Description, prediction, explanation

Description, prediction and explanation are all important in science. We welcome descriptive, predictive and explanatory studies, so long as the work is clear about its aims and uses appropriate methods to achieve its goals.

Traditionally, the social and life sciences have emphasized explanation: the identification of causal relationships between phenomena, such that intervening to change the cause would necessarily change the outcome. Furthermore, social and life scientists prioritize mechanistic evidence that can explain causal relationships between events or traits.

Description and prediction have traditionally had a secondary role in the natural and social sciences. Description has largely been considered valuable only insofar as it provides the starting point for causal inference. Prediction can be useful, but in and of itself has little to do with the pursuit of 'scientific truth' or the identification of laws that govern natural and social phenomena.

These historical emphases and preferences have also determined what is considered a 'significant scientific advance' in these fields — that is, the identification and accurate estimation of causal effects.

The past few years have seen a change in this attitude, partly because of the rise of big data and partly because of the life and social sciences gaining increased interaction with computer science and machine learning, where prediction is the central goal.

Editorially, we do not believe that descriptive studies are inherently of lower value or interest, especially when they uncover previously unknown phenomena or describe phenomena at scale through new measures and tools. Similarly, forecasting studies are invaluable in predicting future outcomes — for instance, predicting risk of disease, identifying young people at risk of poorer educational outcomes or predicting the impacts of anthropogenic climate change — when causal relationships are poorly understood or even when the relationships between effects cannot plausibly be causal.

If your manuscript asks a directional question (does *x* cause/impact/affect *y*?), but reports only correlational evidence, it will most probably be returned to you without review, explaining the reason for our decision. Experiments are the key tool for causal inference. However, for several of the key questions regarding human behaviour, manipulating the independent variables of interest experimentally may be unethical, illegal or unfeasible. In those cases, we will expect that authors make use of identification strategies developed for observational data — for example, difference-in-difference designs, regression discontinuity or instrumental variables[1].

If your manuscript aims to forecast future outcomes in a domain of broad interest and significance, we will expect that it includes out-of-sample validation of your predictions in an independent dataset if a suitable test dataset exists. In cases where no other suitable dataset exists, work can rely on cross-validation using the same dataset and partitioning the dataset into training and test components.

We value descriptive studies, especially when robust descriptions of specific phenomena are lacking or new phenomena of broad significance are discovered and the dataset is large and sufficiently diverse or representative. Mechanistic evidence is not a requirement for publication in those cases, nor is forecasting: although descriptive studies may form the starting point for causal inference or prediction in the future, this isn't a requirement for their publication. However, if the phenomenon in question has been well described in the past and the specific field expects mechanistic evidence as the next step, we will expect that the work goes beyond description.

Researchers have argued that the boundaries between prediction and explanation are far less sharp than traditionally conceived: identifying causal effects provides a basis for prediction of future outcomes in the same contexts. However, explanatory models are almost invariably built without consideration of predictive accuracy, especially beyond the specific context. Recent proposals have made a case for 'integrative empirical modelling' that combines causal inference and prediction of future outcomes[2]. We find these proposals valuable and strongly encourage the submission of research that makes use of integrative empirical modelling. ❐

Published online: 19 October 2021
https://doi.org/10.1038/s41562-021-01230-5

References

1. Marinescu, I. E., Lawlor, P. N. & Kording, K. P. *Nat. Hum. Behav.* **2**, 891–898 (2018).
2. Hofman, J. M. et al. *Nature* **595**, 181–188 (2021).