European Journal of Public Health, Vol. 23, No. 3, 350–351

Viewpoints

Public health surveillance with electronic medical records: at risk of surveillance bias and overdiagnosis

Public health surveillance consists in the systematic collection, analysis and interpretation of data, closely integrated with the timely dissemination and use of these data to help prevent and control diseases and injuries.[1] Surveillance is traditionally conducted with data tailored to address a specific health problem, for instance, data on risk behaviours from periodic health surveys or data on cancers collected in dedicated registries. These data and methods of collection are 'designed' for specific surveillance activities.[2]

In the past two decades, methods of surveillance have radically changed with the development of information technology, making easier the collection of data for public health surveillance.[3] In particular, data from health care providers have become highly accessible through electronic medical records (EMRs), which offers fantastic opportunities for the surveillance of several conditions.

Surveillance of chronic diseases using EMRs

EMR is a digital record of patient's health and health care from hospitals, nursing homes, physician's offices or pharmacies.[3] All kinds of medical events are recorded in EMRs, being diagnoses, laboratory results, treatments, drug prescriptions, vital signs or clinical outcomes. EMRs increase the availability of clinical information, can help decision making and improve quality and efficiency of care. Furthermore, shared EMRs are cardinal for the evolution of health professionals' practices in the care of chronic diseases.[4]

EMRs are designed to support health care providers' activities at the point of care but are not primarily designed for surveillance. Nevertheless, the large amounts of raw data—also qualified as 'organic'[2]—that can be retrieved from EMRs are highly attractive for the surveillance of several conditions, and of chronic diseases in particular. For instance, EMR-based surveillance system could allow identify new diabetes cases and also help assess the process, quality and costs of care of these patients.[5] No other tool has the potential to provide such comprehensive and timely information on chronic diseases.

There are, however, limitations to what can be expected from EMR-based surveillance systems due to data misreporting and to the lack of standardization on how health events are defined and recorded in EMRs. In particular, two related issues challenge the interpretation of EMRs data, that is, 'surveillance bias' and 'overdiagnosis'.

Surveillance bias

Surveillance bias occurs when a condition is searched with differential intensity according to the setting of care or the type of patients.[6]

For instance, detection and treatment of deep vein thrombosis (DVT), a preventable complication of major surgery, is recommended. Suppose that, following a quality improvement policy, systematic screening of DVT by duplex ultrasound is conducted among all patients having major surgery in the hospital A. Thanks to systematic screening, asymptomatic (hence less severe) cases of DVT are found in addition to symptomatic cases. Therefore, the rate of DVT—easily computed thanks to EMRs—will be high in this hospital.[6] In the hospital B, where no systematic screening of DVT is conducted, only symptomatic cases will be identified, and the rate of DVT will be lower. The rate difference between both hospitals is solely the result of a surveillance bias: it does not reveal any difference in the epidemiology of DVT, in quality of care or in prevention policies between hospitals.

Surveillance bias is common with adverse effects of drug treatments, complications after a medical procedure or hospital-acquired conditions, e.g. nosocomial infections, for which differential surveillance activity occurs according to the care setting. Data on these events are of major interest for the assessment of quality of care and can be, actually, easily retrieved from EMRs.

An epidemic of diagnoses due to overdiagnosis

Overdiagnosis also challenges the interpretation of EMRs data. It occurs when asymptomatic people are diagnosed with a condition not associated with a substantial risk of adverse outcomes or for which no intervention reduces substantially this risk.[7,8] Many conditions can be overdiagnosed through screening, incidental findings or by widening the boundaries of treatable conditions. Thus, lowering the thresholds of blood glycaemia level to diagnose diabetes or pre-diabetes can lead to overdiagnosis.[9]

One consequence of lowering thresholds is that patients are now diagnosed with conditions that are less harmful compared with the same conditions diagnosed previously. Hence, a diagnosis of hypertension set for a blood pressure >140/90 mmHg is associated with a lower risk of disease than a diagnosis set for a blood pressure >160/100 mmHg. Furthermore, the combination of screening and of lowering thresholds to define conditions sustains a vicious circle, leading eventually to an epidemic of diagnoses (figure 1).[7] This epidemic of diagnoses may not match any epidemic of diseases.

For instance, an increasing share of the population is treated for hypercholesterolaemia. This increase can be easily documented using EMRs from hospitals, physicians' files or pharmacy databases. Nevertheless, this trend does not imply an increase in the frequency of people with risky blood lipid profile. On the contrary, favourable trends in blood lipids have been observed in high-income countries, and such trends were not explained by improved treatments. The increase in the prevalence of hypercholesterolaemia is due to the higher proportion of persons screened and treated, and to the lowering of blood lipid thresholds for initiating treatment.

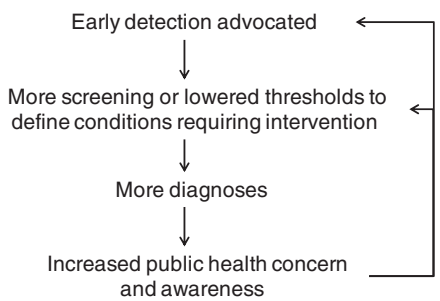

Figure 1 Screening and lowering thresholds to define conditions requiring intervention lead to an increasing number of persons diagnosed with the condition and, eventually, an epidemic of diagnoses (adapted from Welch[7])

The problem is that diagnoses reported in EMRs most often do not capture the changes in the way hypercholesterolaemia is defined or treated, which blurs the analysis of genuine trends in risky blood lipid profile.

Conclusion: high accessibility vs. high quality?

With the development of information technology and EMRs, data from health care providers have become highly accessible, which offers fantastic opportunities for the surveillance of chronic diseases. Nevertheless, the high accessibility does not imply a high quality of data, on the contrary. EMR data should be used with caution: they, indeed, do not speak for themselves. As underlined, surveillance bias and overdiagnosis recall that data from health care providers convey complex health events not defined in a standardized way. At the dawn of the digital era, providing useful information from health care providers for public health surveillance requires more than ever a critical eye on EMR-based surveillance system.

References

1 Lee LM, Thacker SB. Public health surveillance and knowing about health in the context of growing sources of health data. *Am J Prev Med* 2011;41:636–40.

2 Keller SA, Koonin SE, Shipp S. Big data and city living—what can it do for us? *Significance* 2012;8:4–7.

3 Bitton A, Flier LA, Jha AK. Health information technology in the era of care delivery reform: to what end? *JAMA* 2012;307:2593–4.

4 Santschi V, Chiolero A, Burnand B, et al. Impact of pharmacist care in the management of cardiovascular disease risk factors: a systematic review and meta-analysis of randomized trials. *Arch Intern Med* 2011;171:1441–53.

5 Klompas M, McVetta J, Lazarus R, et al. Integrating clinical practice and public health surveillance using electronic medical record systems. *Am J Public Health* 2012;102(Suppl 3):S325–32.

6 Haut ER, Pronovost PJ. Surveillance bias in outcomes reporting. *JAMA* 2011;305: 2462–3.

7 Welch HG. *Overdiagnosed: Making People Sick in the Pursuit of Health*. Boston, USA: Beacon Press, 2011.

8 Moynihan R, Doust J, Henry D. Preventing overdiagnosis: how to stop harming the healthy. *BMJ* 2012;344:e3502.

9 Chiolero A, Paccaud F. Prediabetes and the risk of diabetes. *Lancet* 2012;380:1225.

Arnaud Chiolero[1,2], Valérie Santschi[1], Fred Paccaud[1]
[1]Institute of Social and Preventive Medicine (IUMSP), University Hospital Center, Lausanne, Switzerland and [2]Observatoire valaisan de la santé (OVS), Sion, Switzerland

Correspondence: Arnaud Chiolero, Institute of social and preventive medicine (IUMSP), University Hospital Center (CHUV/UNIL), Biopôle 2, Route de la Corniche 10, 1010 Lausanne, Switzerland.
e-mail: arnaud.chiolero@chuv.ch

doi:10.1093/eurpub/ckt044
Advance Access published on 18 April 2013

European Journal of Public Health, Vol. 23, No. 3, 351–352

All data and all diagnoses deserve a critical eye

I agree completely with the conclusion of this article that 'providing useful information from health care providers for public health surveillance requires more than ever a critical eye on EMR-based surveillance systems'. But at the same time I advocate 'a critical eye' for everything and certainly for all data. We can easily be bewitched by numbers into thinking that they have a precision and validity that we never expect from words, but numbers need to be treated as cautiously as words. Where do they come from? Was there collection quality assured? Were data missing? What biases are at work? Are they timely? If they have been 'adjusted', how well have they been adjusted?

The Need for a Critical Eye

I am writing this in a country (the UK) and on a day (17 March 2013) when data collected from routine systems in hospitals are on the front page of our newspapers.[1] Professor Brian Jarman says that 20 000 people may have died prematurely in a few English hospitals and that the government ignored the signals from the data.[1] Jarman has for many years been producing hospital standardized mortality ratios (HSMRs), 'the ratio of the observed to expected deaths, multiplied by 100, with expected deaths derived from statistical models that adjust for available case mix factors such as age and comorbidity'.[2]

HSMR is on our front pages because a high ratio signalled that something was badly wrong in Mid Staffordshire National Health Service (NHS) Trust, a hospital in the Midlands of England. The high ratio together with other signals of poor performance led to the discovery that care in the hospital had fallen way below acceptable standards and that there had been perhaps 2000 excess deaths. For many people, this 'proved' the usefulness of the HSMR and has led to a call for all other hospitals with high HSMRs to be investigated.

But, as those who developed the HSMR argue themselves, the ratios need to be interpreted with a critical eye.[2] There are problems with the numerator, denominator, risk modelling, interpretation and coding, and some epidemiologists have argued that they are so misleading as to be useless.[3,4]

All of this matters greatly because several hospitals are now being investigated because of persistently high HSMRs and because one of the main ways that the NHS, currently undergoing yet another reorganization, plans to regain its reputation is through making transparent data about everything that is happening in the NHS. We can at least know most of the time whether patients are dead or alive (although not always accurately), but when it comes to subtler diagnoses, there is much room for misinterpretation, as Chiolero and others show. The expectation in the NHS is that many of the data will come from electronic medical records in real time. Many critical eyes will be needed.

Overdiagnosis and Moving from Treatment to Prevention

The great epidemiologist Geoffrey Rose, who had a gift for sound bites, said that there is no disease you either have or do not have, except perhaps rabies and sudden death. With everything else, you might have a little of it or a lot of it. So disease thresholds are arbitrary and can be moved backwards and forwards. Defining normal is also famously difficult. Using a definition of normal as