
PANORAMA PEOPLE

Interview with Professor Christopher J.L. Murray, Director and co-founder of the Institute for Health Metrics and Evaluation (IHME) and the Chair of the Health Metrics Sciences Department at the University of Washington.



Professor Christopher J.L. Murray holds Bachelor of Arts and Science degrees from Harvard University, a DPhil in International Health Economics from Oxford University, and a medical degree from Harvard Medical School. Dr. Murray is a founder of the Global Burden of Disease (GBD) Study, a systematic effort to quantify the comparative magnitude of health loss due to diseases, injuries, and risk factors by age, sex, and geography over time. His career has focused on improving health worldwide by providing the best information on population health.

The memorandum of understanding between IHME and WHO is the foundation for the collaboration between the two organizations. In the words of Dr Tedros Adhanom Ghebreyesus, WHO Director-General: “This new agreement highlights our shared commitment to ensure that health policy is based on the most accurate and up-to-date data available.” For public health experts, evidence based on sound data is key to health policy-making. But data and evidence seem to have dwindling significance in this time of “alternative facts”. How can the collaboration between WHO and IHME contribute to reversing the trend of alternative facts being used in a very broad setting?

I think that the rise of “alternative facts” is about trust and credibility. People – the general public, parliamentarians, and decision-makers – are struggling to know whom to trust for information. It’s not that people do not want evidence. Rather, it’s that they do not know whom to believe, and they may discount evidence that disagrees with their views because they think the source has a political agenda. I think that WHO, together with IHME, can help by marshalling evidence, trying hard to be transparent about its basis, and collectively establishing our work as a trusted source. Even

in the public health sphere we have had some major reversals on recommendations, not because of any agenda, but just because the evidence was not as strong as people thought. For example, in the 1980s and 1990s, saturated fats were identified as a risk for heart disease, but evidence accumulated since has not clearly shown this risk. This type of reversal is also contributing to this lack of confidence in some evidence.

So I think it becomes important as we go forward to think about how we score evidence. There are some things that we can be definitive about, such as the number of people in Denmark dying from one type of cancer or the association between smoking and lung cancer. But there are other things about which we can be much less definitive. If we can become better at communicating the strength of evidence, we can, I hope, avoid these huge changes, which are inevitable in science, but which undermine public confidence in what we do. In my mind this is something on which, by working together, we could make a lot of progress in, across a whole spectrum of issues.

There is trust, there is also communication. Would you say that there is also health literacy – people being unable

to interpret the information that they see? This is also something that has become more prominent recently. Do you feel that there is also a role for WHO and IHME in health literacy?

I think we should never blame the public for not understanding; the responsibility should always be on the technical, scientific and expert community. If we are not getting our message across, then we need to try other ways to do so. This is an important issue in the digital era. There is a large segment of the more-educated population who spend a lot of time on the Internet looking for information about health. As a result, we also need to think about how we get our joint, well-curated evidence high up on search engines, for example. Simple things like that. I think that the onus is on us to figure out how to communicate, how to make sure that when someone is looking for evidence, they will be directed toward well-curated and well-graded evidence.

The GBD approach takes a holistic view of health, considering interdependencies between diseases and determinants of health. What is the potential of the GBD approach and findings, and its use in policy-making in contributing to WHO reaching its targets of the Thirteenth General Programme of Work (GPW 13)?

I think of the GBD as a holistic, comprehensive view of disease, injuries and risks, with a highly standardized approach. There is a lot of good work by experts in various subfields, and GBD brings focus on comparability across conditions, diseases and communities. It tries to address all the biases in different types of data in a very systematic way. If you think about the triple billion targets in the GPW 13, that comparability is really important. If WHO is really going to contribute to a billion more people receiving UHC, then we need to make sure that the measurements across communities, across countries, and over time are meaningful. Recently, there has been a strong emphasis on transparency, especially with the advent of WHO-led GATHER (Guidelines on Accurate and Transparent Health Estimates Reporting). It will be important to ensure that all measurements under GPW 13 are GATHER-compliant, that they result from consultations with countries, and that they are clearly attributable to reference years, in order to be useful in monitoring Member States' progress and WHO's contributions to that progress on the triple billion goals.

We fully agree with you on that. And this is also why in the European context we've started the European Burden of Disease Network, co-hosted with IHME, in which people can exchange technical expertise and views, but which is

also producing a national burden of disease (NBD) manual. How strongly do you feel such a network contributes to the harmonization of information, in the Region or maybe even globally, and what would you expect from a standardized NBD manual, which may be helpful to countries having to report on GPW 13?

With the explosion of the scientific collaboration running the GBD study based at IHME—now exceeding 4000 individuals in 150 countries, and the impressive rise of country-led subnational burden of disease studies using the same GBD-standardized case definitions, methods and approaches, the GBD has become a movement rather than just one study. I think it is really helpful having the European Burden of Disease Network as a vehicle for further accelerating that natural growth in subnational burden of disease studies, and a GBD manual, would be an important step in making it easier for countries to pursue these studies. As the GBD approach evolves so does the growth of related analyses, using, for example, the curated data from the GBD exercise for understanding health systems. We are seeing more of that sort of work. For example, Martin McKee in Europe has been guiding and leading the work on using the GBD approach to understand health-care access and quality. We are seeing many of these GBD extensions and as we make progress on the transparent grading of evidence on risk factors, so will there follow more local applications, for people interested in risks or outcomes relevant to specific settings. The network and the manual will help to accelerate more of what has become a worldwide movement towards these analyses under the broad rubric of the GBD.

The GBD is a leading movement and we should make sure that we have a platform for such a movement. Do you feel that the European network is achieving this aim? Is this the kind of initiative that can perhaps be globally established? What is your view on that?

The GBD approach and study is appealing to many, but with its large number and diverse set of collaborators one of the big advances and evolutions has been how to govern a group of scientists with different opinions on many different topics in a way that still leads to standardization and a highly protocolled approach to measurement. There are many tools that try to make sure that all the different scientific viewpoints are heard and yet common core standards are retained. And that is why we have this elaborate governance for the GBD. The Scientific Council studies all decisions across the 4000 collaborators, whenever there is a disagreement. The council takes those decisions. We have a mechanism involving a completely independent oversight group, chaired by Peter

Piot at the London School of Hygiene and Tropical Medicine, with a diverse global membership. The members meet every six months to review and make suggestions on methods and new directions. Moreover, we have other mechanisms that are also helpful for soliciting and channeling technical feedback and ideas to strengthen the GBD, for example, the European Burden of Disease Network. Given the whole process of scientific peer-review and exchange, these mechanisms are important for maintaining what is vital: a transparent collaboration that any legitimate analyst or scientist can join.

These mechanisms together comprise a healthy environment for the GBD and I think that a recent and very important factor is the collaboration with WHO: WHO is able to convene on regular basis, the world's expertise, or a region's expertise, on a disease or a risk, or a topic. So making the GBD a joint effort with WHO is yet another way to obtain the best scientific critique. That's what we want; to make sure that every scientific critique, new direction or new request is heard, and that there is a robust scientific discussion across the collaboration.

At WHO we frequently get the question: how are we applying new technologies? The GBD approach is rooted in complex models and has evolved significantly over the years. Complex modelling and the use of various algorithms have not only become the backbone of computing in the 21st century, but are also something that the general public has become aware of through such things as “artificial intelligence” or “machine learning”. As scientific professions, public health and medical research are both driving and driven by these new technologies, how do you see the future evolution of the GBD in this context and which developments might have the most significant impact?

We are already seeing a lot of testing and even the application of machine learning in the GBD. Given the thousands of people involved who have a quantitative background or focus, it is not surprising that innovative ways have been put forward. One of those already being deployed, and which will only get better over time, is machine vision. The most advanced aspect of machines may be learning where you can replicate, or do better than a human or a panel of humans can in identifying problems in models and data. For example, there are thousands of outcomes measured as a part of the GBD research each year – over time, by country, for national and subnational estimates – and people need to look at those graphs to identify where the model and the data do not align well. And research shows that machine vision, as long as we create a decent training data set, can do that extremely well.

The second area that everyone is rushing to apply is, of course, substituting classical statistical models, whether Bayesian or not, with machine learning algorithms. And often machine learning algorithms are better at out-of-sample predictions. The challenge – and this is what everybody is grappling with – is that we have a machine learning algorithm that may predict out-of-samples better, but we are not 100% sure why. Many people are not convinced about the picking up of associations that are likely not causal or related. We have examples where a machine learning algorithm might say that the association between tobacco and an outcome is the opposite of what it should be. Statistically you can demonstrate that this is a better prediction because, say, in this case, smoking is a confounding factor or a predictor associated with something else that is causal. But this leads us into a very uncomfortable area of simply using predictive accuracy as the basis for adopting machine learning. We are in that phase right now; the whole machine learning–causality debate is under rapid evolution. There are various hybrid approaches people are considering, and other types of applications for machine learning, in which it may not be essential to get the causality right. But this is a highly active area in which we are going to see constant evolution.

If we go back to the first question that we were talking about – trust and credibility – we will need to be sure of ourselves to move from machine learning to help us detect errors, to machine learning as a substitute for statistical models. There is a large risk of getting the causal connections wrong and, therefore, potentially undermining confidence in the results. But 10 years from now, I am sure we are going to have many more applications of machine learning, within the huge space of the GBD analyses. I think the tools are getting better and better and we are starting to use hybrid approaches that rule out certain types of predictors that are essentially the wrong cause in the relationships. There is a lot of energy right now in the whole machine learning area and at the same time quite a lot of skepticism about adaptive causality.

Is there anything else you would like to add?

I think there is huge potential for furthering evidence and filling that niche where people want a trusted source, through WHO and IHME collaboration, which goes beyond the traditional remit of the GBD study. We were asked a couple of years ago by the independent advisory committee, chaired by Peter Piot, to progressively introduce star rating systems for different types of evidence. We have introduced it, for example, to grade the strength of cause of death data based on objective criteria and we've been trying to foster a system on grading the strength of evidence for different risk–outcome pairs, for example,

vegetables and ischaemic heart disease, or smoking and lung cancer. The amount and strength of evidence for smoking and lung cancer is much higher than for other risk–outcome pairs and we are developing a highly standardized and objective grading system for strength of evidence, which we hope to start rolling out this year. Once you go down that path of grading evidence and see the diversity of the strength of the evidence, you can imagine that this standardized approach could, or should, also be applied to health and policy interventions. That to me is one of the most exciting agendas for the future: taking the tools that we can develop, focus on standardization, and expanding that into the space of evaluating policies and interventions. This is relevant to WHO and IHME collaboration because one of the components of the MOU, one I think Dr Tedros is the most excited about, is the collaboration on country policy dialogue. That is, using the work IHME has been doing

on forecasting future health scenarios and twinning it with a good sense of evidence on different interventions, which you can trace out for countries. This raises questions, such as what are the scenarios that can be pursued for which the evidence is strong or for which the evidence is perhaps not as strong but the benefits could be quite large? There is great potential there. We have already seen in the early versions of this work, in showing it to various governments, a lot of enthusiasm for that mixture of evidence and forecasting, twinned with the evidence on the interventions for tackling the health problems faced by countries.

Disclaimer: The interviewee alone is responsible for the views expressed in this publication and they do not necessarily represent the decisions or policies of the World Health Organization. ■