



Precision medicine in the clouds

Vogt, Henrik ; Green, Sara; Brodersen, John

Published in:
Nature Biotechnology

Publication date:
2018

Document version
Peer reviewed version

Citation for published version (APA):
Vogt, H., Green, S., & Brodersen, J. (2018). Precision medicine in the clouds. *Nature Biotechnology*, 36(8), 678-680.

Precision medicine in the clouds

To the editor:

A billion-dollar question is whether precision medicine (aka personalized, “P4” or systems medicine) can substantially increase the utility of individualized disease prevention and population health^{1,2}. In this respect, the first results from the “Pioneer 100 Wellness Project (P100)” featured in last August’s Nature Biotechnology issue, is a landmark^{3,4}. The study sheds light on an approach that has primarily existed as a vision and precedes the US National Institutes of Health’s (NIH; Bethesda, MD) “All of Us Study”, which will include a million participants in a similar scheme (<http://www.allofus.nih.gov/>). The researchers behind the study claim to demonstrate how measurement of personal data clouds over time can improve our understanding of health and disease and to identify “actionable possibilities”, by which individuals can enhance health through preventive strategies.

P100 is an exploratory study of associations in networks of biomarkers and risk factors established from large and dynamic data clouds of 108 participants. Over the course of a 9-month period, participants underwent whole genome sequencing (yielding 127 polygenic scores for disease risks plus three copy number variations), three-times testing of metabolome (643 metabolites), proteome (262 proteins) and microbiome (4616 taxonomic units), 218 other clinical tests and measurements, as well as daily activity tracking via “quantified self” technologies. Associations between these biomarkers yielded a total of 3470 connections in a correlation network. The P100 study is a project designed to display the potentials for novel technologies and gather support for precision or P4 medicine (predictive, preventive, personalized and participatory), and it has gained prominent publicity in *Nature*⁵. The P100 will also be scaled up to include 100,000 participants in the 100K Wellness Project³. Against this background, we here discuss whether the P100 study actually supports the prospect of substantial benefits from data-driven disease prevention, and argue that it exposes severe challenges.

The graphics of the correlational networks presented by Price *et al.*³ do offer an interesting research potential for exploring connections in molecular networks and for identifying candidate biomarkers³. However, as yet there are only a few examples where such

approaches have led to novel discoveries of therapeutic potential, especially when it comes to disease prevention which is the main aim of P100. Systems medicine has very ambitiously been promoted as a “holistic” approach that will “tackle all components of the complexity of non-communicable diseases”⁶. It is, of course, not obvious that scientific models can ever encompass *all* factors that will determine an individual’s health. The P100 also illustrates that the purportedly “holistic” ambition of quantitative systems medicine is far from being realized as it employs a reductionist method defined less by true integration than collection and correlation of diverse, but mostly molecular, data.

In terms of clinical research method, the P100 does not involve a classic randomized trial with a control group (RCT). Instead its leading researchers opt for an ambitious, but ill-defined n-of-1 approach where each participant is extensively monitored and considered his/her own control. Quote Leroy Hood: “We hope to develop a whole series of stories about how actionable opportunities have changed the wellness of individuals”⁵. In line with this reasoning, the authors present changes in clinical biomarkers during the 9-month study as evidence for clinical utility. However, to measure an unbiased, valid effect in an n-of-1 randomized clinical trial, the study would have to include pairs of organized periods so that one period of each pair applies the experimental therapy and the other period applies usual care or placebo, and both the clinician and the patient have to be blinded⁷. In the P100 study, these strict methodological criteria do not apply. For example, this makes it difficult to examine whether people alter their behavior in response to just being observed (the Hawthorne effect). Proponents of preventive precision medicine may argue that there are other ways of providing evidence for treatment efficacy as the number of measured variables increases and the number of research subjects that are similar enough for a personalized approach, moves towards $n=1$. This may involve continuous monitoring to observe significant changes in each particular person. The prospect of using such big data “narratives” as evidence is intriguing. However, the P100 study fruitfully highlights how important epistemological and methodological questions must be clarified and answered before it is clear how it is determined who is at risk and what treatments actually work.

In general, it is not clear what the key concepts of “actionability” or “clinical utility” refer to in the P100 study. We do not know how much action the participants were actually able to take when faced with “actionable possibilities”, or how useful it was for them. The competence

of the health coaches and the rationale for their “personalized” health recommendations are unknown, although the latter are described as being “evidence-based” and mostly consist of advice on diet, exercise, stress management, dietary supplements, or physician referral.

Importantly, the effects are measured mostly in terms of correction of surrogate endpoints. As a general point, such changes provide no guarantee for changes neither in the hard endpoints of morbidity and mortality, nor in quality of life. As an example, a study carrying out a 10-year follow-up on the large, five-year Inter99-study⁸, which tested lifestyle counselling based on screening for risk factors, showed no significant difference between the intervention and control groups with respect to ischemic heart disease, stroke, or mortality, despite successful changes in risk factors as in the P100 study⁹.

In P100, changes in vitamin D levels and markers for pre-diabetes are interpreted as evidence of health benefits. 88% of the participants were labeled as out of range at baseline in their vitamin D levels. Although the participants “out of range” managed to increase their vitamin D levels with 7.2 ng/mL per round, it is questionable whether vitamin D supplements change hard endpoints¹⁰.

As another staggering figure, 48% of the previously well individuals were deemed to have “out-of-range” HbA1c levels (glycated hemoglobin) and were labeled as pre-diabetic. Again, what these participants would gain from lowering HbA1c by 0.085% per round is unclear. A third example is mercury levels. It is blatantly unclear how valid and useful it is to label 81 out of 108 participants with a mercury problem that on average was lowered by 0.002 mcg/g/round.

More generally, vitamin D deficiency and pre-diabetes are examples of controversies in preventive medicine, both being criticized for questionable clinical utility, connections to alternative medicine, overmedicalization and unsustainable burdens on the healthcare system^{11,12}.

One case story is given special attention as evidence of potential benefit in the P100 publication³. A 65-year-old male presented symptoms to his physician and was found to have abnormally high ferritin levels. Genetic testing showed that he was homozygous for the primary genetic risk factor for hemochromatosis. In a commentary⁴, Butte claims that this particular example teaches us that “genomes can provide useful medical information to healthy individuals”. However, testing *after* experience of symptoms is a very different matter from

screening asymptomatic people, which is what the P100 is otherwise about. It is not clear that the man had fared any worse with regular primary care. It should also be noted that generalized screening of asymptomatic individuals for hereditary hemochromatosis is yet another controversial case in preventive medicine¹³.

Another important question is whether precision medicine can actually “empower” people to change their lifestyles in response to risk information. Although there are studies showing some positive results, a general trend in available evidence is that lifestyle interventions and health checks among the apparently healthy are of limited benefit^{8,9,14–16}. One common and difficult challenge is that people simply often do not follow recommendations. Proponents of prevention precision medicine may argue that a personalized approach is needed precisely to provide the individual with more accurate risk assessments and advice so that they may change their behaviors according to *their* particular case. However, a recent systematic review with meta-analysis has showed that there is no clear evidence for a significant positive impact of communicating individualized genetic risks of disease on risk-reducing health behaviour¹⁵.

P4 medicine now promises to change this picture through *continual*, multi-level feedback on personal health. However, the P100 illustrates that the approach may be up against quite fundamental challenges: Only 64% of the participants complied with Fitbit activity tracking for 40 days out of 9 months (and even fewer for sleep tracking)—a striking result considering that the participants (largely middle-class volunteers) were presumably unusually motivated. The prospect of significant benefits from the preventive precision medicine should also be considered in light of increasing evidence that risk information and extensive self-tracking generally have little effect on health-related actions¹⁷⁻¹⁸.

This is not to say that counselling or coaching based on risk estimates and screening does not have an effect for *some* people, as several studies show⁹. However, taken together, such results call for modesty concerning the prospects of extensive self-monitoring to improve population health greatly, especially in providing strategies that benefit those who live their lives in health-detrimental environmental situations, who are at high risk and in the greatest need of medical care¹⁹. As the authors of the abovementioned Inter99 study state, the implication of the generally weak results of screening and lifestyle counselling is that health behavior depends on the

individual's social context and life conditions and that interventions should therefore be mostly be structural rather than targeted at individuals⁹.

Although extensive measurement and testing may have some benefits, these must be weighed against harms and costs. The P100 study seems solely geared towards showing benefits, we must consider possible downsides of this approach on a theoretical basis.

The most important finding in the P100 study may be that *all* of the 108 participants were found to have multiple “actionable possibilities” and thus were labeled as in need of medical attention. To define *everyone* as at risk is clearly not very precise. It is also potentially harmful. To underscore, these people considered themselves well at baseline, but upon entering the study, they were all regarded as in *suboptimal* health. This entails a change in identity, and may result in pathologization and worry^{20,21}.

Crucially, the P100 highlights the problem of *overdiagnosis* (i.e., discovery of *real* but asymptomatic abnormalities that will never become symptomatic health problems). Overdiagnosis is a serious problem because it often leads to costs and harms associated with overtreatment and further over-testing, while it cannot, by definition, benefit those who would never develop symptomatic disease.

Recent meta-analyses show that the problem of overdiagnosis is not only due to a lack of predictive power in medical testing, but to medical overactivity more generally. The problem increases with lowered diagnostic thresholds, intensified screening for early disease and risk states, diagnostic technologies with higher resolution, and through more frequent and intensified measurements of multiple parameters²². The P100 approach includes all of these potential “drivers of overdiagnosis”. Previously published papers commenting on the P100 study have raised concerns about the increased risk of overdiagnosis^{1,23,24}. The P100 researchers acknowledge false-positive tests, but have not addressed overdiagnosis, neither in previous papers nor in the current publication^{3,25}.

We should also consider opportunity costs: The P100 strategy means a large investment for each participant (and society), both in terms of money, time and attention. What could have been gained in terms of health and other benefits had the resources been used otherwise?

What have we learned from the P100 study? This question is difficult to answer. The study does not follow available guidelines for assessment of public health impact of medical

testing. This is particularly disconcerting as the project is the pilot for the much larger 100K Wellness Project. P100 promotes a novel concept that is branded as “scientific wellness”. However, in practice the study represents a form of data-driven screening for early disease and risk states that entails an unprecedented over-medicalization in terms of measurements and interventions²⁴. The benefit the previously well participants reap from this is literally in the (data) clouds.

The P100 study highlights the need to discuss how high a price we are willing to pay to explore the potential of P4 medicine, and whether we currently have unrealistic expectations to the new technologies. Overstating the potential may influence personal and political choices, and divert attention and resources from strategies that may be more effective. Researchers who develop technologies for preventive medicine should take its profound challenges seriously—especially overdiagnosis. To make informed choices about the future of healthcare, we need studies that substantiate what we may call “actionable evidence”, i.e., studies showing changes in hard endpoints that take seriously the risk of unintentional harm. In the context P4 medicine or precision medicine, the balance between promises and evidence should be better calibrated.

Acknowledgement: Henrik Vogt wishes to thank professors Linn Getz and Irene Hetlevik at the General Practice Research Unit, Norwegian University of Science and Technology, as this article builds on his ph.d. project which they supervised.

Henrik Vogt¹, Sara Green² & John Brodersen³

¹University of Oslo, Centre for Medical Ethics, Institute of Health and Society, Oslo, Oslo 0318, Norway

²Section for History and Philosophy of Science, Department of Science Education & Centre for Medical Science and Technology Studies, Department of Public Health, University of Denmark, Copenhagen, Denmark

³*Section of General Practice & Research Unit for General Practice, Department of Public Health, University of Copenhagen, Copenhagen, Denmark*

Correspondence should be addressed to e-mail: vogt.henrik@gmail.com

- 1 Green, S. & Vogt, H. *Humana.Mente Journal of Philosophical Studies* **30**, 105---145 (2016).
- 2 Khoury, M. J. & Galea, S. *JAMA* **316**, 1357---1358 (2016).
- 3 Price, N. D. *et al. Nat Biotechnol* **35**, 747---756 (2017).
- 4 Butte, A. J. *Nat Biotechnol* **35**, 720---721 (2017).
- 5 Gibbs, W. W. *Nature* **506**, 144---145 (2014).
- 6 Bousquet, J. *et al. Genome Med* **3**, 43 (2011).
- 7 Guyatt, G., Zhang, Y., Jaeschke, R. & McGinn, T. in *Users' Guides to the Medical Literature: A Manual for Evidence--Based Clinical Practice* (eds G. Guyatt, D. Rennie, M. Meade, & D.J. Cook) (McGraw---Hill Education, New York, 2015).
- 8 Aadahl, M. *et al. Prev Med* **48**, 326---331 (2009).
- 9 Jorgensen, T. *et al. BMJ* **348**, g3617 (2014).
- 10 Bjelakovic, G. *et al. The Cochrane database of systematic reviews*, CD007470 (2014).
- 11 Welsh, P. & Sattar, N. *BMJ* **348**, g2280 (2014).
- 12 Yudkin, J. S. & Montori, V. M. *BMJ* **349**, g4485 (2014).
- 13 Nadakkavukaran, I. M., Gan, E. K. & Olynyk, J. K. *Pathology* **44**, 148---152 (2012).
- 14 Krogsboll, L. T., Jorgensen, K. J., Gronhoj Larsen, C. & Gotzsche, P. C. *The Cochrane database of systematic reviews* **10**, CD009009 (2012).
- 15 Hollands, G. J. *et al. BMJ* **352** (2016).
- 16 Look Ahead Research Group. *The New England journal of medicine* **369**, 145--- 154 (2013).
- 17 Bloss, C. S. *et al. PeerJ* **4**, e1554 (2016).
- 18 Jakicic, J. M. *et al. JAMA* **316**, 1161---1171 (2016).
- 19 Burke, W. & Trinidad, S. B. *Genome Med* **3**, 47 (2011).
- 20 Hansson, A., Brodersen, J., Reventlow, S. & Pettersson, M. *Health, Risk & Society* **14**, 148---152 (2012).

This is a preprint version of a correspondence article published in Nature Biotechnology. Please see the following link for the final version: <https://www.nature.com/articles/nbt.4210>.

21 Jorgensen, P., Langhammer, A., Krokstad, S. & Forsmo, S. *Family Practice* **32**, 492---499 (2015).

22 Moynihan, R., Doust, J. & Henry, D. *BMJ* **344** (2012).

23 Diamandis, E. P. *BMC Med* **13**, 5 (2015).

24 Vogt, H., Hofmann, B. & Getz, L. *Medicine Health Care and Philosophy* **19**, 307--- 323 (2016).

25 Hood, L., Lovejoy, J. C. & Price, N. D. *BMC Med* **13**, 4 (2015).