Published: November 2015

The importance of communication

Brian Williams - Epidemiologist affiliated to SACEMA.

Communicating science to a lay audience is never easy but communicating science to other scientists can be almost as hard. Through the SACEMA Quarterly we hope to help scientists to learn to communicate with others even in just a small way.

More than 50 years ago Robert Oppenheimer, one of the giants of 20th century physics, reflecting on the state of science wrote 'Today [as opposed to Plato's Greece], it is not only that our kings do not know mathematics, but our philosophers do not know mathematics and—to go a step further—our mathematicians do not know mathematics. Each of them knows a branch of the subject and they listen to each other with a fraternal and honest respect; and here and there you find a knitting together of the different fields of mathematical specialization … We so refine what we think, we so change the meaning of words, we build up so distinctive a tradition, that scientific knowledge today is not an enrichment of the general culture. It is, on the contrary, the possession of countless highly specialized communities who love it, would like to share it, and who make some efforts to communicate it …. Out of this have grown the specialized disciplines like the fingers of the hand, united in origin but no longer in contact' (1).

Oppenheimer's reflections have even greater relevance now, when the advances in modern science have given us a detailed understanding of the origin of the universe, the fundamental building blocks of matter, the genetic code, the way in which viruses manipulate our DNA to ensure their survival, cures for a myriad of recently incurable diseases and much else besides. But the price of the specialization that this entails has been an increasing inability to understand our colleagues and draw on their expertise in our own areas of research.

Consider these two observations. The first is the demonstration, based on six studies, with follow-up times ranging from 4 to 12 years, that the incidence of TB declines exponentially with body-mass-index (BMI) at a rate of 13.8% ± 0.4% for each unit increase in BMI independently of age, race, gender, country or even smoking status (2). A person with a BMI of 35 kg/m^2 is 12.0 (11.2±12.9) times less likely to develop TB than a person with a BMI of 17 kg/m^2. So BMI is a powerful predictor of a person's risk of developing TB, but what determines this extraordinarily precise rate of decline with BMI? Presumably BMI is associated with a change in the host's immunity that leads to this dramatic and predictable exponential decline in the risk of developing TB with increasing BMI. But, at the time of writing, no-one has been able to offer any kind of explanation.

The second is the relationship between survival and age in people infected with HIV but not on ART. Data from the CASCADE cohort (3) show that the survival of people infected with HIV, but not on ART, follow a Weibull distribution with a shape parameter of about 2.25 across all ages (4). Even more striking is the fact that the median survival declines linearly at a rate of 2.00 ± 0.06 months for each one year increase in age (4) so that children infected at the age of 2.5 years will live for a median of 15.8 years while people infected at the age of 65 years will live for a median of 5.4 years. But the reasons for this simple, linear relationship between the age at infection with HIV and the survival once infected remain a deep mystery.

While these simple statistical relationships are striking, unexpected and precise, their explanation will demand a combination of immunology, virology and reasonably sophisticated mathematical modelling.

The problem is that all too often the immunologist/virologist will say: our subject is far too complicated to be reduced to a few simple relationships expressed in mathematical form, while the mathematician will say: solving non-linear partial differential equations is far too complicated to explain to people who don't understand the language of mathematics.

So we go back to our laboratories and our computers and continue to work in isolation from our colleagues. While we all believe in 'inter-disciplinary research', the reality often falls short of the intent. How then can we begin to learn each others languages, hear what others are saying, use our joint knowledge and understanding to throw light on important problems, and hopefully make the world a slightly better place?

In this issue of the Quarterly we cover a wide range of issues, but underlying all of them is the need to communicate; among scientists, with policy makers, with the business community and with lay-people.

Alide Dasnois discusses the decades-long struggle of mine-workers to get compensation for occupational lung diseases. The Occupational Disease in Mines and Works Act (5) allows that 'Any person who works or who has worked at a mine or works, or any other person acting on behalf of such a person, may at any time apply to the Director for a medical examination of such a person for the purpose of determining whether such a person is suffering from a compensatable disease....The Director of the Medical Bureau for Occupational Disease is obliged to cause such a person to be medically examined as soon as possible.... ex-mineworkers are entitled to a bi-annual benefit examination and to transport costs on an annual basis.' In 1995 the assessors of a judicial commission of enquiry, set up by the government and Chaired by Justice Ramon Leon, to investigate the regulation of occupational health and safety in the mining industry, visited three hostels on two mines (Leon, Davies, Salomon and Davies, 1996). Each room was occupied by between 12 and 20 men, giving an average of just over five square metres per person. The assessors were 'shocked by the conditions in which food was prepared' and by ablution facilities 'so squalid as to shock the most hardened'. But 20 years into our democracy the law remains honoured in the breach, neither the government nor the unions seem interested in applying their minds to this historic injustice, the findings of the Leon Commission largely ignored and the men whose labour made this country rich continue to die. How then shall we convince the politicians of the incontrovertible link between hard-rock mining and occupational lung diseases and the right of millions of rural men to compensation and laws which have been in place for nearly half-a-century?

Part of the answer may lie in Rafael E. Luna's review of *We Have a Narrative* by Randy Olson. Echoing an article in the Quarterly by Brian Williams on the relationship between narratives and paradigms and their importance in scientific thinking it is not only in doing science that narratives matter but without good narratives it will be hard to convince those who are not scientists of the need to pay attention to our findings.

In *Lessons from the 2012 national HIV household survey to improve mathematical modelling for HIV policy*, Eaton, Johnson and Rehle show the importance of mathematical modelling in the interpretation of routine and surveillance data for HIV. The reason, of course, why dynamical models are so important in public health is this: statistical analyses will tell us where we are now, how many people are currently infected with HIV, and even how many people have recently been infected with HIV. But to estimate trends over time and, in particular to project future trends over time in incidence, prevalence, mortality, ART coverage, AIDS related opportunistic infections, and estimate costs and cost-benefits, we need to combine the data that we have with our detailed knowledge of the natural history of the diseases under investigation in order to constrain future projections. And it is the use of dynamical models, fitted to all the data that are available, but constrained by the natural history, that give dynamical modelling its great power. Now the hope is to use these results to inform and shape public health policy on how best to manage the epidemic of HIV and TB that continues to ravage our country.

The work described by Eaton *et al.* is complimented by an article by Manda on how best to address the problem of non-response when estimating HIV Prevalence using survey data. The problem is ubiquitous in all survey data and where the non-response rate is high can very easily lead to misleading results.

In *Predicting the effect of climate change on the abundance and distribution of tsetse flies* Hargrove and Vale show how very detailed measurements of all aspects of the ecology of tsetse flies, collected over a period of 25 years or more, combined with a detailed understanding of their natural history, in particular the way in which temperature affects different stages of their development, can be used to drive a mathematical model and predict their likely extinction in the Zambezi Valley by 2050, another unintended consequence of global climate change.

Finally, Musenge takes us back into the realm of formal mathematical analysis in his discussion of what colloquially known, at least by mathematicians, as the *Big N problem*. The problem arises when we have hierarchical data. For example, we might have data on countries, provinces within countries, town within provinces, people within towns, risk groups among people, and diseases among risk groups. In many instances it is trivial to write down a formal way to fit or to analyze the data but if it takes forever to run even on a supercomputer then the formal solution is not helpful. To illustrate the problem many scientific techniques, from calculating the structure of DNA, to analysing climate data, to the discovery of Quasars, depend on doing what is known as a ‘Fourier Transform’ on the data and the time to do such a calculation with N data points, in its simplest formulation, takes N^2 computations. However, first Gauss and later many others showed that one can use a modification of the technique that only takes $N\log_2 N$ computations. While this difference may not seem great, suppose that one needs to do a Fourier transform on a million data points, which in this day and age is not unusual. Then the saving in computations is $N/\log_2 N$ which, if N is 10^6 gives a saving of 50 thousand times. A computation that would take one day to run using a ‘slow’ Fourier Transform would take 2 seconds to run using a ‘fast’ Fourier Transform. It is easy to see that with hierarchical data the number of data points can easily become very large and finding efficient ways of carrying out analyses on such data sets may often mean the difference between being able to do the analysis and not being able to do the analysis.

Brian Williams *- Epidemiologist affiliated to SACEMA. Area of research interest: mathematical biology*.

References:

1. Oppenheimer JR. The tree of knowledge. Haper's. 1958:55-7.
2. Lonnroth K, Williams BG, Cegielski P, Dye C. A consistent log-linear relationship between tuberculosis incidence and body mass index. International Journal of Epidemiology. 2010; 39: 149-55.
3. CASCADE Collaboration. Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active anti-retroviral therapy. A collaborative analysis. Lancet. 2000; 355: 1131-7.
4. Williams BG, Granich R, Chauhan LS, Dharmshaktu NS, Dye C. The impact of HIV/AIDS on the control of tuberculosis in India. Proceedings of the National Academy of Sciences USA. 2005; 102: 9619-24.
5. ODMWA. Occupational Disease in Mines and Works Act, 1973 (Act No. 78 of 1973) and Regulations. Issue No.1 (Aquilia Publications CC)]; 1994.