EDITORIAL

Prediction of optimal outcomes in organ transplantation

Daniele Poole[1]* , Stefano Skurzak[2] and Mandeep R. Mehra[3]

Organ transplantation represents a durable therapeutic option for end-stage organ diseases. Long-term outcomes, cost-effectiveness and health utility benefits for transplantation remain unparalleled despite the emergence of novel alternative treatment paradigms across organ systems. Archetypes of such advances include mechanical cardiac support, lung pumps and the nascent field of genetically derived therapy and regenerative medicine [1]. Principles of organ transplantation center around three critical concepts: utility (maximize benefit, avoid harm), justice (diversity and equity), and individual respect (autonomy for decision making). Application of these principles has led to increasing indications, higher age limits and broader use of organs worldwide, with the exception of pancreas and intestine transplants [2]. Yet, maximizing these altruistic principles remains unfulfilled in some areas because problems of access, payment systems and allocation strategies create implementation heterogeneity. The ability to benchmark and predict outcomes to achieve greatest benefit constitutes a backbone for achieving the prime goals of transplantation.

Key drivers of successful transplantation balance adequate numbers of organs with appropriate allocation to the right individual. The first requires increasing awareness for organ donation and maximal recoverability of offered organs. The second task is optimal matching of donor organs to recipients under the utility, justice and autonomy framework [3]. This transplant benefit concept was introduced to equitably balance urgency (the risk of dying while awaiting an organ) with utility (life-years gained after transplant). Thus, economic modeling around this charter requires reliable prognostic models to predict wait-list survivorship and post-transplant morbidity and mortality. Although, for physicians, correct outcome predictions may not have immediate effects on single patient management, they provide standards to compare with to assess center performance, allowing the design of targeted corrective interventions of clinical practice.

Prognostic model generation has been facilitated by mandated registries for organ donation and transplant which collect epidemiological data, patient-level information, and allow timely monitoring of transplantation activity. Several organ-specific registries exist, supported by academic communities of various organ systems [the International Society of Heart and Lung Transplantation (ISHLT) for heart and lung, the European Liver Transplant Registry (ELTR) for liver, the European Renal Association – European Dialysis and Transplant Association (ERA-EDTA) for kidney, the International Pancreas Transplant Registry (IPTR) for pancreas, the International Intestinal Transplant Registry (IITR) for intestine], national and international collectives (Eurotransplant, Scandiatransplant, Collaborative Transplant Study) [4, 5]. While overall survival appears homogenous across different geographic areas in short (1 year) and intermediate terms (5 years) [5], marked heterogeneity of case-mix, access opportunity, processes of care, and allocation strategies persist. For this reason, direct comparison of outcomes may not reflect performances of different organizational models. Thus, generalizable prognostic models are essential to guide local resource allocation and practice.

In this regard, assessment of donor–recipient matching in terms of survival is quite complex. Among organ recipients, those less ill have longer survival compared to those at risk of imminent death, but comparisons made between patients with the same severity on the waiting

*Correspondence: daniele.poole@alice.it
[1] Anesthesia and Intensive Care Operative Unit, S. Martino Hospital, Belluno, Italy
Full author information is available at the end of the article

Springer
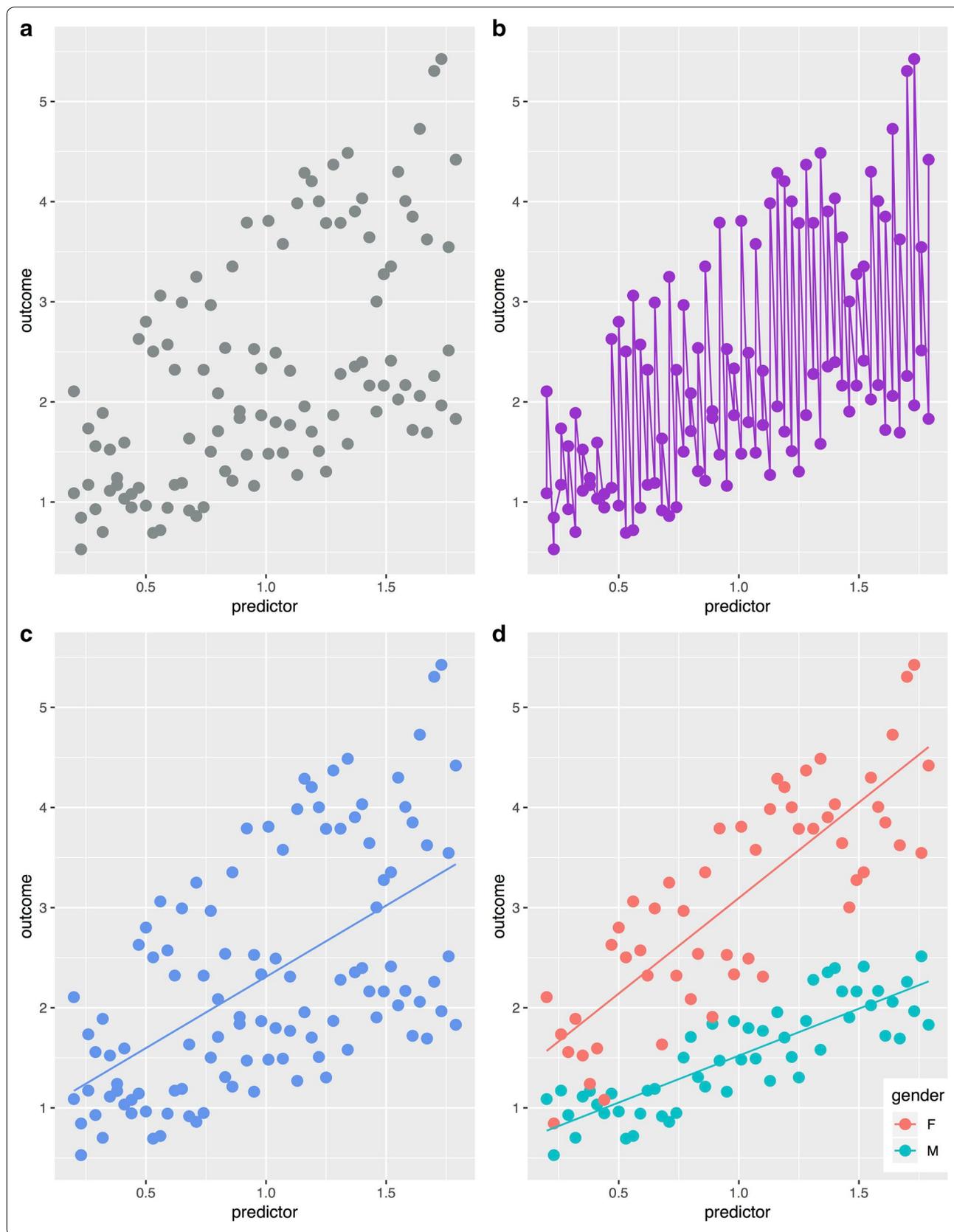
(See figure on previous page.)

Fig. 1 Let us assume that we are developing a model from a large body of variable data that has significant scatter (**a**). We may develop an overfitted model, which describes every element of the dataset at hand (**b**). We may also develop an underfitted model when we do not account for the most important predictors of outcome (**c**). Actually, we notice great variability around the regression line (the variability "explained" by the model, the R^2, is only 33%). Over- and underfitted models, for different reasons, will provide skewed predictions when tested on external cohorts (this is why external validation is mandatory before using a prognostic model). A good model, should account for most important predictors keeping an eye on parsimony. In our example, separation of risk by "gender" in the analysis, improves the model fit considerably (**d**; $R^2 = 79\%$)

list with relatively long survival, even if not transplanted, reduce observed benefits of transplantation. Another problem is that models with few variables do not maintain predictive ability when applied beyond the constraints of the development cohort (effectiveness gap in external validity). Such underfitted models fail to account for critical differences in the case-mix of individuals and may lead to erroneous decisions (Fig. 1c).

As an example, the Model for End-Stage Liver Disease [6] uses three variables while the Heart Failure Survival Score [7] includes seven parameters. Both stratify patients in severity to evaluate transplant candidates, but are heavily influenced by important unmeasured prognostic factors. The opposite problem, overfitting, occurs when excess variables are used to model common and rare features of the cohort (Fig. 1b).

Another crucial internal validity assessment step is to verify that models provide reliable predictions within subgroups defined by variables included in the model. If this condition is not met, the model will be skewed and will misrepresent the prediction outcome when tested on populations with a different proportion of the subgroup [8]. Further, discrimination (the ability to separate low- from high-risk patients) and calibration (which measures the closeness of observed event rates and those predicted by the model) should be assessed [9], the latter often being overlooked in transplantation. Models may have one and not the other of these two properties. For example, if in a cohort of patients we define predicted mortality using observed mortality (say, 25%), we will have perfect calibration (25% predicted vs. 25% observed), but the model will not be able to discriminate between survivors and non-survivors because they will all have the same score.

Finally, prognostic models in this field are best designed to measure short-term survival rather than long-term outcome. This is due to rapidly advancing paradigms of transplantation within which variables change their prognostic weight over time and previously unaccounted variables become more relevant as clinical practice evolves. One example is the Kidney Donor Risk Index [10], developed a decade ago on data collected over a 10-year time span.

The Scientific Registry of Transplant Recipients (SRTR) develops prognostic models for transplanted patient and graft survival, which are recalibrated every 6 months (the variables' prognostic weight is updated as new data become available) and completely overhauled at longer intervals [11]. SRTR was founded as a government request to monitor transplant activity in the US [12], while ISHLT runs an international registry from a clinician initiative [13], developing prognostic models with benchmarking outcomes but also to clearly enhance clinical practice.

Registries designed and conducted by clinicians may provide data of higher quality, which facilitate reliable prognostic model development. Even in such circumstances, shorter-term data, relevant to the early management principles of therapeutic success (and of importance to intensive care clinicians), remain more vigorous since they are more complete and less prone to missing values over time. Thus, reliability of prognostic models must be validated in external cohorts [14].

Under this perspective, the statistical approach used is less important. Popularity of unsupervised machine learning approaches is partly connected to the illusion that self-learning processes by a machine can lead to perfection. However, such perfect models, regardless of the statistical methods used, do not exist because the same data limitations of fit persist. It is essential to have a close alliance between clinicians and statisticians for use of tools to generate models that, in the end, are designed to continually enhance clinical outcomes [8, 14].

Author details
[1] Anesthesia and Intensive Care Operative Unit, S. Martino Hospital, Belluno, Italy. [2] Servizio di Anestesia e Rianimazione 2 Città della Salute e della Scienza di Torino, Turin, Italy. [3] Brigham and Women's Hospital Heart and Vascular Center and Harvard Medical School, Boston, MA, USA.

Compliance with ethical standards

Conflicts of interest
Prof Mehra is a consultant to Abbott, Medtronic, NupulseCV, FineHeart, Portola, Janssen, Bayer, and Mesoblast. No direct conflicts relevant to the current manuscript are inherent in these consulting agreements. He is also editor-in-chief of the *Journal of Heart and Lung Transplantation*. The views expressed are his own and do not represent the journal or the society that it represents, the International Society for Heart and Lung Transplantation. Daniele Poole and Stefano Skurzak have no conflict of interest.

References

1. Brody H (2016) Regenerative medicine. Nature 540:S49
2. GODT—Global Observatory on Donation and Transplantation. http://www.transplant-observatory.org/organ-donation-transplantation-activities-2015-report-2/. Accessed 15 Oct 2018
3. Persad G, Wertheimer A, Emanuel EJ (2009) Principles for allocation of scarce medical interventions. Lancet 373:423–431
4. GODT—Global Observatory on Donation and Transplantation. http://www.transplant-observatory.org/registries. Accessed 15 Oct 2018
5. CTS—Collaborative Transplant Study. http://www.ctstransplant.org. Accessed 15 Oct 2018
6. Malinchoc M, Kamath PS, Gordon FD, Peine CJ, Rank J, ter Borg PC (2000) A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. Hepatology 31:864–871
7. Aaronson KD, Schwartz JS, Chen TM, Wong KL, Goin JE, Mancini DM (1997) Development and prospective validation of a clinical index to predict survival in ambulatory patients referred for cardiac transplant evaluation. Circulation 95:2660–2667
8. Poole D, Carrara G, Bertolini G (2017) Intensive care medicine in 2050: statistical tools for development of prognostic models (why clinicians should not be ignored). Intensive Care Med 43:1403–1406
9. Poole D, Carlisle JB (2016) Mirror, mirror on the wall…predictions in anaesthesia and critical care. Anaesthesia 71:1104–1109
10. Rao PS, Schaubel DE, Guidinger MK, Andreoni KA, Wolfe RA, Merion RM, Port FK, Sung RS (2009) A comprehensive risk quantification score for deceased donor kidneys: the kidney donor risk index. Transplantation 88:231–236
11. Snyder JJ, Salkowski N, Kim SJ, Zaun D, Xiong H, Israni AK, Kasiske BL (2016) Developing statistical models to assess transplant outcomes using national registries: the process in the United States. Transplantation 100:288–294
12. Leppke S, Leighton T, Zaun D, Chen SC, Skeans M, Israni AK, Snyder JJ, Kasiske BL (2013) Scientific Registry of Transplant Recipients: collecting, analyzing, and reporting data on transplantation in the United States. Transpl Rev (Orlando) 27:50–56
13. Stehlik J, Bavaria JE, Bax J, Cronenwett JL, Edwards LB, Fairman RM, Ferrari R, Jacobs JP, Johnson MR, Kraiss LW et al (2016) Heart, lung, and vascular registries: evolving goals, successful approaches, and ongoing innovation. J Heart Lung Transpl 35:1149–1157
14. Altman DG, Royston P (2000) What do we mean by validating a prognostic model? Stat Med 19:453–473