Big hopes for big data

Large-scale multi-modal information on patients' health is ever increasing, providing an opportunity to use big data for taking individualized medicine to a global scale.

There has been an explosion of healthcare-related data in the past decade. With the digitalization of medical records, increasing affordability of molecular testing, advent of medical informatics and widespread use of wearables, the sheer volume of data available for analysis is staggering. Enter big data. Clever analysis of these terabytes of information is already making inroads at the individual's level——helping to redefine what means to be 'healthy', uncovering unknown disease risk factors and allowing more-accurate diagnostic and prognostic predictions. At this turning point, we bring our readers a special Focus that highlights the potential of big data–based approaches to affect health beyond the level of the individual, by means of innovative solutions to also address challenges in health systems at the planetary level.

Big data, by nature, is infinitely versatile—it is as powerful as the datasets being brought together and the ability of algorithms to make the most of it. In this issue, a News Feature gives a bird's-eye view of the incredible breadth, depth and scale of big data. A Review by Eran Segal and colleagues critically discusses perspectives and challenges to be overcome for the use of big data in health to be truly beneficial. Michael Snyder and colleagues demonstrate how combining longitudinal clinical data with deep molecular profiling of healthy participants can reveal unforeseen ageing markers that can be used to improve management of health. Two articles in this issue, one by Majid Ezzati and colleagues and another one by a group led by Elizabeth Sowell, show how much can be learned by bringing healthcare datasets with factors that are not typically recorded in hospital settings. Their work unveils new associations at the intersection of individual-level health parameters and records of atmospheric temperature or lead exposure. Such studies reveal environmental determinants of health that can inform public-health policies to elevate population well-being and reduce inequalities.

The advent of new technologies is bringing innovation and expanding the possibilities of clinical research by providing new solutions for trial participants and bringing together disparate populations. Molecular information is increasingly incorporated into mainstay clinical oncology to match patients with therapies—what is commonly known as 'precision medicine'. That information is incredibly valuable to researchers and provides new insights into the mechanisms of disease and treatment response. However, tumor sequencing has so far been accessible only in major cancer centers and has thus remained inaccessible to many. Direct-to-consumer genetics company 23andMe has announced they will be leveraging their data to boost patient recruitment in clinical trials at local care centers in the USA, thereby reducing geographical barriers to state-of-the-art clinical research for the broader population. Harnessing increasingly widespread wearable devices to generate clinical data could reduce the costs and logistics of medical research, as recently shown by a study using data from an internet-connected smartwatch to detect atrial fibrillation. Finally, the Count Me In initiative, a framework of patient-partnered research that allows patients diagnosed with angiosarcoma across the USA and Canada to securely share their medical records, samples and genetic information with researchers, exemplifies how this approach can accelerate research for patients with rare conditions who are in need of more-effective treatment approaches.

With the collection and treatment of the data on this scale, challenges need to be faced and pitfalls need to be avoided. In its relatively recent history, big data has already shown that failure to notice these shortcomings can have dire consequences on health equity. Computational models can be very powerful—they can harmonize health parameters and increase the granularity of an individual's health status. However, in their design, these models must be devoid of bias or they could exacerbate, rather than reduce, racial, gender, economic or geographical disparities in care—as discussed by Emily Courey Pryor and colleagues in this issue. Moreover, to prevent misuse of personal data, specific legislation will need to be put into place to guarantee that medical data are being handled securely.

As big data makes its way into clinical care, expectations are high. Green shoots of what big data could do for medicine are appearing; digital innovation and variety of healthcare information are being leveraged to reveal novel insights into human health and disease. Going forward, we encourage the global research community to think outside the box toward a novel way of considering health-relevant datasets, toward a bigger scale of thinking and toward a new way to generate data, always putting equity of health first and foremost. We look forward to bearing witness to this next revolution. ❐