of included studies in Cochrane reviews. J Clin Epidemiol 2019;112: 59−66.

[2] Higgins JPT, Lasserson T, Chandler J, Tovey D, Churchill R. Methodological Expectations of Cochrane Intervention Reviews. Cochrane: London, Version 1.07. 2018. Available at https://community.cochrane.org/mecir-manual. Accessed July 29, 2019.

[3] Metzendorf MI, Richter B. Selective searching for high quality health-related evidence syntheses − more bias or time gained? Poster. Global Evidence Summit, 13-16 Sept 2017, Cape Town, South Africa. Available at https://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=50245. Accessed July 31, 2019.

[4] Metzendorf MI, Richter B, Bandeira-Echtler E, Hausner E, Waffenschmidt S. Descriptive analysis of non-randomized studies included in Cochrane Reviews regarding their availability in PubMed. Poster. 25th Cochrane Colloquium, 16-18 Sept 2018, Edinburgh, UK. Available at https://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=50244. Accessed July 31, 2019.

[5] Halladay CW, Trikalinos TA, Schmid IT, Schmid CH, Dahabreh IJ. Using data sources beyond PubMed has a modest impact on the results of systematic reviews of therapeutic interventions. J Clin Epidemiol 2015;68:1076−84.

[6] Hartling L, Featherstone R, Nuspl M, Shave K, Dryden DM, Vandermeer B. The contribution of databases to the results of systematic reviews: a cross-sectional study. BMC Med Res Methodol 2016;16:127.

http://dx.doi.org/10.1016/j.jclinepi.2019.09.001

Statistical thinking, machine learning

We read with interest *"A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models"* by Christodoulou et al. [1] appreciating its rigor and importance. We question, however, the conceptual dichotomy of logistic regression (LR) vs. machine learning (ML).

Liberally interpreting Breiman's and Mitchell's works [2,3] among others [4], Christodoulou et al. posit that ML differs from regression in that ML lacks underpinning theory or prior domain knowledge, yet acknowledging the difference more as of a continuum than a clear cut [5].

LR exploits a Bernoulli-distributed outcome as a linear function of predictors through the log-odds, with coefficients estimated by maximum likelihood and Newton's method. Incorporated random or other mixed effects can account for data hierarchies. Variations of LR, from probit to Bayesian LR, historically are not regarded as ML [6]. In LR features are usually chosen by content experts based on empirical knowledge or theory, subject to sample size and event rates. In the presence of collinearity, or when theory is vaguer, techniques for selecting variables can be used, for example, regularization. To some extent, feature selection already falls in the ML realm.

Many ML techniques are not much different to LR other than in name and model fit routines. For instance, a support vector machine (SVM) with linear kernel is similar to LR, minimizing the hinge instead of logistic loss [7]. A naïve Bayes classifier can be considered a particular case of LR [8]. In an artificial neural network, a single-layer neuron with sigmoid activation is also LR [9].

We agree with the authors that theory-driven modeling differs from abductive data-driven discovery [10]. However, in the LR/ML definition given by Christodoulou et al. it is debatable that regularization and boosted/bagged LR are included in the LR family, whereas generalized estimating equations are kept separate, as well as linear/nonlinear SVM approaches are pooled together. Furthermore, optimization methods such as genetic algorithms seem to be mixed up with prediction.

We also think that study design should be more of a defining element. Differences in study design can be instrumental in determining *why* and *how* an approach is chosen. A stratified study might be better approximated by a linear model than a general population sample. As mentioned in the review, the number of predictors and sample size can affect the model applicability and performance; of note, the review compared risk of bias rather than performance.

Given these overlaps of ML and (bio)statistical approaches, a more useful categorization can be linear vs. nonlinear (including higher order interactions in LR) [11,12] or model ranking by complexity [13]. After all, clinicians tend to prefer parsimonious, interpretable models such as linear scores or decision rules with few variables as opposed to black boxes calculating nonlinear functions of many predictors [14,15]. We interpret Brian Ripley's *"machine learning is statistics minus checking of models and assumptions"* (useR! 2004, Vienna) as granting ML a broad, computationally empowered scope, yet necessarily rooted in well-founded causal inference, not isolated, specious prediction.

Jiang Bian
Department of Health Outcomes and Biomedical Informatics
University of Florida
Gainesville, FL, USA

Iain Buchan
Department of Public Health and Policy
University of Liverpool
Liverpool, UK

Yi Guo
Department of Health Outcomes and Biomedical Informatics
University of Florida
Gainesville, FL, USA

Mattia Prosperi*
Department of Epidemiology
University of Florida
Gainesville, FL, USA

*Corresponding author. Clinical and Translation Science Building, Suite 4234, 2004 Mowry Road, Gainesville, 32610-0231 FL, USA. Tel.: 352-273-5860; fax: 352-273-5365.
E-mail address: m.prosperi@ufl.edu (M. Prosperi)

Declarations of interest: none.

Author contributions: M.P. and J.B. conceived the original discussion, Y.G. and I.B. contributed to the discussion from biostatistics and epidemiology perspectives. M.P. and J.B. wrote the initial version. All authors revised the article carefully.

References

[1] Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019;110:12−22.
[2] Breiman L. Statistical modeling: the two cultures (with comments and a rejoinder by the author). Stat Sci 2001;16:199−231.
[3] Mitchell TM. Machine learning. 1st ed. New York, NY: McGraw-Hill Education; 1997.
[4] Boulesteix A-L, Schmid M. Machine learning versus statistical modeling. Biom J 2014;56:588−93.
[5] Beam AL, Kohane IS. Big data and machine learning in health care. JAMA 2018;319:1317−8.
[6] Cramer JS. The early origins of the logit model. Stud Hist Philos Sci C 2004;35:613−26.
[7] Vapnik VN. The nature of statistical learning theory. New York, NY: Springer New York; 2000.
[8] Ng AY, Jordan MI. On discriminative vs. generative classifiers: a comparison of logistic regression and naive Bayes. In: Dietterich TG, Becker S, Ghahramani Z, editors. Advances in neural information processing systems 14. Cambridge, MA: MIT Press; 2002:841−8.
[9] Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform 2002;35:352−9.
[10] Bergadano F, Cutello V, Gunetti D. Abduction in machine learning. In: Gabbay DM, Kruse R, editors. Abductive reasoning and learning. Dordrecht: Springer Netherlands; 2000:197−229.
[11] Prosperi MCF, Altmann A, Rosen-Zvi M, Aharoni E, Borgulya G, Bazso F, et al. Investigation of expert rule bases, logistic regression, and non-linear machine learning techniques for predicting response to antiretroviral treatment. Antivir Ther 2009;14:433−42.
[12] Fraccaro P, Nicolo M, Bonetto M, Giacomini M, Weller P, Traverso CE, et al. Combining macula clinical signs and patient characteristics for age-related macular degeneration diagnosis: a machine learning approach. BMC Ophthalmol 2015;15:10.
[13] Vapnik VN, Chervonenkis AY. On the uniform convergence of relative frequencies of events to their probabilities. In: Vovk V, Papadopoulos H, Gammerman A, editors. Measures of complexity: festschrift for alexey chervonenkis. Cham: Springer International Publishing; 2015:11−30.
[14] Prosperi M, Min JS, Bian J, Modave F. Big data hurdles in precision medicine and precision public health. BMC Med Inform Decis Mak 2018;18:139.
[15] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019;1:206−15.

http://dx.doi.org/10.1016/j.jclinepi.2019.08.003

Statistics versus machine learning: definitions are interesting (but understanding, methodology, and reporting are more important)

We thank Bian et al for their interest in our study. Since the study was published, the distinction between logistic regression and machine learning has fueled a lot of discussion. We had addressed this issue already in the initial submission but corroborated on it based on the reviewers' comments. We state in the article that we do not believe there is a clear dichotomy but rather that algorithms lie on a continuum regarding flexibility, and reliance on the data versus subject knowledge. Nevertheless, several publications and discussions explicitly make this distinction and often conclude that machine learning leads to better predictive performance compared with traditional statistical methods. This justifies the pragmatic definition in our article.

We feel that discussing semantics and definitions can be insightful, but it should not be the focal point resulting from our study. The key messages of our study on clinical risk prediction modeling are as follows: (1) by itself, using highly flexible algorithms do not necessarily lead to improved performance, (2) methodological conduct in developing, validating, and fair comparison of different algorithms needs to improve, and (3) the reporting of prediction model studies should adhere to current guidelines such as transparent reporting of a multivariable prediction model for individual prognosis or diagnosis [1]. The sensible development of a prediction model depends on the specific context, should bear in mind the clinical setting in which the algorithm is intended to be used, and needs to be carefully and fully described.

Nevertheless, we respect the comments on the practical choices that we made in our study. We compared "logistic regression" with "machine learning." The category "logistic regression" included standard maximum likelihood logistic regression and penalized logistic regression (lasso, ridge, elastic net) but excluded bagged/boosted logistic regression and other algorithms that we labeled as traditional statistical methods. Penalization (regularization) has origins deep in the statistical literature. We refer to Stein's paradox described in 1955 at the Berkeley Symposium on Mathematical Statistics and Probability [2]. This inspired the development of penalized linear regression methods such as ridge regression in 1970 and the lasso in 1996 [3,4]. The category "machine learning" included everything except the "logistic regression" category or other traditional statistical methods. We agree that support vector machines with linear kernels or Naïve Bayes have limited flexibility by design. Hence, it makes sense to rank algorithms by complexity or flexibility [5]. This is informative and helps researchers to choose an algorithm that is reasonable for the specific purpose at hand and in balance with the amount of data and subject knowledge available.

Ben Van Calster*
Department of Development and Regeneration
KU Leuven
Leuven, Belgium
Department of Biomedical Data Sciences
Leiden University Medical Centre
Leiden, The Netherlands

Jan Y. Verbakel
Department of Public Health and Primary Care
KU Leuven
Leuven, Belgium

DOI of original article: http://dx.doi.org/10.1016/j.jclinepi.2019.08.003.
Conflict of interest: none.
Funding: Research Foundation−Flanders (FWO) [grant G0B4716N]; Internal Funds KU Leuven [grant C24/15/037]; Cancer Research UK [grant 5529/A16895]; the NIHR Biomedical Research Centre, Oxford, UK.