Received: 20 September 2018 | Accepted: 20 September 2018

DOI: 10.1111/eci.13031

COMMENTARY

WILEY

Precision medicine for individual patients should use population group averages and larger, not smaller, groups

Benjamin Djulbegovic[1] | **John P. A. Ioannidis**[2]

[1]Department of Supportive Care Medicine, Department of Hematology, Program for Evidence-based Medicine and Comparative Effectiveness Research, City of Hope, Duarte, California

[2]Stanford Prevention Research Center and Meta-Research Innovation Center at Stanford (METRICS), Stanford University School of Medicine, Stanford, California

Correspondence

John P. A. Ioannidis, Stanford Prevention Research Center, Stanford, CA.
Email: jioannid@stanford.edu

Funding information

Laura and John Arnold Foundation

Advocates of precision medicine propose that individual patients require evidence tailored to their singular, unique and personal profile. According to this way of thinking, population-based research, randomized clinical trials (RCTs) and evidence-based medicine (EBM) are often seen as outdated tools for application of research evidence to individuals.[1] Their major deficiency is claimed to be their insistence on making evidence-based recommendations based on "group averages" from populations (eg, trial data).[1] The public is warned that if your doctor recommends treatment based on a RCT, "it is time to find a physician who knows that you and averages are not the same."[2] Here, we argue that to accomplish its mission successfully, precision medicine must still use group averages; anything else would defy scientific principles of risk evaluation. The question is not whether to use group averages or not, but which are the best, reliable groups. In most medical situations, the informative groups would have to be larger and broader, not smaller, than what is done in current practice.

Risk is a group phenomenon and is knowable only as a population-based measure.[3,4] When we say 30% of patients are at risk of developing pulmonary embolism, we mean that if we consider 100 "similar" patients, 30 of them, on average, will develop a clot in their lung. The risk of the phenomenon of interest in various groups of patients is often knowable with sufficient accuracy.[3] For example, clinical decision tools such as Geneva or Wells criteria reliably assess the risk of pulmonary embolism in various population groups with different features.[5] This type of decision-making is referred to as decisions under risk.[4,6]

However, we can never say with perfect certainty which individual patient will develop pulmonary embolism. While risk assessment can be quite accurate at the level of population, any individual patient has probability of either 100% or 0% to have a clot.[3,4] That is, risk in any individual patient remains ultimately unknowable.[3] This is referred to as decision-making under uncertainty.[4,6] This is analogous to the famous thought experiment of quantum physicist Erwin Schroedinger: we cannot know if a cat placed inside of a box is dead or alive unless one opens the box and inspects it.[3] According to the quantum theory of interpretation of the probability, the cat is both dead and alive until the act of measurement occurs.[3] Similarly, a patient is believed to have both disease and no disease, or to respond to treatment or not until we assess what actually happened to the patient.[3]

Moreover, epistemic uncertainty regarding application of trial or other population data ("group averages") to individual patients represent a classic problem of inductive reasoning, which remains unsolvable on theoretical ground.[4] We can never be certain whether treatment effects we observed in the past will be seen in the next patient we see, and inferences from trial observations can never be assured when applied to yet-unobserved individual patients in clinical practice.[4]

The ultimate unknowability of the effects in individual patients does not mean that all risk predictions are equal. Indeed, the entire goal of precision medicine is to reliably reduce the population to smaller groups, in which risk can still be assessed with equal or higher accuracy than in the larger groups. However, unless the risk group becomes an n-of-1, the fate of individual patients remains unknowable.

There are very few situations in medicine where the risk category can reliably shrink to a single individual. Perhaps, the closest to this would be monogenetic diseases with

perfect penetrance where carrying a specific mutation practically confers 100% risk of developing disease. Even then, however, the exact manifestations of the disease, their timing, and what can be done for these patients may still vary for different individuals carrying the very same mutation.

The question is not whether group averages should be used, but how do we ensure that these group averages are unbiased and properly representative of the different patients to whom we want to apply them. Typically, the key challenge is whether we should strive to expand or shrink the boundaries of the group which is used to make inferences for our individual patients. Expanding the boundaries of the group that is considered directly relevant for decision-making increases the sample size and thus makes the estimates more precise (having smaller variance). A disadvantage is that more patients included in the larger group may deviate more from the patient to whom we want to apply our knowledge. Therefore, the trend in the precision medicine era is mostly to shrink the boundaries and make the group as similar as possible to the single patient at hand. However, making treatment or other management choices based on a severely restricted sample can be highly misleading. This is a reason why predictive models often fail external validity assessments—in our desire to develop fine-grained, precise risk estimates, the models typically overfit and generate wrong risk estimates when applied outside of the training data set.[7]

An increasingly common situation is that physicians and patients are tempted to search electronic medical records to find other patients who are matched to a very large number of characteristics. While an electronic health record database may include, for example, tens of thousands of patients with colorectal cancer, when a patient with colorectal cancer tries to match his own profile to other previous patients who not only had colorectal cancer, but also had the same values for 100 other characteristics, the eligible patients to compare against may be few, if any. Uncertainty grows to rampant levels, as the eligible sample becomes smaller. What happened to a few previous patients carries tremendous statistical uncertainty in application of such research evidence to future individual patients.

Worse, most of the "insights" derived in such small samples have major bias due to regression to the mean. The smaller the sample and the more spectacular the observed difference between two treatments options in this sample, the greater the expected disappointment from regression to the mean that will arise in the new patient(s) whom we want to treat. For example, suppose treatment A is more toxic and would be considered preferable to treatment B only if it is twofold better in clinical response. If treatment A is found to be slightly more than twofold better in clinical response than treatment B in a very small sample, the true difference will probably be much smaller. Hence, paradoxically the seemingly inferior treatment B based on the assessment from the small sample may be the best choice for the next (our) patient to be treated. Conversely, using methods to collapse risk categories in fewer clinically relevant categories may help improve risk assessment by including a larger number of patients who are similar enough to the patient we want to manage.[7]

Often, we use subgroup analyses to guide our treatments to individual patients. However, most subgroup difference claims are spurious.[8] Most of the time, the status of medical evidence is such that if we want to use the best evidence for the individual patient, we need to use larger samples from broader groups, not smaller samples from more restricted categories. The limiting factor to get best evidence is not the inability to narrow down to highly specific tiny subgroups, but the inability to have large enough samples with unbiased results, even when we broaden inclusiveness. This means that we typically need larger new population-based studies, and in particular, RCTs that tend to be more unbiased than non-randomized data, especially when randomization and allocation concealment are properly safeguarded. Even if only non-randomized data are available, the best decision may still be obtained by expanding the inclusion criteria to include more observations (and thus reduce uncertainty) rather than shrink the population considered to be relevant (and suffer from overfit). Either way, group averages would be used.

Precision medicine and EBM share an interest for the individual patient.[9] In fact, the main definition of EBM by David Sackett includes explicit reference to "making decisions about the care of individual patients." But it is unlikely we will be able to help individuals without having access to group averages from large, and ideally larger (not smaller), populations. Despite the uniqueness of each particular patient, there is simply no better method of treating individual patients but using research evidence obtained from groups of people. We think better research designs and more honest communication with patients would occur if the research community and patients realize this important limitation of our scientific inference rather than pursue unwarranted precision hype.

ACKNOWLEDGEMENTS

The Meta-Research Innovation Center at Stanford (METRICS) has been funded by the Laura and John Arnold Foundation.

CONFLICT OF INTEREST

None.

REFERENCES

1. Greenhalgh T, Howick J, Maskrey N. Evidence based medicine: a movement in crisis? *BMJ*. 2014;348:g3725.
2. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med*. 2018;210:2-21.

3. McEvoy JW, Diamond GA, Detrano RC, et al. Risk and the physics of clinical prediction. *Am J Cardiol*. 2014;113(8):1429-1435.
4. Djulbegovic B, Hozo I, Greenland S. Uncertainty in clinical medicine. In: Gifford F, ed. *Philosophy of Medicine (Handbook of the Philosophy of Science)*. London, UK: Elsevier; 2011:299-356.
5. Raja AS, Greenberg JO, Qaseem A, Denberg TD, Fitterman N, Schuur JD. Evaluation of patients with suspected acute pulmonary embolism: best practice advice from the clinical guidelines committee of the American College of Physicians evaluation of patients with suspected acute pulmonary embolism. *Ann Intern Med*. 2015;163(9):701-711.
6. Knight FH. *Risk, Uncertainty, and Profit*. Boston, MA: Hart, Schaffner & Mark; Houghton Mifflin Co; 1921.
7. Woike JK, Hoffrage U, Martignon L. Integrating and testing natural frequencies, naive Bayes and fast-and-frugal trees. *Decision*. 2017;4:234.
8. Wallach JD, Sullivan PG, Trepanowski JF, Sainani KL, Steyerberg EW, Ioannidis JA. Evaluation of evidence of statistical support and corroboration of subgroup claims in randomized clinical trials. *JAMA Intern Med*. 2017;177(4):554-560.
9. Ioannidis JPA, Khoury MJ. Evidence-based medicine and big genomic data. *Hum Mol Genet*. 2018;27(R1):R2-R7.

How to cite this article: Djulbegovic B, Ioannidis JPA. Precision medicine for individual patients should use population group averages and larger, not smaller, groups. *Eur J Clin Invest*. 2018;e13031. https://doi.org/10.1111/eci.13031