Overcoming health inequity: potential benefits of a patient-centered open-source public health infostructure

Superando a falta de equidade em saúde: benefícios potenciais de uma estrutura de informação em saúde pública, centrada no paciente e de domínio público

> Ernesto T. A. Marques Jr. 1,2 Romulo Maciel Filho ¹ Paul Nordstrom August ^{1,3}

Abstract

 Centro de Pesquisas Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, Brasil.
 Department of Medicine, The Johns Hopkins School of Medicine, Baltimore, U.S.A.
 Kenzen Systems Pte Ltd., Singapore City, Singapore.

Correspondence

Correspondence
E. T. A. Marques Jr.
Laboratório de Virologia
e Terapia Experimental,
Centro de Pesquisas Aggeu
Magalhães, Fundação
Oswaldo Cruz.
Av. Moraes Rego s/n, Campus
da Cidade Universitária,
Recife, PE
50670-420, Brasil.
emarques@cpqam.fiocruz.br
emaraues@jhmi.edu

The past 20 years have witnessed an astonishing increase in computational power and an incredible reduction in the cost of contemporary computer systems, but the public health infostructure in most countries has not changed significantly. This article discusses the potential benefits of applying patient-centered infostructure at the primary medical "points-of-care" services, based on networked integrated open-source technology and programming standards to develop tools to detect and reduce health inequalities. Such systems, which could be implemented from the local to the national level, would enable the expansion of evidence-based medicine, clearer identification of health inequalities, and more accurate cost-benefit analyses. In addition, the public health sector could link such databases to traditional Electronic Patient Record (EPR) systems at a greatly reduced cost by promoting the use of standards-based formats for data transfer and storage. Ultimately, the new health infostructure would help decrease health inequity. In fact, developing countries like Brazil, India, and South Africa are well-positioned to take advantage of the open-source movement and "leapfrog" countries burdened by legacy systems.

Information Systems; Equity in Access; Primary Health Care

Introduction

Modern medical "points-of-care" and support services such as radiology and pathology have been transformed by information technology in developed countries, with computers on every hospital ward, outpatient clinic, nursing station, and general practitioner's desktop. Despite all the glistening hardware and terabytes of data (and countless billions of dollars, euros, or pounds spent), a public health official would have a hard time detecting substantial improvements in public health across a population, or even within a local health care region. This lack of improvement is largely due to the fact that most medical information systems in use today are not designed to meet the needs of public health managers or facilitate the development of evidence-based medicine or public health. They use multiple proprietary (and often expensive) database software and a variety of formats that often do not communicate well with each other, if they can communicate at all. The result is an inefficient medical infostructure where end users are unable to transform the terabytes of data into useful information, which could guide better health care spending priorities and implementation strategies 1.

The public health situation in developing countries is quite different from that of developed nations for numerous and obvious reasons; however, the main problem for the public health ad-

ministrator is the same: lack of good, updated information to guide priority decisions and identify problems of delivery inequity and exclusion 2,3. This problem becomes even more critical in situations where resources are more limited, since they demand stricter definitions of public health priorities. More accurate cost-effectiveness analyses can allow identification of the problems with greater social impact and aid the implementation of effective strategies to solve them 4,5. In the last decade, costs of computing power hardware have reduced tremendously, whereas the costs of proprietary software licenses and maintenance fees have remained almost unchanged or increased, thus limiting the use of such software systems outside developed nations. Today the financial cost of implementing a hardware infostructure in few "sentinel" points-of-care units, strategically selected to be representative of larger regions, is a very small fraction of the total current public health expenses, likely less then 1%. This investment, even in very poor countries, could easily be recovered 6,7.

At this stage, it is critical that health policymakers in general and particularly those in parts of the developing world with a critical mass of software engineers (e.g. Brazil, India, South Africa) take advantage of the incredible power of open-source software and the surrounding community to launch a massive effort for creating a low-cost patient-centered health infostructure, designed to meet the information needs of public health organizations and make the structure available for global distribution. This also requires the use of standard data transfer and storage on all proprietary software purchased by governmental agencies. Combined, these actions have the potential to dramatically improve patient care, public health interventions, epidemiological studies, and biomedical research on neglected infectious diseases by creating more relevant banks of clinical samples and data.

While most public health officials have lacked such information tools 1, private managed care, health insurance, and life insurance companies have long since used medical, hospital, and pharmacy billing and patient demographic data to price their premiums and make decisions on insurance coverage and analyze specific treatments costs. Needless to say, health care is one of the world's most profitable industries. Meanwhile, legislation to protect patient information privacy and other ethical issues involved in the database design and protection is one of the most daunting challenges for the development of such medical and public health information systems. However, modern encryption, anonymization techniques, and databases with large numbers of subjects can prevent singling out an individual patient while ensuring privacy. Such technical approaches for preserving patient privacy are discussed elsewhere 8. As an example, we will briefly highlight a very low-cost information system designed specifically to optimize pre-clinical and clinical-epidemiological dengue vaccine research in Recife, Northeast Brazil, a less developed region of the country.

Limitations of existing health infostructures

Within a typical modern hospital the most one can usually hope for is a standard Electronic Patient Record (EPR) system linked to pathology, radiology, pharmacy, and other internal hospital databases. Rarely do these electronic data sets include or compile information from a patient's family doctor or general practitioner, community nursing, community clinics, social services, environmental, socioeconomic, occupational, or other relevant data sets created or managed elsewhere - for instance the census, geoprocessing, remote imaging, and public transportation systems, among others.

Data collection at the primary care level is generally much more difficult. Even when electronic information systems are installed in these services, they normally depend on a variety of proprietary, closed database systems that rarely allow data mining or data transfer between similar practices using the same software, let alone to local hospitals, official health organizations, or other potential data users. The lack of online or timely electronic communication between different health care providers leads inevitably to costly and sometimes even hazardous redundant procedures (e.g. unnecessary radiological tests). The absence or lack of access to synchronized medical and treatment histories is a clear problem.

Patients have usually visited the family doctor, been referred to the specialist, had numerous blood samples taken and other tests performed (often redundant), and have limited knowledge of, access to, or control over the medical data collected for them. Patients have no way of knowing what data exist, who has access to them, or whether there are errors (to name but a few problems), and each time patients visit a medical service, the clinicians have to begin collecting the same information all over. Meanwhile, public health specialists are attempting to simultaneously manage the health system, identify problems, find solutions, improve services, and grapple with a restricted budget, while basing their

decisions on information that is difficult to obtain, costly, old, unreliable, and scattered across numerous different fragmented health services that need to be reentered into a new database in order to be analyzed.

During an ongoing dengue vaccine development program at the Laboratório de Virologia e Terapia Experimental, Centro de Pesquisas Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, Brasil [Virology and Experimental Therapy Laboratory, Aggeu Magalhães Research Center, Oswaldo Cruz Foundation], several of these problems were encountered while creating a research database of patient clinical and demographic information. Patients were recruited at three hospitals in Recife - Hospital Esperança, Hospital Santa Joana, and the Maternal and Child Health Institute of Pernambuco (IMIP) - and the clinical pathology exams (biochemistry, hematology, and liver function tests) were performed at two different private laboratories, Cerpe Laboratórios and Laboratório de Medicina Diagnóstica (DILAB), and a public laboratory from the National Public Health Laboratories System under the Pernambuco State Office of the Ministry of Health (LACEN-PE). One laboratory could only furnish relevant hematology and biochemistry results in a non-standard format which required additional "parsing". Another laboratory was only able to provide image files of laboratory reports that could not be electronically imported as useful data, and the data for these patients had to be manually reentered into the database. Another lack of standardization was found in the regional and national dengue surveillance forms from the local public health authorities, which were different and cumbersome to fill in, leading to probable underreporting of dengue cases by busy clinicians. Importantly, most of the information was already being entered into some form of database at each point-of-care, spending more time and possibly leading to data entry errors, thus demonstrating a complete lack of coordination and standardization of patient medical information. Figure 1 illustrates the isolated nature and limitations of most existing medical information systems. Each institution may have had their own individual databases, but none of those systems had the flexibility to allow any form of data mining or knowledge discovery, and each official notification form had to be manually rewritten and sent to the epidemiological surveillance office, where it was reentered into another database. Thus, by the time the initial information is finally processed for analysis by public health specialists, several weeks have passed and many opportunities to act have been lost.

A patient-centered infostructure

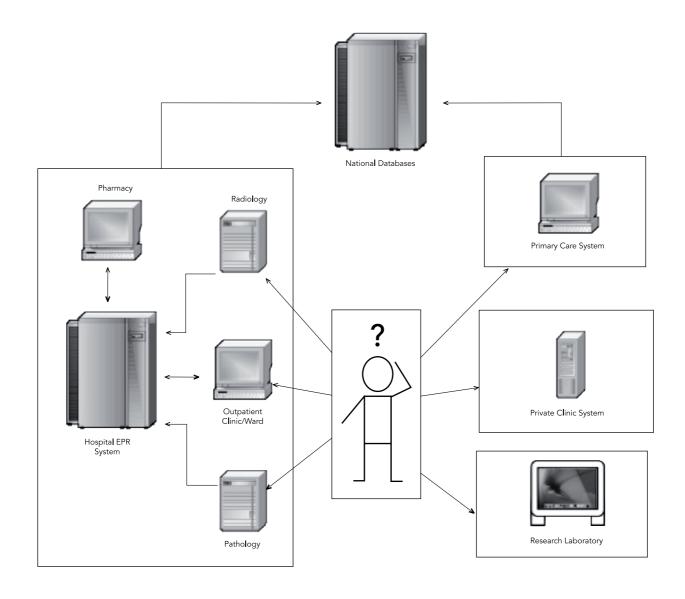
An alternative model places patients at the center of the infostructure, with ownership, access, and control over their biomedical data. This system is made possible by the use of powerful, inexpensive "LAMP (Linus, Apache, MySQL, and PHP) stack servers" running the open-source software Linux (operating system), Apache (web server), MySQL (database), and PHP (scripting language). The enormous advantage of this approach - apart from its inherent low cost (the primary expense being the server hardware) - is the ease of installation and the common set of tools and resources it offers. In marked contrast to the proprietary systems marketed by most commercial EPR and other informatics vendors, the open-source developer community thrives on discussion, sharing of knowledge and codes, and debugging strategies. Numerous websites exist expressly for the purpose of trading knowledge and solutions: see for example Planet MySQL (http://www.planetmysql. com), Planet PHP (http://www.planetphp.com), Planet Apache (http://www.planetapache.com), and MySQLForge (http://forge.mysql.com), among many others.

The global Wikipedia project (http://www. wikipedia.org) provides a useful model for the potential of LAMP-based systems and demonstrates how the use of a common interface facilitates complex data entry, editing, and viewing. Based in Tampa (Florida, USA) with additional nodes in Amsterdam (Netherlands) and South Korea, the Wikipedia handles hundreds of millions of hits per day. The "MediaWiki" software that powers the various Wikimedia products (Wikipedia, Wiki News, etc.) is itself open-source and can be installed on any LAMP server for local or customized use; e.g. a company or university could set up its own internal wiki (see http://www.medi awiki.org). By using the tremendous power of the LAMP stack coupled with appropriate database and interface development we can envisage a radically new health infostructure focused on the patient but enabling much more efficient data transfers between health care providers, policymakers, and biomedical researchers. A schematic model is presented in Figure 2.

Data "ownership" by the patient may seem unnecessary or an obstacle to implementation, but it can be greatly beneficial and crucial for long-term development. If successful, the medical staff knows what information is available and patients have control over which information can be released to a particular medical point of care, while the public health system has access to anonymous information. Given that under such

Figure 1

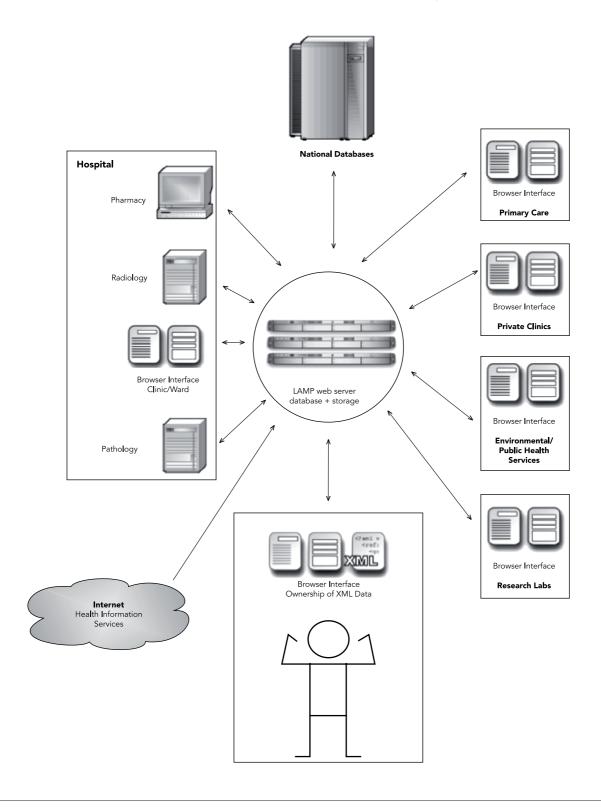
Schematic view of a typical current health infostructure. The patient is seen in multiple locations, each of which stores data in a proprietary system. There is little if any data transfer between points-of-care, and patients usually have no access to or control of their data. Selected data may be sent to national registries or research databases by hospitals, primary care providers, and other clinics, but often not in a coordinated or reliable fashion.



an infostructure a patient's data can be accessed and used in more ways than under current systems, it is critical that patients be informed and that they consent to such data sharing.

The design, development, and implementation of such highly complex electronic information systems optimized for public health uses is a great ethical and design challenge, but the technological tools to overcome these challenges exist. One key to a successful information system is careful selection of the information collected and managed by the system database. A second important key is the design of several user interfaces optimized to meet the needs and different levels of access to information by distinct users. Physicians have particular needs that vary by specialty, while nurses and social workers also have specific information needs to appropriately interact with their patients. However, the greatest challenge is to design user interfaces optimized for

An open-source LAMP-based health infostructure. Patient data are stored and distributed (with appropriate data security and access controls) from a regional node with connections to hospitals, public and private clinics, research laboratories, and national databases (e.g. cancer registries). Data collected at one point-of-care are accessible at others, chiefly by web-browser interfaces (XML data transfers are also possible.) The data set belongs to the patient, who has control over access and can view logs of transactions. Internet links to other information services would enhance patient and clinician knowledge.



different patients. Each step of the development process must be discussed carefully by a multidisciplinary group regarding the usefulness of information content and its ethical implications, as well as the definition of who has access to such specific sets of information, which information need to be anonymized, and which interfaces should be offered to patients and/or caregivers.

Currently, a large amount of information is collected at different points-of-care and annotated on several pieces of paper stapled together and distributed both within the institution and to other institutions. The process begins with patient registration, followed by triage, physical exam, medical observations, orders of exams and the results, medical prescriptions, and followup visits. Most patients are overwhelmed with scheduling and going through the entire process in an efficient and timely manner. Typically the administrative staff, nurses, and social workers need to assist patients, particularly the poor, less educated, and elderly who need much more help to actually use the medical services. The support personnel usually need to compile patient and medical information to assist patients in overcoming each step of the medical care process and its associated bureaucracy until patients finally reach the appropriate procedures. Managers are rarely able to adequately supervise and evaluate the efficacy of such assistance, basically provided by all medical points-of-care (but informally) and are thus unable to determine whether the resources allocated for this purpose are really used for the patients that need them the most.

The role of the medical support personnel continues to be important to assist patients through the information system. Some patients require little or no help, while others are never able to use the system alone. However, the information system makes the work of the entire support personnel much more efficient by providing rapid access to the information they need to help the patients, while their work is supervised and evaluated. A specific interface for users that require assistance can also be developed. Vulnerable patients probably benefit the most from such systems, since their caregivers deal with a more organized system.

One can envisage a system in which patients allow their data to be accessed by epidemiological and other types of research. Naturally, each system "opt-in" or "opt-out" has to be analyzed and approved by ethics committees before implementation. Electronic messages and/or patient consent forms can be sent to potential volunteers, inviting then to make their anonymous patient data available to be used in research and explaining their privacy rights. Data entry and updating by patients is greatly encouraged when they have a sense of autonomy and understand how the data are used. Ethics committees are also able to implement much closer oversight of the research process, protect the most vulnerable patients, and guarantee the safety of their information. Such information systems do not automatically create ethical dilemmas that do not already exist. The main difference is that medical administrators and staff become much more accountable for decisions that impact patient care.

Designing a public health infostructure involves several steps, beginning with the selection of the information to be collected and the ethical implications. This actually means an improvement on the current situation, since under the current system, all medical records already have some sensitive information, but no one discusses which information is appropriate for inclusion. A key advantage of an electronic system over the current paper records is the ease in extracting and analyzing information from the records. In our current system, ethics committees have to analyze research proposals and authorize researchers to search the records, but committees have no control over what the person actually sees while going through paper medical records. In contrast, with such an electronic information system, the hospital ethics committees has much greater control over exactly which information is released and to whom. After determining the initial database content, a period of simulated user testing and focus group testing is required to develop and optimize the interfaces according to different system users' needs. Ideally, patients involved in later stages of development in the information system choose whether they want to participate, and prototype testing is performed in medical schools, due to the system's experimental and medical training characteristics. After confirming the prototype, routine use can be implemented in a small number of selected patients until the system is functional and approved for unrestricted use.

Obviously, not all patients can afford computers or have the skills to use them, so in low-income areas, medical points-of-care like hospitals, clinics, and public health programs can establish places for patients or their legal proxies to access and update data with the assistance of the clinic personnel, if necessary, while awaiting their physician appointments. The simple timesaving and clinical benefits of having core medical data (e.g., on allergies, medical history, and current medications) available at any point-of-care potentially pays for system implementation costs.

The great medical and economic benefits of this streamlined infostructure are already obvious, but in a sense it is just the beginning. The open-source community is constantly creating exciting new tools, many of which can be adapted for biomedical and public health use. One creative example is the GarbageScout system developed by Jim Nachlin (see http://www.garbagescout.com). Using readily available open-source tools, Nachlin created an innovative and user-friendly system for tracking discarded items of value in several American cities.

Such tools can become invaluable aids in public health interventions and epidemiological studies. For example, communities can be encouraged to submit photos and locations of hazardous standing water in dengue-prevalent areas. Actual dengue cases could automatically be plotted on local maps, drawn from hospital and primary care diagnoses and laboratory results.

Thanks to the low set-up cost of LAMP servers and the open-source nature of the proposed software, the system could be installed at virtually any level, from a small local health district through State or regional districts and even at the national level. Ideally the software would be developed by an international team and released in a variety of languages, just as existing open-source packages are done regularly and freely.

The "Leapfrog Phenomenon"

Interestingly in terms of overcoming global health inequities, the developing nations are in a unique position to exploit the potential of a patient-centered, LAMP-based health infostructure by virtue of starting with a fairly blank slate. Just as many developing countries (notably those in Southeast Asia and South America) have "leapfrogged" expensive and obsolescent landline telephony and become nations of mobile phone users, countries that have still not invested in proprietary, non-integrated, and technologically obsolete health information systems will have the easiest time implementing a new model.

Dengue vaccine development database

Dengue vaccine development poses several major challenges. In order to enable more efficient vaccine development, a database was specially designed to curate data from a dengue patient cohort relevant to pre-clinical and clinical dengue vaccine research. This database only contains data from volunteers who enrolled in the dengue research study approved by the Ethics Committee of the Aggeu Magalhaes Research Center, Oswaldo Cruz Foundation, and the Comissão Nacional

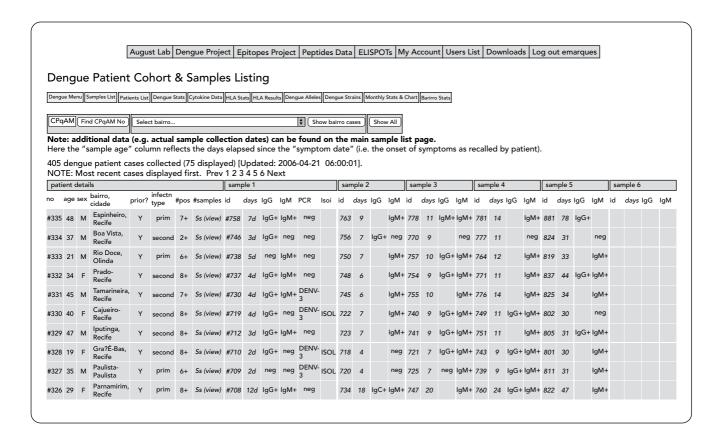
de Ética em Pesquisa [CONEP; National Research Ethics Council] and who signed informed consent. The same principles applied in this database can be applied to databases with broader applications for use by public health officials and not necessarily only for research or publication purposes, and thus may not require informed consent. The current database contains data collected at several points in time and more than 200 data fields ranging from clinical symptoms, standard biochemistry laboratory data (albumin, liver function tests, etc.), serology (dengue IgM and IgG), and hematological data (platelets, hematocrit, leukocyte count), viral genotype (polymerase chain reaction, or PCR, quantitative PCR, or qPCR, and viral isolation), viral load, patient HLA types for the DR, DQ, A, B, and C loci, plasma cytokine levels (interleukins 2, 4, 5, 10, interferon-g, and TNF-a) and enzyme-linked immunospot (ELISPOT) responses to various viral peptide pools, geopositioning (GPS) references for the home address, mosquito vector infestation levels, and mosquito infective status, to name a few. A facilitator at each sentinel hospital initially gathers the data and interacts with volunteers, attending physicians, and the pathology laboratory. Facilitators have a specific questionnaire prepared to efficiently guide their interaction with the different actors on the information chain without excessively occupying their time. Samples and data forms are then delivered to a central laboratory, the data are digitalized, and the samples are distributed through the laboratory for proper preparation and storage. A case manager with access to the database coordinates the patient follow-up, interacting with patients, the hospital facilitator, and clinicians to ensure that all the necessary information flows efficiently and appropriately (Figure 3). All reporting forms are generated electronically and sent to the epidemiological surveillance departments. Considering the three hospitals selected as sentinels, real-time monitoring of dengue cases could be used as a warning system to indicate when and where control measures are required to contain a dengue outbreak. Figure 4 shows the seasonality of the number of suspected and confirmed dengue cases in each month.

Conclusion

Countries that have reached a critical mass like Brazil, India, and South Africa can develop a public health information system with the use of powerful open-source software, codes, and standards at very low cost that can be implemented in sentinel clinics and enable evidence-based

Figure 3

Dengue patient follow-up schedule and planner. The patient case manager can follow the status of each case and monitor exam results and diagnoses through a specific database view.



public health management. We believe that public health databases alone cannot replace well-designed epidemiological studies, but can allow more efficient and faster measurements of cost-efficacy, inequality, and exclusion, and generate data-driven hypotheses and target epidemiological studies to the questions not covered by the database. These tools can provide efficient compilation of the necessary data to allow public health managers to identify problems and define priorities using high-quality information provided in real time by "sentinel"

points-of-care. Once the information system is implemented at the sentinel points-of-care, it can provide a continuous cyclical flow of information, as demonstrated in Figure 5. Once sufficient information is gathered, epidemiologists and public health officials can analyze the data and define actionable measures. As these measures are implemented, new data are entered into the dynamic information system, the impact of the measures is quantified, and a new cycle of measures is launched, leading to continuous optimization of the health care system.

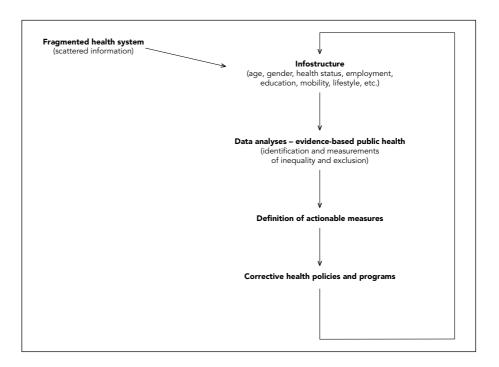
Figure 4

Monthly distribution of suspected and confirmed dengue cases.

Month	Patients	Pos pats pct pos		patients positive (+)/negative (o)
	2	1	50.0	+ o
Feb 2004	10			000000000
Mar 2004	19	8	42.1	++++++0000000000
Apr 2004	17	15	88.2	++++++++++++00
May 2004	23	12	52.2	++++++++++00000000000
Jun 2004	13	6	46.2	+++++000000
Jul 2004	7	4	57.1	++++000
Aug 2004	11	1	9.1	+000000000
Sep 2004	6	1	16.7	+00000
Oct 2004	5	1	20.0	+0000
Nov 2004	7	2	28.6	++00000
Dec 2004	5	3	60.0	+++00
Jan 2005	7	3	42.9	+++0000
Feb 2005	7	1	14.3	+000000
Mar 2005	9	3	33.3	+++000000
Apr 2005	14	2	14.3	++00000000000
May 2005	21	11	52.4	+++++++++000000000
Jun 2005	23	20	87.0	+++++++++++++++++++++++++++++++++++++++
Jul 2005	37	27	73.0	+++++++++++++++++++++++++++++++++++++++
Aug 2005	17	14	82.4	++++++++++++
Sep 2005	13	3	23.1	+++000000000
Oct 2005	9	5	55.6	+++++0000
Nov 2005	15	8	53.3	++++++000000
Dec 2005	8	5	62.5	+++++000
Jan 2006	23	13	56.5	+++++++++++0000000000
Feb 2006	14	8	57.1	++++++000000
Mar 2006	30	12	40.0	++++++++++00000000000000000000000000000
Apr 2006	33	6	18.2	+++++0000000000000000000000000000000000

Figure 5

Schematic representation of a cyclical information flow.



Resumo

Este artigo discute os benefícios de se desenvolver um sistema de informação em rede, centrado nos pacientes dos serviços médicos primários, usando tecnologias open-source e definições de padrões de programação e o desenvolvimento de ferramentas capazes de detectar desigualdades. Esses sistemas, que podem ser implementados em nível local e gradualmente se expandirem para uso nacional, capacitariam a expansão da prática da medicina baseada em evidência, a identificação mais clara das desigualdades e análises mais precisas de custos-benefícios. Os setores de saúde pública também poderiam interligar esse sistema aos prontuários eletrônicos tradicionais a custos muito reduzidos por meio da promoção do uso dos sistemas padrões de estocagem e transferência de dados nos produtos de informática comerciais usados nos serviços de saúde. Em sua forma final, esse novo sistema de informação em saúde seria capaz de assistir à gestão da redução das desigualdades nas medidas de saúde pública. De fato, países em desenvolvimento como Brasil, Índia e África do Sul estão muito bem posicionados para darem um salto na frente de outros países que já estão comprometidos com sistemas de informações antiquados.

Sistemas de Informação; Equidade no Acesso; Atenção Primária à Saúde

Contributors

E. T. A. Marques Jr. participated in drafting the article, supervising the design, and developing the database. P. N. August participated in drafting the article, developing the design and database, and conducting the programming. R. Maciel Filho participated in drafting the article, supervising the design, and developing the database.

Acknowledgments

The authors wish to thank Dr. J. Thomas August for valuable discussions that greatly improved the ideas presented in this report.

References

- 1. Berwick DM, DeParle NA, Eddy DM, Ellwood PM, Enthoven AC, Halvorson GC, et al. Paying for performance: Medicare should lead. Health Aff (Millwood) 2003; 22:8-10.
- 2. Baron JA, Weiderpass E. An introduction to epidemiological research with medical databases. Ann Epidemiol 2000; 10:200-4.
- Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. BMJ 2003; 326:1070.
- 4. Eddy D. Bringing health economic modeling to the 21st century. Value Health 2006; 9:168-78.
- 5. Eddy DM. Linking electronic medical records to large-scale simulation models: can we put rapid learning on turbo? Health Aff (Millwood) 2007; 26: w125-36.
- 6. Eddy DM. Evidence-based medicine: a unified approach. Health Aff (Millwood) 2005; 24:9-17.
- 7. Schlessinger L, Eddy DM. Archimedes: a new model for simulating health care systems - the mathematical formulation. J Biomed Inform 2002; 35:37-
- Peterson MG. Privacy, public safety, and medical research. J Med Syst 2005; 29:81-90.

Submitted on 19/Set/2006 Final version resubmitted on 24/Apr/2007 Approved on 12/Jun/2007