“Get back to where you once belonged”: monitoring the AIDS pandemic in the 21st century

“Volte para o lugar de onde saiu”: monitorando a pandemia da aids no século XXI

Abstract This paper reviews the concepts and methodological strategies that have shaped the monitoring of the AIDS pandemic, today in its third decade. A deliberate effort was made to highlight aspects usually forgotten by the canon. The paper aims to track the footsteps of the evolving strategies in the field of surveillance & monitoring, with the help of disciplines such as epidemiology, molecular biology, social, and behavioral sciences. The deep divide that opposes societies severely affected by the epidemic and affluent societies much less affected by the epidemic is contrasted with the scarce human and financial resources of the societies facing harshest epidemic vis-à-vis the comprehensiveness of the response to the epidemic in their affluent counterparts in terms of the scope and high standards of their initiatives on monitoring, prevention, management & care. The pressing need to implement feasible alternatives to the current sophisticated and expensive ones is briefly discussed. Beyond the renewed challenge posed to the creativity of scientists and health professionals, the AIDS pandemic is described as a major public health crisis, compromising the social fabric in some contexts, and as a never fulfilled calling for an ethics of solidarity between different societies and different social strata of each given society.

Key words Aids, Monitoring and evaluation, History of science

Resumo O artigo procede a uma retrospectiva histórica dos conceitos e estratégias metodológicas que informaram e informam o monitoramento da pandemia de aids, hoje na sua terceira década. Procurou-se iluminar aspectos raramente abordados pela história canônica da epidemia, buscando traçar as raízes e desdobramentos das atuais alternativas de vigilância e monitoramento, no campo da epidemiologia, da biologia molecular e das ciências sociais e comportamentais. O profundo cisma que opõe sociedades duramente afetadas pela epidemia e que contam com recursos humanos e financeiros escassos a serem empregados tanto no monitoramento, como na implementação ampla e adequada de estratégias de prevenção e tratamento às pessoas vivendo com HIV/aids nas sociedades afluentes e substancialmente menos afetadas pela epidemia é criticamente revisado, em prol de soluções factíveis e prementes, dado o alto custo e a complexidade de boa parte das alternativas-padrão em curso nos países mais ricos. Mais do que um desafio à engenhosidade de cientistas e profissionais de saúde, a pandemia de aids é revis-tada como uma crise de grandes proporções da saúde pública, da vida social e um chamamento renovado à frustrada ética solidária entre as diferentes sociedades e segmentos sociais.

Palavras-chave Aids, Monitoramento e avaliação, História da ciência
Origins

In 1981, young men from Los Angeles, San Francisco and New York were hospitalized with severe medical conditions that did not respond to standard antibiotic regimens. We now know these cases as the tip of a vast iceberg, the visible beginning of the worst pandemic mankind has ever faced. The history of the Acquired Immunodeficiency Syndrome (AIDS) epidemic has been told many times by contemporary experts from a range of disciplines, and will not be retold here. Instead, the present paper aims to provide some historical perspective on the development and current status of AIDS surveillance.

Two aspects of the history of the AIDS epidemic will be particularly useful in guiding us through the scientific labyrinth that AIDS has opened for the contemporary epidemiologist. The first tip of our Ariadne’s thread refers to the origins of AIDS. While the first AIDS cases were recognized in the mid-1980s, subsequent studies have shown that the etiologic agent of AIDS, a retrovirus named Human Immunodeficiency Virus (HIV), has been around much longer. As demonstrated by the American molecular biologist Bette Korber and coworkers, HIV most likely emerged in the first half of the 20th century and has circulated in some African communities since then. The virus disseminated at a relatively slow pace in the 1940s and 1950s, but for many different reasons, still under debate, gained momentum in the 1970s, when a much larger dissemination took place. In the absence of a good surveillance system in most African countries, HIV spread unnoticed. Thus when the Western scientists finally “discovered” the emergent syndrome, originally linked to Homosexual men, Heroin injectors, Hemophiliacs and, later on, Haitian immigrants living in the USA (the infamous “4 Hs”); see details in Bastos, the actual African H (Heterosexual transmission) was already the driving force of a devastating epidemic. In the absence of reliable surveillance, the epidemic progressed unabated in Africa, to “emerge”, many years later.

The Sub-Saharan epidemic progressed to become a major threat to public health and a deep crisis. In addition to devastating the health of the African population, a major catastrophe by itself, HIV has exacerbated many other medical and social horrors, such as endemic tuberculosis, persistent malaria, civil wars, droughts and famines. The resulting multifarious syndemic (see details about the concept of syndemics and their impact on public health at http://www.cdc.gov/syndemics/overview.htm), with AIDS as one of its main components, has compromised the frail local economies of the Sub-Saharan countries, disrupted the social fabric of these communities, and contributed to an unexpected reversal of much of the hard-won social and health progress of the 1970s. Over the course of the epidemic, those countries most severely affected have seen their gains in terms of decreasing child mortality, augmented life expectation and tuberculosis control (both in terms of a decreasing incidence of TB and the proper management of TB with first-line drug regimens), partially or totally reversed to the unacceptable levels of the immediate post-colonial period.

The second tip of our Ariadne’s thread involves the way in which AIDS was initially detected in the abovementioned American hospitals. The mysterious syndrome, later called AIDS, was identified as a new medical condition as a result of the comprehensive exchange of information and experience between the staff of the hospitals where the young men were hospitalized, in particular experts in the fields of pharmacovigilance and nosocomial infection surveillance, and the researchers from the Centers for Disease Control and Prevention (CDC) (later renamed as Centers for Disease Control and Prevention). Recognition of both the severity of the new syndrome and its potential as a new pandemic in the making arose directly out of this exchange of information.

Pharmacovigilance and nosocomial infection surveillance gradually emerged as key areas of biomedical expertise after the triumphant post-World War II years, when most practitioners and many public health officials supposed, in what we now know was a terrible mistake, that infectious diseases would be soon “conquered” by the vigorous use of an increasing arsenal of antibiotics. Bravado was slowly replaced by prudence, as the indiscriminate use of antibiotics was found to be associated with the emergence of resistant strains among a wide range of micro-organisms. It was in this context that the first outbreaks of nosocomial infections were thoroughly documented and analyzed and the importance of nosocomial surveillance was established.

Over the same period, the booming field of new and powerful medicines (antibiotics included) brought with it more bad news. The lack of comprehensive post-marketing surveillance of medicines such as Thalidomide created a worldwide nightmare of children born with severe body deformities from Thalidomide-using mothers.
Recent recognition of the risks associated with non-steroid anti-inflammatory drugs serve as a reminder to the new generation of doctors and patients that post-marketing pharmacovigilance is a permanent necessity.

In addition to playing a key role in the identification of the new syndrome, the fields of nosocomial infection surveillance and pharmacovigilance also played key roles in identifying the etiology of the new disease. When AIDS first emerged in three major metropolitan areas in the USA, it was initially hypothesized that the syndrome was due to unexpected side effects of the antibiotics employed in the unsuccessful treatment of the young men (who died in the interval of a few weeks). Although this hypothesis was later ruled out, the role of chemical cofactors such as poppers (inhaled nitrites) persisted for years in the medical literature, and from times to times still haunts the laboratories of virologists and data-banks of epidemiologists.

In the same way, the initial hypothesis that the severely ill young men were infected with resistant strains of different micro-organisms was thoroughly explored and then ruled out. The micro-organisms affecting these patients, such as *P. jirovecii* (then called *P. carynii*), were found to be weakly pathogenic or non-pathogenic for healthy individuals; this phenomenon was later defined under the umbrella of opportunistic diseases. Thus, the hypothesis that AIDS was the result of an underlying immunodeficiency, as later proved true, was corroborated not only through the results of laboratory tests and clinical assessments, but also due to the provisional elimination of concurrent hypotheses.

Neither adverse drug effects nor antimicrobial resistance proved to be the etiologic agent in AIDS, and immunodeficiency took the center of the stage. However, in the years that followed, both viral and microbial resistance and the complex side effects of drugs aiming to treat HIV infection, co-infections and opportunistic diseases would reappear as major players in the evolving epidemic, returning to the stage from the back door, under a different context. The complex interplay of immunodeficiency (and the chance to reconstitute the integrity of immune response), the emergence and spread of resistant strains of HIV and of a myriad of microbes causing opportunistic infections and co-infections, and the side effects of a growing arsenal of medicines are today intertwined under the spotlight of contemporary research and will likely remain there for the foreseeable future.

**Brookmeyer’s first breakthrough:**

**Back-calculation**

In the mid-1980s, the first large AIDS cohorts such as MACS (Multicenter AIDS Cohort Study) demonstrated that AIDS could not be studied adequately using the traditional tools and concepts of infectious disease epidemiology, which had been shaped for decades by the study of acute communicable diseases. In contrast, the natural history of HIV infection spanned years, with acute infection followed by a long latent period (during which the virus was actively replicating but the host remained asymptomatic) before the advent of the full-blown clinical syndrome of AIDS.

As a result, monitoring the incidence of AIDS cases could only provide a picture of the past of the epidemic, out of phase with its current state by as much as 8-10 years (the latent or incubation period of the syndrome). In the absence of surveillance for the virus itself (instituted much later in some countries, and still under criticism in some settings due to ethical concerns and operational difficulties), the American statistician Ron Brookmeyer proposed a simple but insightful alternative: the equation defining the incidence of AIDS could be deconvoluted, taking in consideration the underlying distribution of the incubation period. With this new “egg of Columbus”, useful surveillance data were generated for the purposes of modeling, reconstruction of epidemic trajectories, and estimation of the impact of alternative interventions, such as the consistent use of condoms at the population level or the putative impact of an anti-AIDS vaccine.

The first back-calculation models relied on a basic assumption about the stationarity of incubation periods, or in other words, the stability of incubation periods over successive cohorts of newly infected patients. As long as such an assumption provided a reasonable approximation of the real world, different cohorts of individuals, infected in different periods of time, could be modeled in an integrated way, using a fixed underlying distribution of incubation periods. Problems with this approach arise, however, if the underlying distribution of the incubation period changes over time, singularizing each successive cohort of patients.

The assumption of a stationary incubation period was subsequently relaxed through a range of strategies, employing classic or Bayesian statistics. However, dramatic changes in the ecology and treatment of HIV infection after the mid-1990s posed a conceptual challenge to even the
best conceived modeling strategies: Does HIV natural history still exist or is it just a new item for the museum of contemporary medicine?

HIV “natural history”: all that is solid melts into air

Back-calculation is based on a key concept: despite individual variations, the underlying distribution of the incubation period can be used to get the deconvolution of the equation defining AIDS incidence. This underlying distribution has a real meaning if HIV disease can be described under the paradigm of the natural history of HIV infection. However, the course of HIV disease has been impacted profoundly by prophylaxis for opportunistic infections (such as sulfamethoxazole for the prevention of P. jiroveci infection), early treatment of tuberculosis, involuntary or deliberate modification of the environment where HIV+ people live, the changing ecology of different micro-organisms that may harm patients with immunodeficiency, and last but not least, the introduction of potent anti-retroviral medicines. These interventions have changed the so-called natural history of HIV to such an extent that the very meaning and consistency of the concept has been called into question. Indeed, there is no longer something we could call a natural history of the HIV infection, but rather a complex interplay between the social ecology of human communities and their networks, the ecology of virus populations and the arsenal of human-made interventions in terms of prophylaxis and treatment.

The last attempts to use back-calculation in the post-HAART area have resulted in substantial inaccuracy, despite the incorporation of a range of statistical techniques to deal with different sources of variability. One of the last attempts to model the dynamic of the AIDS epidemic in a context with a reliable information system (Europe) yielded imprecise estimates. To account for the impacts of the wide range of factors affecting disease progression, impacts which are strongly dependent on local contexts as a result of variability in the local ecology of micro-organisms, in the availability of different medicines, and in guidelines regarding medication use, back-calculation was forced to incorporate so many new and varying parameters that any modeling attempt was caught between excessive model complexity and oversimplification. It was time to move beyond back-calculation.

Brookmeyer strikes back: some basic epidemiology can help a lot

Brookmeyer, however (this time in partnership with his colleague Tom Quinn), would have a second “Columbus egg” insight. Using the classic epidemiologic concept that prevalence (P) is the equivalent of incidence (I) times the duration (D) of disease (P = I x D), he and Quinn found that the timing of the emergence and detection of different biomarkers of HIV infection provided a way to find an interval roughly equivalent to the duration of HIV infection (and thus to estimate incidence).

The classic epidemiological equation requires some basic assumptions, such as the stability of the population under analysis (in addition, the actual relation involved the prevalence odds instead of prevalence itself). However, a hypothetical stable population and simple prevalence constitute two reasonable approximations when you are eager to track a deadly epidemic. In the same way, it is not necessary to call for Einstein’s relativity every time one needs to solve a problem in classic physics, for which Newton’s much simpler equations constitute a quite precise approximation.

The method of Brookmeyer and Quinn relied on the measurement of a peptide called P 24 and the measurement of antibodies against HIV, both of which rise progressively following infection. The authors used the interval between the detection of these two biomarkers as a proxy for the duration of a segment of the history of HIV infection. But the devil in statistics always lives in the impossibility to tame randomness. The rate at which these two biomarkers emerged in the course of HIV infection was quite variable, resulting in variable estimates of duration. As a result, Brookmeyer’s new tool was successfully applied primarily to very large populations, such as a large Indian population that pioneered the application of the new technique.

The main limitation of the technique was thus related to the very large confidence intervals of any incidence estimate when the populations under analysis were significantly smaller than the original studies of thousands (themselves originally formulated as a valid alternative to the other source of inaccuracy in the measurement of incidence - the high dropout rates of most longitudinal studies). Epidemiologists were thus lost between the Scylla of imprecise incidence estimates from cross-sectional studies (in which patient drop out over time obviously does not exist) using Brookmeyer and Quinn’s breakthrough tech-
nique, and the Charybdis of longitudinal studies were incidence could actually be measured (instead of estimated), but in which high attrition rates resulted in underestimation of HIV incidence up to 60%, as discussed in the paper by Brookmeyer et al. Such high attrition rates were especially relevant among studies carried out in low and middle income countries, were the epidemic was particularly vigorous and dynamic, and surveillance a major challenge.

However a succession of researchers, profiting from Brookmeyer’s original insight, had gradually emerged over time, and have been working to expand and refine the original paradigm. The first of Brookmeyer’s heirs was Robert S. Janssen, from CDC.

The children of Brookmeyer: detuned assays and BED

The work of Janssen and coworkers could be defined as a third Columbus egg, but not in the sense of Brookmeyer’s previous insight. Janssen et al.’s original paper (published in 1998) was based on the same premise as Brookmeyer and Quinn’s paper, namely that incidence can be estimated in the context of a cross-sectional study if biological markers can function as a proxy for the duration of a segment of HIV infection over time. But unlike Brookmeyer and Quinn’s original formulation, Janssen et al. did not contrast two different markers, but rather found a way to see the same phenomenon with two different “glasses”. Janssen and coworkers used kits with altered optical densities to “watch” the progressive increase of antibodies against HIV over time, with “tuned” and “detuned” glasses. The so called “detuned assays” (the ones with altered optical densities and therefore altered diagnostic properties) could only detect larger amounts (densities per ml) of antibodies vis-à-vis the regular (“tuned”) kits, then creating a fake interval of time (from the perspective of the actual biological phenomenon and its respective timing) superimposed upon a single photograph of an ongoing process (the progressive increase of antibodies).

The new technique experienced a boom in the late 1990s and early years of the 21st century, using a range of kits and applied to many different contexts in terms of human and virus populations. Since risky behaviors constitute the inescapable source of new infections, behavioral surveillance emerged as an interesting alternative to laboratory-intensive surveillance, especially for those contexts in which epidemics were relatively concentrated among a handful populations at increased risk. In generalized epidemics where heterosexual transmission affects a substantial proportion of the population, behavioral surveillance requires population-based large surveys, posing again the questions of high costs and complex logistics.
Behavioral surveillance faces major difficulties. The first of these was addressed by the definition of a comprehensive basket of indicators and the development of standardized questionnaires by major international institutions and funding agencies. Complexity arises, however, in the pressing need to make comparisons across different societies and cultures, and the need to tailor assessments to local characteristics and cultural mores.

One major paradigm shift must be mentioned here. With the decisive input from Jonathan Mann and coworkers, the conceptual framework underlying the definition and use of behavioral and social indicators move beyond the straight-jacket of the classical concept of risk (itself problematic in the field of infectious disease epidemiology and rather tailored to the assessment of non-communicable diseases) toward the multidimensional paradigm and toolbox provided by the concept of vulnerability.

Many times implicitly rather than explicitly used by those who employ the toolbox eventuating from Mann’s seminal concepts, the conceptual framework originally proposed by Mann and his coworkers provides the very foundations of contemporary baskets of social and behaviors indicators, such as the one proposed by the UNGASS (The United General Assembly Declaration of Commitment on HIV/AIDS), used worldwide as a core resource toward a cross-cultural monitoring of the pandemic.

From the point of view of epidemiologic methods, emphasis has been put on two other major difficulties: how to infer trends from serial cross-sectional studies and how to properly sample hidden populations. Since the seminal paper by Ades on serial seroprevalence surveys, one must be cautious while interpreting and inferring long-term trends from panels of successive cross-sectional studies. A key problem refers to the differential inclusion of infected and uninfected individuals in serosurveys that combine the assessment of sociodemographic and behavioral data with basic laboratory findings (e.g. regular Elisa tests). In the context of generalized epidemics with limited access to antiretrovirals, differential inclusion of infected and uninfected individuals may merely reflect the effects of disease progression and death. In the same way, the dynamic character of background demographic factors (e.g. changes at the level of the demographic structure, some of them imposed by AIDS itself) can determine changes impossible to distinguish from actual incidence trends over time. On the other hand, in contexts with broad access to antiretroviral drugs, the opposite trend has been observed, as discussed in detail by Boily et al. The re-composition of high-risk networks after the introduction of highly active anti-retroviral therapy (HAART) may change the availability of infected partners seeking risky behaviors and reverse former protected behaviors of individuals inserted in such renewed networks.

Last but not least, most populations particularly vulnerable to HIV infection are marginalized and stigmatized populations, out of the reach of regular health services. As hidden populations, such individuals are difficult to enumerate, and will often be missed by simple random samples. In the same sense, in regular household surveys such populations are seldom reached, and once reached frequently deny behaviors that could be stigmatized by interviewers, family members or neighbors, or even criminalized, as in the case of the consumption of illicit drugs.

Contemporary responses to these methodological concerns include the use of sampling methods that try to circumvent such difficulties with the thorough exploration of network dynamics and/or the assessment of the spatial and temporal characteristics of populations, weighting for the inter-network versus intra-network variability and the patterns of spatio-temporal clustering. Spatio-temporal sampling and respondent-driven sampling have been used for this purpose.

The great divide: AIDS in two worlds

Whereas most areas deeply affected by the epidemic still lack even the most basic resources in terms of surveillance, as well as severe limitations in their ability to provide for the care and management of patients living with AIDS, resource rich countries, almost invariably harboring concentrated epidemics, have progressively moved to new frontiers of surveillance such as the proactive surveillance of HIV resistance. The best contemporary alternative to this divide has been the establishment of surveillance networks integrating developing and developed countries, such as the WHO-sponsored network on viral resistance. Unfortunately, the achievements of such global networks have thus far fallen short of their ambitious goals, due to political and logistical difficulties, as well as due to the scarcity of funds.
The worst surveillance is no surveillance

Despite the methodological and technical problems mentioned above, my firm position is that the worst surveillance is the absence of any surveillance. No lack of accuracy or precision should function as a disincentive for the continuous monitoring of the epidemic worldwide, especially in the most deprived settings. My long-term experience with the analysis of secondary data with graduate students from Lusophone African countries at the National School of Public Health, Oswaldo Cruz Foundation, Brazil, has been a matter of concern with respect to the overall low quality and scarcity of reliable data. However, it has also been a stimulus to reform and improve such national information systems. Each one of the trained students who return to their native countries and join the efforts of the local health authorities represents a prospect of change toward a more humane and balanced use of scientific progress.

Get back: resistance and the new pharmacovigilance

As treatment rolls-out in Sub-Saharan Africa, complex challenges emerge, such as the monitoring of the emergence and eventual transmission of HIV resistant strains. However, such concerns should not be used to deter current initiatives to increase access. Middle-income countries such as Brazil have shown that the forecasted “viral anarchy” need not take place, and that resistance levels similar to those observed in the US and Europe can be maintained over time, tracked by careful surveillance.

The time has come to reform the bases of AIDS surveillance. This reform should include a new emphasis on pharmacovigilance, concerned not only with monitoring new antiretroviral drugs launched in the world market, but also with monitoring new regimens for patients failing first-line regimens. Pharmacovigilance should also be extended to examine the complications arising among the cohort of aging AIDS patients, who face the diseases of middle-age as well as the accumulated side effects of major anti-retroviral drugs used over long periods of time (e.g. metabolic diseases).

As HIV-infected people live longer, co-infections with other viruses and bacteria with long clinical courses become more relevant, as is currently happening with hepatitis C. In the same way, the long-term maintenance of people with different comorbidities such as drug dependence poses new challenges such as the drug interactions between anti-retroviral medicines and methadone.

Final remarks

As discussed before, there are no easy solutions in the field of AIDS surveillance. Besides the continuous challenge to redefine methods and strategies in order to deal with the complexity and dynamic character of the pandemic, a major challenge is always posed by those population most in need, living in low and middle-income countries hardly affected by AIDS and a myriad of other social and medical problems.

As treatment roll-outs, finally reaching the millions of people living in Subsaharan Africa and Southeast Asia living and dying from AIDS, new layers of complexity, such as growing secondary resistance to different anti-retroviral medicines, will be added to the already complex situation. Much has been achieved by middle-income countries such as Brazil in its efforts to provide universal treatment and to establish a comprehensive system of monitoring and evaluation. The challenges ahead will require from researchers, health professionals, activists and policy makers the best of their talents and commitment, but the pandemic never gave to anyone really committed to curb it a single moment of solace.

So, it’s time to renew the energies of professionals and the civil society at large, and to train a new generation of researchers and health professionals, the ones who will seek the renewed alternatives for the renewed problems posed by the time to come.
Acknowledgments

Thanks are due to Cecilia Minayo for the kind invitation to write this paper, to Maya Petersen for her insightful comments and invaluable help, and to the anonymous reviewer who raised important topics I tried to incorporate in the revised version of the manuscript.

References


Artigo apresentado em 10/03/2008
Aprovado em 27/03/2008
Versão final apresentada em 02/04/2008