The role of data collection and management in achieving better public health outcomes

Magna Aidoo

In the practice of modern administration, data is defined as a volume of systematically recorded facts, collected and organised according to a regular scheme (Gordon, 2007). Such information has for centuries been recorded in paper form (for example, in tables, files and index cards), but from the middle of the twentieth century has increasingly been managed in computerised systems such as databases.

The practice of health care involves a great deal of recording, management and use of data. Some of this is precise, structured and even quantitative in form (for example, pathology lab results), and some is less formalised, stored often as text. Patients have personal medical records that are created and maintained by their physicians. If someone has an episode in hospital, a clinical record will exist for them even if it is only to support administration and billing.

There is a strong historical relationship between data recording and epidemiology, starting with Graunt's seventeenth century analysis of the 'Bills of Mortality' (registrations of causes of death), which aimed at understanding bubonic plague. Statistical correlation studies can even point the way towards identifying the causes of diseases, as in Snow's geographical study of the London cholera epidemic of 1854, which helped build the case for 'germ theory', identified the sewage and water systems as routes of transmission and led to a revolution in public health policy.

Having accurate and comprehensive data about the health of the population and the major causes of ill health to this day assists governments and other agencies in formulating their health care and health promotion policies and allocating resources on the basis of perceived priorities. In pursuit of these goals, strictly medical data may be coupled with data about the population's demographic and economic profiles – for example, to identify at risk populations.

Methods of health data collection

There are four main ways of capturing health-related data:

- **Civil registration systems:** Ideally covering the entire population, this method includes registers of deaths and is particularly valuable as a source of epidemiological evidence where death registers include a reliable account of the cause of death. However, such schemes are rare in low- and middle-income countries, they under-report many chronic and debilitating but non-fatal conditions and they provide little depth of evidence.

- **Data from personal medical records systems.** This method is based on the assumption that each patient has a longitudinal collection of personal health information that has been entered into the record by qualified health providers, including such elements as physician observations, lab results, blood pressure or blood glucose readings, prescribing information and so on. Such information is gathered and organised primarily to support continuing, efficient and quality health care, but when it is aggregated and anonymised it can provide an excellent evidence base for detecting trends, adducing causes and informing health practice policy. However, this approach is only realistically possible with a well-integrated system of electronic health records, which only a few higher-income countries are coming close to implementing.

- **Data collected during the routine care of chronic conditions.** This method is a variant of the previous one, but less 'global' in scope. It is based on the observation that in many countries there are special agencies or health campaigns, often better organised and resourced than the rest of the local health service, which are set up to cope with particular conditions such as malaria or HIV/AIDS, or which focus on particular care pathways such as maternity, childbirth and child development. Such sectoral projects frequently do a good job of recording personal health data pertinent to the care that they deliver, and they may provide a locus for capturing further data and screening for a variety of conditions.

- **Data obtained by special surveys.** In situations where the health informatics infrastructure is weak and health records are not easy to find or access, information about the health of populations may be collected by special surveys. According to their design, surveys may at the same time gather extra data such as demographics, eating habits and other lifestyle factors. Survey data are easy to anonymise for data analysis and surveys can be mobilised relatively easily and cheaply, though it should always be borne in mind that the survey population is only a sample and poorly designed surveys may skew the data they collect.

The 'sentinel method' of health surveillance generally means when a particularly well-organised institution such as a local hospital undertakes the analysis of its patient record data, and this is regarded as a sample indicative of the health of the wider population. However, while recommended by some, this is not a data collection method in its own right.

Surveillance, from CDs to NCDs

Historically, epidemiology concerned itself with the surveillance of communicable diseases, particularly those that are easily transmissible and resulted in virulent outbreaks bringing death on a large scale, such as bubonic plague in the seventeenth century, cholera in the nineteenth century and influenza in the early twentieth century. In recent years swine flu and SARS have provoked similar alarm, as indeed did HIV/AIDS until its true mode of transmission was properly understood. Since 1969 the World Health Organization (WHO) has maintained a list of 'notifiable' infectious diseases, which countries must monitor and have a treaty obligation to report. These are cholera, plague, yellow fever, smallpox, relapsing fever and typhus and, since 2005, polio and SARS. WHO also runs the Epidemic and Pandemic Alert and Response (EPR) system for international surveillance of a range of emergent infectious threats, including several haemorrhagic fevers, influenzas and forms of hepatitis.

Influenza remains subject to a high degree of surveillance and contingency planning due to its rapid rate of mutation, and because it has caused a number of major pandemics with many fatalities. Tuberculosis (TB) is also a subject of concern as it is the world's number one infectious killer of women of reproductive age, and a major proximate cause of death for people with AIDS.

It is now time for more resources and effort to be turned towards data collection and surveillance of non-communicable diseases (NCDs). Coronary and vascular disease, cancers, diabetes and chronic obstructive pulmonary disorders have been identified as the main emerging threats. The number of deaths from such causes is already double that of deaths from all infectious diseases (including HIV/AIDS, TB and malaria), maternal and perinatal conditions and nutritional deficiencies combined (Daar et al., 2007). The WHO predicted in 2005 that without concerted action, 388 million people worldwide would die of chronic NCDs by 2015. Sixty per cent of all deaths globally are due to NCDs, and of these 80 per cent occur in low- to middle-income countries, including a significant number of Commonwealth countries.

NCDs can have pervasive and costly impacts on societies. Failure to control diabetes, for example, leads to poor quality end-of-life, blindness and amputations, and these burdens have overstretched some health-care systems.

In a global status report for 2010, WHO argued that obtaining accurate data is vital to reversing the global rise in premature death and disability from NCDs (WHO, 2011). However, in many countries there is currently little relevant mortality data and poor surveillance over these conditions. What data does exist on NCDs is said to be poorly integrated into national health information systems (ibid).

The WHO 2008–2013 action plan for the prevention and control of NCDs contains recommendations under three headings: the monitoring of risk factors (including lifestyle factors) that contribute to prevalence of NCDs; monitoring outcomes in terms of morbidity and mortality; and assessing the capacity of the national health system to respond to the challenges. It is recommended that this last should include assessment of the national capacity to prevent NCDs, to achieve early detection and to provide appropriate health care, including medicines (WHO, 2008).

Data collection that inquires into the lifestyles and consumption patterns of patients is important in understanding NCDs. Statistical correlation of such data can lead to the identification of 'risk factors'[1] and indicate where research can usefully be concentrated to discover the mechanisms of illness, which in turn helps to devise effective means of intervention. Data collection and statistical analysis also provide evidence for judging the effectiveness of treatments – for example, the use of statins and vitamin supplements to prevent cardiovascular disease.[2]

Data collection around NCDs goes hand in hand with the issues of screening and the understanding of risk factors. Effective monitoring of the health of individuals afflicted with NCDs (or those at risk of them) is possible only if they have been identified and recruited into a care programme or monitoring scheme. In the case of cancer, the high cost of treatment is an argument in favour of deploying screening programmes for early detection and intervention. For cardiovascular conditions and diabetes, monitoring body mass is indicative and blood testing can determine what interventions may be appropriate. In devising screening programmes, it can be helpful to identify which sectors of the population are most at risk and find institutional settings in which people can be easily recruited to a screening programme (such as in schools and universities, the civil service and certain large employers).

Challenges – and possible solutions

Looking back to the list of data collection methods above, we may think that some appear inspirational but are currently impractical in many settings. We might dream of an ideal system of health and social data collection and management, covering the entire population from cradle to grave, encompassing every health issue, and so standardised and interconnected that medical research and public policy alike is perfectly informed. The reality is very different. To a greater or lesser extent, many aspects of the health of individuals are not recorded. Not infrequently, their records are fragmented across different locations and periods of their lives, or are held in 'silo' database systems that cannot 'talk' to each other. Many are stored in paper form, which cannot easily be anonymised, aggregated and analysed in support of research and planning.

In many Commonwealth countries the emphasis for now must be on practical solutions that can be developed quickly and without great expense, reaching as many people as possible, neglecting no significant sectors of the population and delivering clear benefits in short order. One of the most promising ideas appears to be to engage with established projects and programmes that have prolonged contact with sectors of the population. Around maternal health, for example, there are great opportunities arising from antenatal contact between a clinic or health visitor and the expectant mother. There are opportunities for screening for various conditions, early diagnosis and general monitoring of the health of both infant and mother.

No country should underestimate the host of problems that can stand in the way of effective health data collection and health surveillance. It is not enough merely to place the task of data collection on the health front-line staff. They also need to be motivated to do so, to see how it benefits their own work and

the health of the population and to pay attention to data quality and accuracy.

Data is of little help in population-wide health surveillance if it cannot be efficiently extracted, abstracted, aggregated and analysed. This argues in favour of a rolling-out of low-cost but effective systems for electronic records capture and management. We should be ready to take advantage of modern resource-efficient technologies such as mobile phones and cellular networks, already being ingeniously applied to health information management in various parts of the world. For example, at the St Gabriel's Hospital in Malawi a system based on open-source software keeps track of distantly located patients, informs community health workers of proper drug dosages and co-ordinates support for HIV-affected patients through a computer-to-SMS-message gateway.

The other technical challenge in computer-assisted health data collection, as decades of experience in the industrialised world have shown, is to ensure that data is not trapped within closed systems, but is – as the jargon has it – 'inter-operable' across different systems, capable of being combined and compared while protecting the privacy of patients.

Finally, we need to find better ways of making sense of data as it emerges, ensuring that it helps improve the quality of public planning. For example, data visualisation 'mash-ups' that merge data from more than one source, as in health data superimposed onto geographical maps, make it easier to spot trends, correlations and emerging patterns and threats. Educational campaigns to improve statistical literacy, not only among medical professionals but also among politicians, planners, other policy makers and the general public, may also make data use more effective.

References

Daar, A.S. et al. (2007). Grand challenges in chronic non-communicable diseases. *Nature*, 450: 494–96.

Gordon, K. (2007). *Principles of data management: facilitating information sharing*. Swindon: British Computer Society.

World Health Organization (WHO) (2011). Global status report on non-communicable diseases 2010. Geneva: WHO.

World Health Organization (WHO) (2008). 2008–2013 action plan for the global strategy for the prevention and control of noncommunicable diseases. Geneva: WHO.

Endnotes

[1] A good example is the British doctors study, initiated in 1951 by Richard Doll and Austin Bradford Hill, which pointed to a link between tobacco smoking, lung cancer, and coronary thrombosis.

[2] See the Medical Research Council and British Heart Foundation's Heart protection study, results of which were published in stages between 2002 and 2005: www.hpsinfo.org

Dr Magna Aidoo is currently Head of Health at the Commonwealth Secretariat. Her background is in public health, with a focus on mental health and HIV/AIDS. Until recently she worked in the UK National Health Service, where she was known for her advocacy and contribution to strategies on diabetes. In addition to her international work, Dr Aidoo has for 20 years been Associate Lecturer at Goldsmiths, University of London, teaching Organisational Management, Research Methods and Community Education. Dr Aidoo is also is on UK Public Health Register.