of the data for validation)
- Can't learn about "what if" we change some variables

Causal Inference

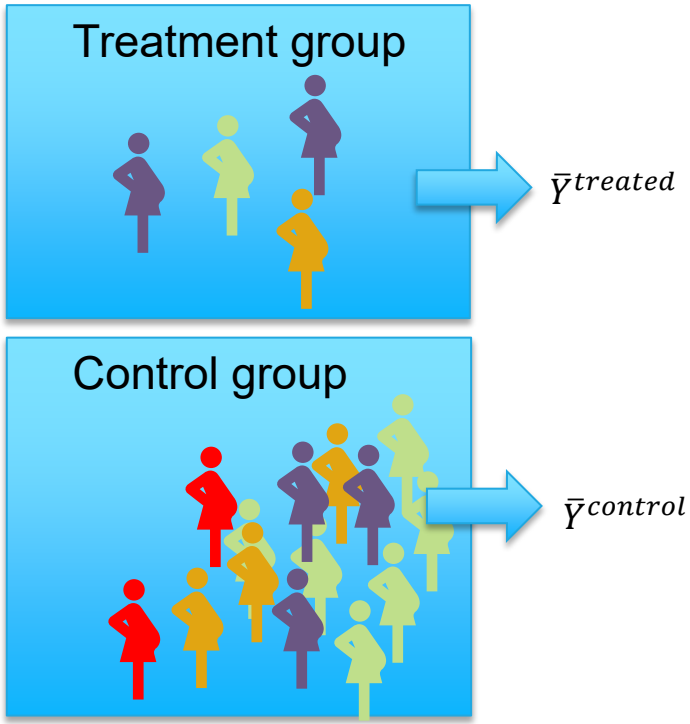
What *would happen* to patient outcomes, if we intervened? (e.g. administer a treatment)



- "Ground truth" can never be observed -> we can make progress, but using assumptions

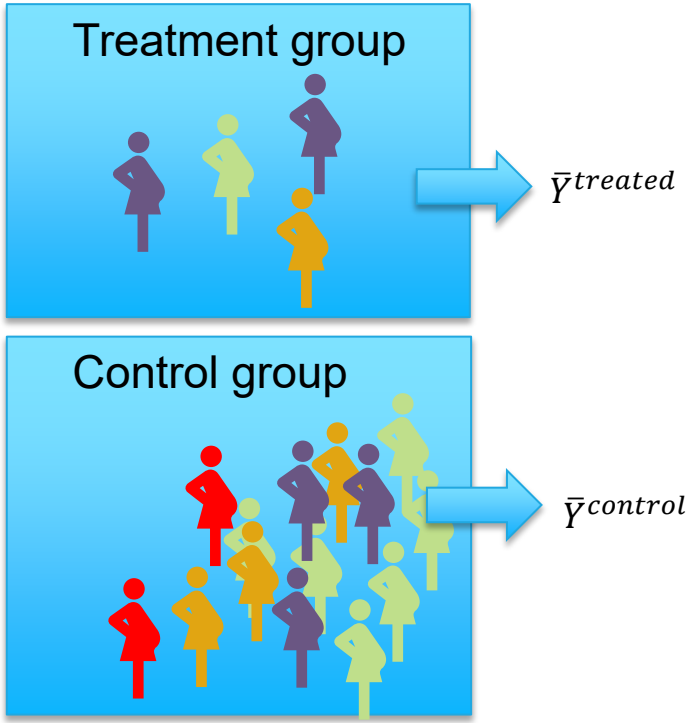
"Causal Machine Learning" harnesses supervised ML methods to answer causal questions

Old questions: what is the treatment effect of an intervention, *on average*?



"Causal Machine Learning" harnesses supervised ML methods to answer causal questions

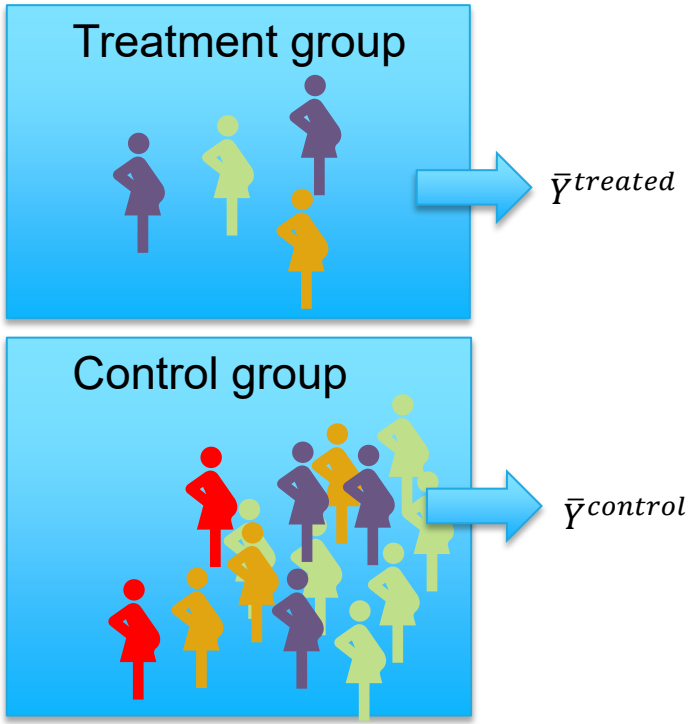
Old questions: what is the treatment effect of an intervention, *on average*?



- When using observational data, confounding by indication leads to bias
  - Average treatment effect ≠ $\bar{Y}^{treated}$ - $\bar{Y}^{control}$
  - To reduce bias:
    - » Select balanced treated and control samples (e.g. via propensity scores) and/or
    - » Adjust for confounders in outcome regression
  - Misspecification in these models-> inaccurate treatment effect estimates

"Causal Machine Learning" harnesses supervised ML methods to answer causal questions

Old questions: what is the treatment effect of an intervention, *on average*?

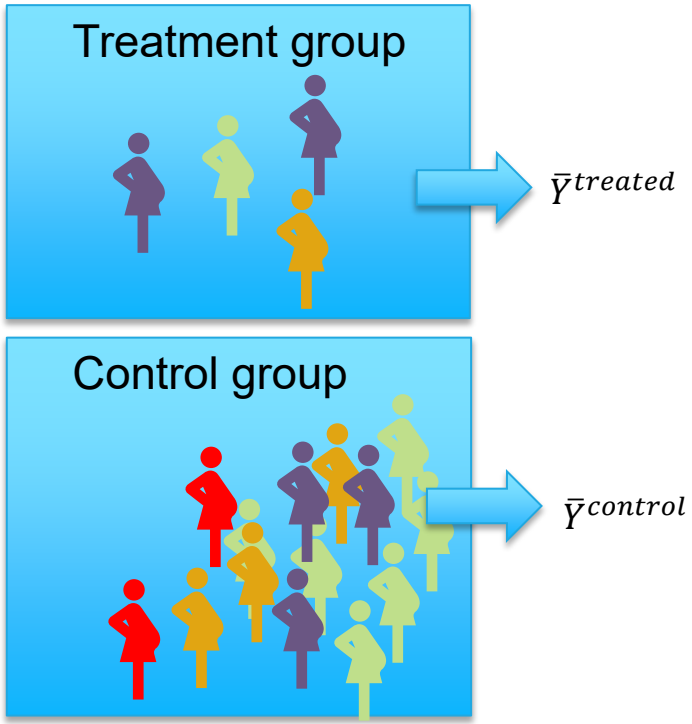


- When using observational data, confounding by indication leads to bias
  - Average treatment effect ≠ $\bar{Y}^{treated} - \bar{Y}^{control}$
  - To reduce bias:
    - » Select balanced treated and control samples (e.g. via propensity scores) and/or
    - » Adjust for confounders in outcome regression
  - Misspecification in these models-> inaccurate treatment effect estimates
- ML can help with flexible model specification
  - E.g. random forests can naturally account for interactions

Statistical theory tells us how to best combine ML estimates to reduce bias in estimated treatment effect

"Causal Machine Learning" harnesses supervised ML methods to answer causal questions

Old questions: what is the treatment effect of an intervention, *on average*?



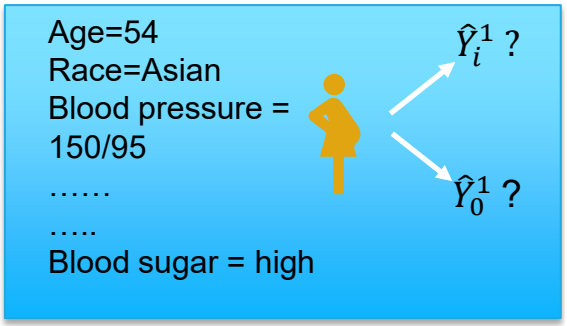
- When using observational data, confounding by indication leads to bias
  - Average treatment effect ≠ $\bar{Y}^{treated} - \bar{Y}^{control}$
  - To reduce bias:
    - » Select balanced treated and control samples (e.g. via propensity scores) and/or
    - » Adjust for confounders in outcome regression
  - Misspecification in these models-> inaccurate treatment effect estimates
- ML can help with flexible model specification
  - E.g. random forests can naturally account for interactions

Statistical theory tells us how to best combine ML estimates to reduce bias in estimated treatment effect

Methods to look for: Targeted minimum loss based estimation (TMLE), double machine learning

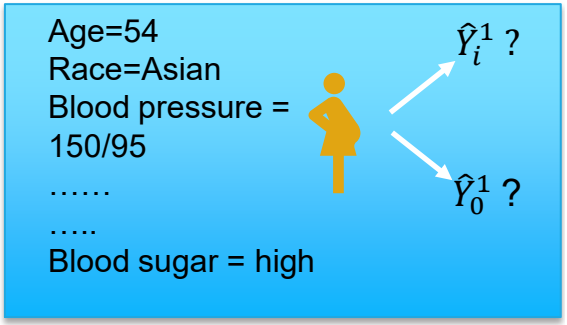
"Causal Machine Learning" harnesses supervised ML methods to answer causal questions

- **New questions**, aiming to inform personalized decision making:
  - which subgroups have the highest response to treatment?
  - what patient characteristics predict treatment response?
  - what is the optimal treatment/dosage for a given patient?



"Causal Machine Learning" harnesses supervised ML methods to answer causal questions

- **New questions**, aiming to inform personalized decision making:
  - which subgroups have the highest response to treatment?
  - what patient characteristics predict treatment response?
  - what is the optimal treatment/dosage for a given patient?

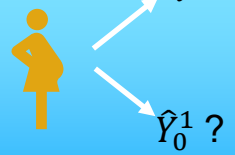


- Can use the ML estimates of propensity scores and regression functions
- Combine them to estimate
  - Conditional average treatment effects $\hat{E}[Y_i^1 - Y_i^0 | X = x]$
  - Individual treatment effects $\hat{Y}_i^1 - \hat{Y}_i^0$
  - Optimal treatment decision rules

"Causal Machine Learning" harnesses supervised ML methods to answer causal questions

- **New questions**, aiming to inform personalized decision making:
  - which subgroups have the highest response to treatment?
  - what patient characteristics predict treatment response?
  - what is the optimal treatment/dosage for a given patient?

Age=54
Race=Asian
Blood pressure = 150/95
……
…..
Blood sugar = high



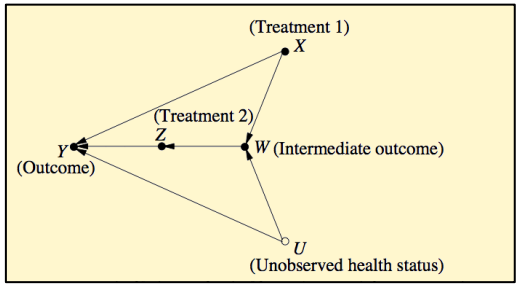
- Can use the ML estimates of propensity scores and regression functions
- Combine them to estimate
  - Conditional average treatment effects $\hat{E}[Y_i^1 - Y_i^0 | X = x]$
  - Individual treatment effects $\hat{Y}_i^1 - \hat{Y}_i^0$
  - Optimal treatment decision rules

Methods to look for: Causal Forests, counterfactual treatment outcomes, individualised treatment effects, optimal dynamic treatment regimes, etc.

Assumptions

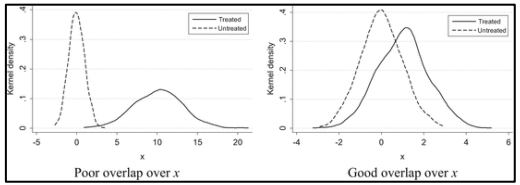
No unmeasured confounders

- Do I have all important variables that affect treatment assignment and predict outcomes in my data?
- Discuss with experts, draw causal diagrams!



Positivity / overlap

- Can the patients in my data realistically receive both treatment options?

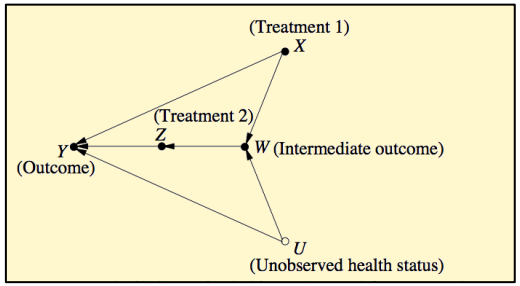


Cerulli G. 2015. https://doi.org/10.1007/978-3-662-46405-2_2

Assumptions

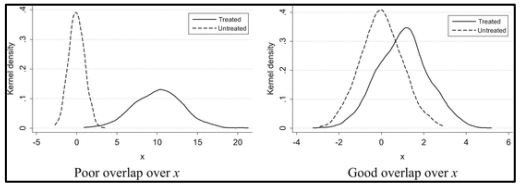
No unmeasured confounders

- Do I have all important variables that affect treatment assignment and predict outcomes in my data?
- Discuss with experts, draw causal diagrams!



Positivity / overlap

- Can the patients in my data realistically receive both treatment options?



Cerulli G. 2015. https://doi.org/10.1007/978-3-662-46405-2_2

Which ML methods to use?

- Be transparent about choices, use ensembles of methods such as the “Super Learner”
- Use simulation evidence and available guidance (different tasks may require different learners)

The reading list

- Causal machine learning field fast expanding across several fields
  - *statistics*
  - *econometrics*
  - *computers science*

- With direct applicability to HEOR!
  - *Available software in R and Python*

- Some methodological expansions needed (e.g. uncertainty estimation)

Theory and Methods

Estimation and Inference of Heterogeneous Treatment Effects using Random Forests

Stefan Wager & Susan Athey

Pages 1228-1242 | Received 01 Dec 2015, Accepted author version posted online: 21 Apr 2017, Published online: 06 Jun 2018

Download citation https://doi.org/10.1080/01621459.2017.1319839 Check for updates

Full Article Figures & data References Supplemental Citations Metrics Reprints & Permissions Get access

Clinical Pharmacology & Therapeutics

Tutorial | Open Access

From Real-World Patient Data to Individualized Treatment Effects Using Machine Learning: Current and Future Methods to Address Underlying Challenges

Quantifying Ignorance in Individual-Level Causal-Effect Estimates under Hidden Confounding

Andrew Jesson [1] Sören Mindermann [1] Yarin Gal [1] Uri Shalit [2]

SECTION

3

How Do Payers & HTAs Deal With the Black Box?

Pall Jonsson, MSc, BS, PhD

Programme Director - Data, National Institute for Health and Care Excellence (NICE), Manchester, England, UK

Unintended complications will happen

My account Search UK edition
The Guardian
For 200 years
News website of the year
News Opinion Sport Culture Lifestyle More
UK World Coronavirus Climate crisis Football Business Environment UK politics Education Society Science Tech Global development

Skin cancer

AI skin cancer diagnoses risk being less accurate for dark skin – study

Research finds few image databases available to develop technology contain details on ethnicity or skin type

Nicola Davis *Science correspondent*

Tue 9 Nov 2021 23.30 GMT

Studies suggest image recognition technology can classify skin cancers as successfully as humans. Posed by model. Photograph: ChesiireCat/Getty Images/iStockphoto

HTAs want certainty

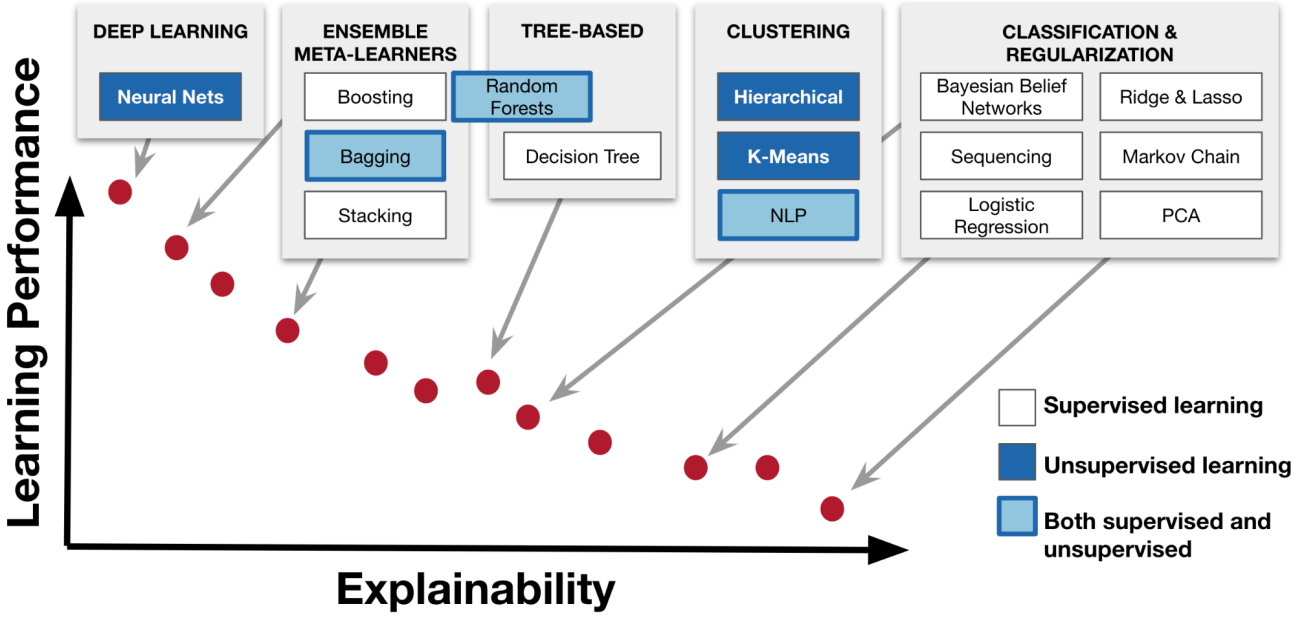
Box 1. The new definition of HTA

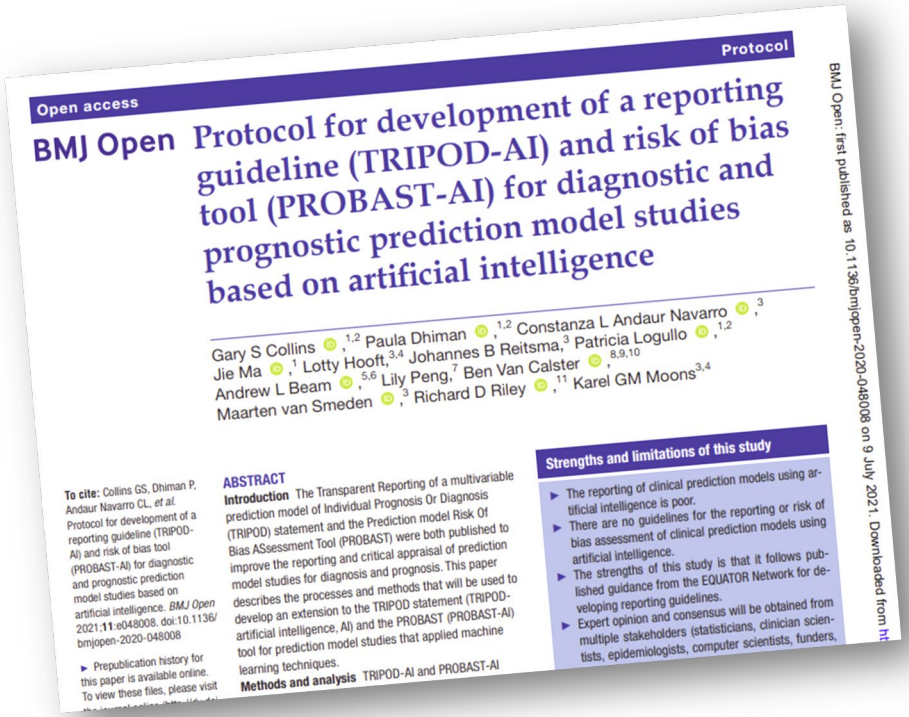
The definition of HTA is provided below, with important clarifying information provided in four accompanying notes.

HTA is a multidisciplinary process that uses explicit methods to determine the value of a health technology at different points in its lifecycle. The purpose is to inform decision-making in order to promote an equitable, efficient, and high-quality health system.

O'Rourke et al. (2020). *The new definition of health technology assessment: A milestone in international collaboration.* IJTAHC36, 187–190.

Explainability of ML varies





Open access Protocol

BMJ Open

Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence

Gary S Collins ,[1,2] Paula Dhiman ,[1,2] Constanza L Andaur Navarro ,[3] Jie Ma ,[1] Lotty Hooft,[3,4] Johannes B Reitsma,[3] Patricia Logullo ,[1,2] Andrew L Beam ,[5,6] Lily Peng,[7] Ben Van Calster ,[8,9,10] Maarten van Smeden ,[3] Richard D Riley ,[11] Karel GM Moons[3,4]

To cite: Collins GS, Dhiman P, Andaur Navarro CL, *et al*. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open* 2021;**11**:e048008. doi:10.1136/bmjopen-2020-048008

► Prepublication history for this paper is available online. To view these files, please visit

ABSTRACT

Introduction The Transparent Reporting of a multivariable prediction model of Individual Prognosis Or Diagnosis (TRIPOD) statement and the Prediction model Risk Of Bias ASsessment Tool (PROBAST) were both published to improve the reporting and critical appraisal of prediction model studies for diagnosis and prognosis. This paper describes the processes and methods that will be used to develop an extension to the TRIPOD statement (TRIPOD-artificial intelligence, AI) and the PROBAST (PROBAST-AI) tool for prediction model studies that applied machine learning techniques.

Methods and analysis TRIPOD-AI and PROBAST-AI

Strengths and limitations of this study

- ► The reporting of clinical prediction models using artificial intelligence is poor.
- ► There are no guidelines for the reporting or risk of bias assessment of clinical prediction models using artificial intelligence.
- ► The strengths of this study is that it follows published guidance from the EQUATOR Network for developing reporting guidelines.
- ► Expert opinion and consensus will be obtained from multiple stakeholders (statisticians, clinician scientists, epidemiologists, computer scientists, funders,

Medicines & Healthcare products Regulatory Agency

Health Canada Santé Canada

FDA U.S. FOOD & DRUG ADMINISTRATION

Guidance

Good Machine Learning Practice for Medical Device Development: Guiding Principles

Published 27 October 2021

The U.S. Food and Drug Administration (FDA), Health Canada, and the United Kingdom's Medicines and Healthcare products Regulatory Agency (MHRA) have jointly identified 10 guiding principles that can inform the development of Good Machine Learning Practice (GMLP). These guiding principles will help promote safe, effective, and high-quality medical devices that use artificial intelligence and machine learning (AI/ML).

https://www.fda.gov/media/153486/download

ISPOR ML task force: PALISADE

"Key considerations for communicating the appropriate use of ML to stakeholders and decision makers"

Element	Definition
Purpose	Is the purpose of the algorithm clearly stated at the outset? Is the implementation of the algorithm in a healthcare setting fair and ethical?
Appropriateness	Is there a clear justification that the algorithm is acceptable in the context within which it is being applied?
Limitations	Have the strengths and limitations, in the context of the purpose, been identified? This should cover both the algorithm and any data used.
Implementation	Consideration of access, implementation and resource issues when implemented in healthcare settings
Sensitivity & Specificity	For classification algorithms, has the model performance and accuracy (specificity and sensitivity) been appropriately determined?
Algorithm characteristics	Has the machine learning mechanism been clearly characterised and described? Is there sufficient transparency for the results to be reproducible?
Data characteristics	Is the selection of datasets justified and are the key characteristics known? This should extend to training sets, test sets and validation sets.
Explainability	Are the outputs of the algorithm clearly understandable by both the healthcare professional and the patient?

Thank you!

Pall.Jonsson@nice.org.uk

SECTION

4

Case Study: Diabetic Retinopathy

Federico Felizzi, MS, PhD, BSc

Global HEOR Director, Novartis
Basel, Switzerland

Overview

- Deep learning systems in diabetic retinopathy
- A tree-based decision modeling framework
- Effects of reimbursement policies
- Implications for long-term disease management costs

AI in Diabetic Retinopathy


JAMA | Original Investigation

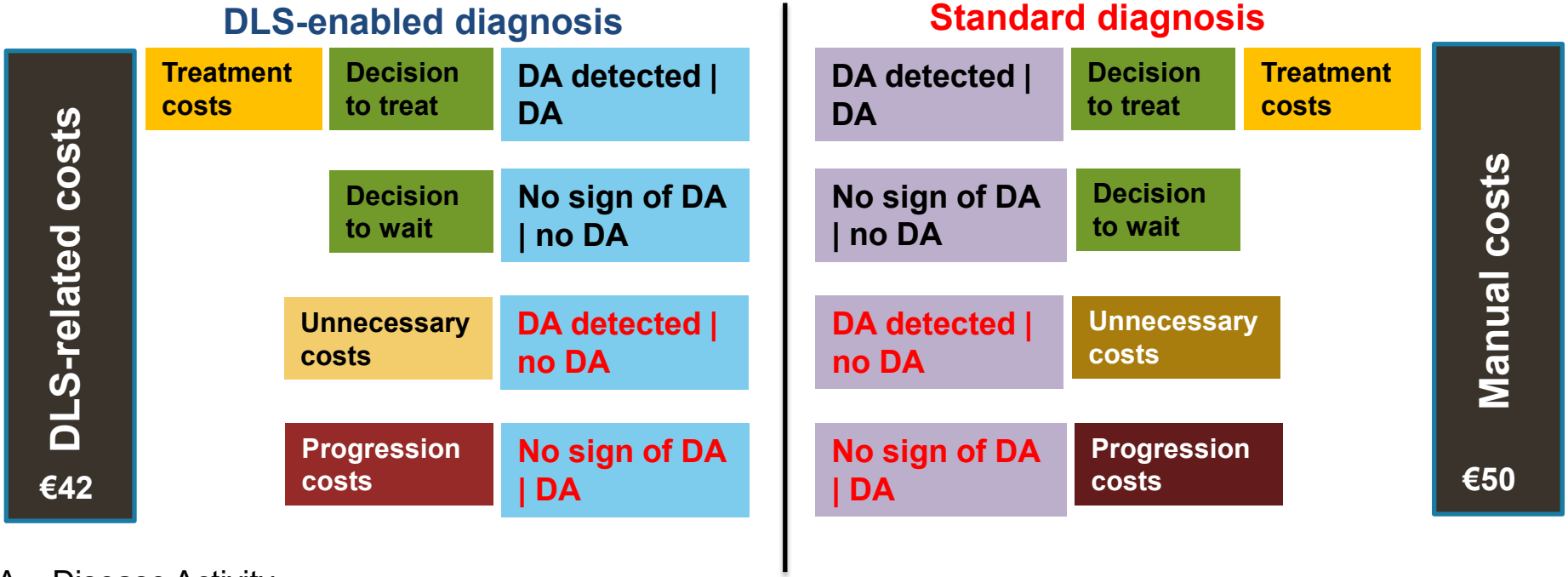
Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes

Daniel Shu Wei Ting, MD, PhD; Carol Yim-Lui Cheung, PhD; Gilbert Lim, PhD; Gavin Siew Wei Tan, FRCSEd; Nguyen D. Quang, BEng;

- Deep Neural Nets trained without specifying lesion-based features
- Consistency of Interpretation
- Near Instantaneous Reporting of results

		DLS	Professional Grader
Referable DR			
	Sensitivity	**90.5%**	**91.2%**
	Specificity	**91.6%**	**99.3%**
Vision Threatening DR			
	Sensitivity	**100%**	**88.5%**
	Specificity	**91.1%**	**99.6%**

A Decision Modeling Framework



DA = Disease Activity

[Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: an economic analysis modelling study (thelancet.com)]

Policies and Implications for Treatment Pathways

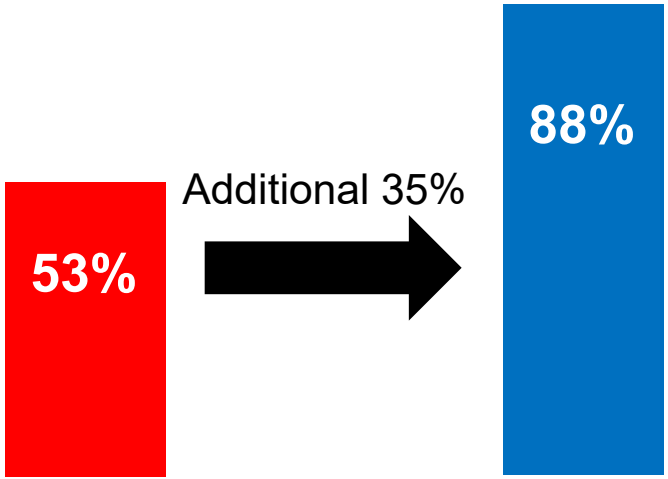
Cost-Utility Analysis of Extending Public Health Insurance Coverage to Include Diabetic Retinopathy Screening by Optometrists

Sasha van Katwyk, MSc[1], Ya-ping Jin, MD, PhD[2,3], Graham E. Trope, MB, PhD, FRCSC[3], Yvonne Buys, MD, FRCSC[3], Lisa Masucci, MSc[4], Richard Wedge, MD[5], John Flanagan, OD, PhD[2,7], Michael H. Brent, MD, FRCSC[3], Sherif El-Defrawy, MD, PhD, FRCSC[3], Hong Anh Tu, PhD[3], Kednapa Thavorn, MPharm, PhD[1,6,*]

[1]Clinical Epidemiology Program, Ottawa Hospital Research Institute, The Ottawa Hospital, Ottawa, Ontario, Canada; [2]Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; [3]Department of Ophthalmology & Vision Sciences, University of Toronto, Toronto, Ontario, Canada; [4]Institute of Health Policy, Management and Evaluation, University of Toronto, Ontario, Canada; [5]Health Quality Ontario, Toronto, Ontario, Canada; [6]Health PEI, Charlottetown, Prince Edwards Island, Canada; [7]School of Optometry and Vision Science Program, University of California Berkeley, California, USA; [8]School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa; Ottawa, Ontario, Canada; [9]Institute of Clinical Evaluative Sciences (ICES UOttawa), Ottawa, Ontario, Canada

	sensitivity	specificity
Trained Ophthalmologist	82%	95%
Optometrist	74%	84%
GP	46%	81%

(re)Introduction of a policy that reimburses DR screening at optometrist practices



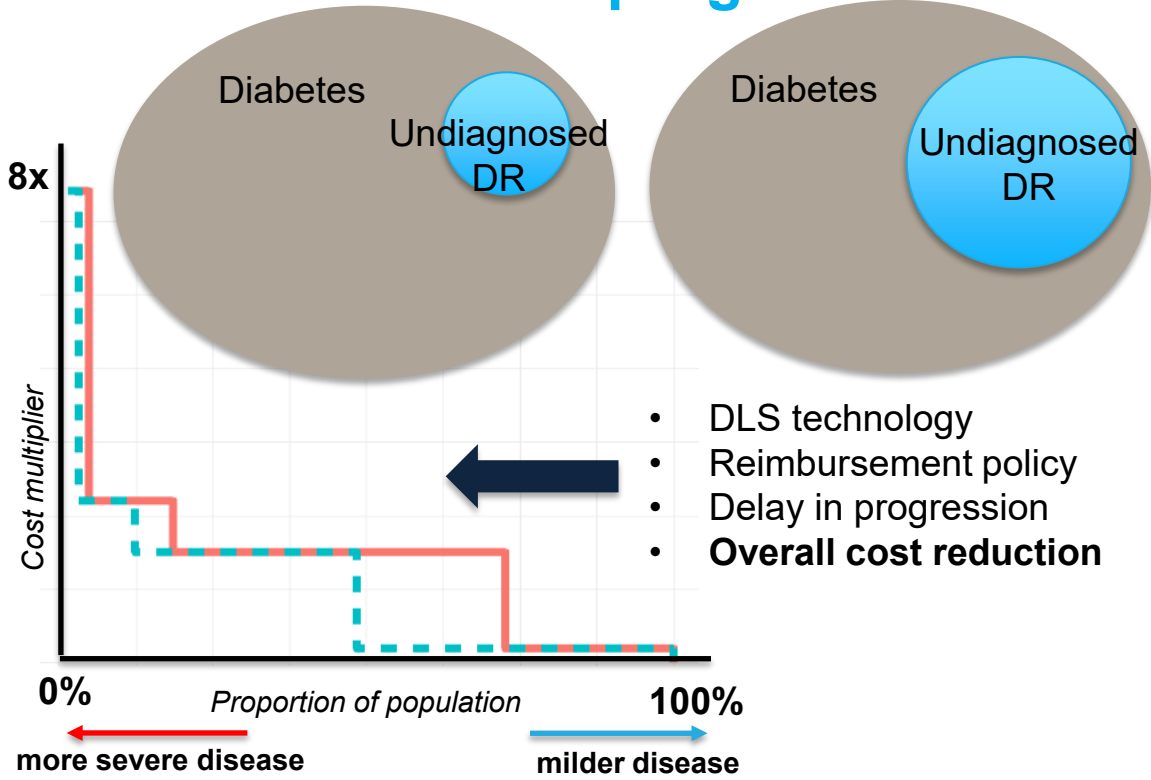
people screened for DR in the **last 2 years** **without policy**

people screened for DR in the **last 2 years** **with policy**

A case-study framework linked to disease progression

Stages of DR	cost multipliers	proportion
no DR	1	68%
mild DR	1.2	24%
moderate DR	2.5	6.1%
severe NPDR	3.2	1.2%
PDR	7.4	0.5%

Treated	16.3%	Progressing to PDR
Not Treated	43.5%	



- DLS technology
- Reimbursement policy
- Delay in progression
- **Overall cost reduction**

Protocol W

JAMA Ophthalmol. 2021;139(7):701-712. doi:10.1001/jamaophthalmol.2021.0606

Overall Conclusions

- Deep learning in retina scans
  - high **sensitivity** and **specificity** and near instantaneous delivery of results
- Implementation in a clinical setting
  - Extension of reimbursement policy to optometrist enables expansion of DR screening in subjects with Diabetes
- Decision modeling framework (tree)
  - Correct disease activity assessment
  - Treatment decision and long-term costs


thank you
danke
謝謝
ngiyabonga
teşekkür ederim
gracias
obrigado
dziękuję
bedankt
go raibh maith agat
merci
спасибо
рахмат
Баярлалаа
dank je
tapadh leat
mauruuru
hvala
sukriya
kop khun krap
terima kasih
arigatō
dakujem
takk
grazie
tạ ơn
mochchakkeram
sagolun
chnorakaloutioun
gracies
gratias ago
sulpay
matondo
misaotra
blagodaram
vinaka
faafetai lava
kiitos
dankie
dhanyavad
nandri
nanri
gracie
bayarlalaa
enkosi
koszonom
sobodi
dekuji
mesi
didi madloba
kam sah hamnida
rahmat
najis tuke
tanemirt
rahmet
ありがとう
감사합니다
ευχαριστώ
diolch
dhanyavadagalu
shukriya
merce
xiexie
trugarez
мерси
barka
mersi
wetalin
tack
paldies
grazzi
mahalo
хвала
asante
manana
obrigada
murakoze
tenki
djiere dieuf
tau
mamnun
дякую
chokrane
akun
dankon
aciu
spas
kia ora


Questions?

5

To contact the presenters:
taskforce@ispor.org