

Published in final edited form as:

Acad Pediatr. 2011 ; 11(4): 280–287. doi:10.1016/j.acap.2011.02.007.

Electronic Medical Records (EMRs), Epidemiology, and Epistemology: Reflections on EMRs and Future Pediatric Clinical Research

Richard C. Wasserman, MD, MPH^{*,+^}

^{*} Department of Pediatrics, University of Vermont College of Medicine, Burlington, VT

⁺ Pediatric Research in Office Settings (PROS), American Academy of Pediatrics, Elk Grove Village, IL

[^] Children's Hospital of Philadelphia, Philadelphia, PA

Abstract

Electronic medical records (EMRs) are increasingly common in pediatric patient care. EMR data represent a relatively novel and rich resource for clinical research. The fact, however, that pediatric EMR data are collected for the purposes of clinical documentation and billing rather than research creates obstacles to their use in scientific investigation. Particular issues include accuracy, completeness, comparability between settings, ease of extraction, and context of recording. Although these problems can be addressed through standard strategies for dealing with partially accurate and incomplete data, a longer term solution will involve work with pediatric clinicians to improve data quality. As research becomes one of the explicit purposes for which pediatricians collect EMR data, the pediatric clinician will play a central role in future pediatric clinical research.

Introduction

The electronic medical record (EMR) is transforming clinicians' day-to-day care of patients.¹ In their simplest form, EMRs are analogous to the individual patient paper charts that were developed and used throughout the 20th century to recall observations, inform others, gain knowledge, monitor performance, and justify interventions.² From the vantage of individual physician clinics and offices, hospitals, and health care systems, the EMR is first and foremost a tool for clinical use. As with the paper patient record, the EMR represents the legal record, plays an important part in charge capture and proper billing, and may in fact have some advantages in this regard.³ From the vantage of patients, the EMR of a single office or health system may contain only a portion of electronic information stored about them. The sum of stored electronic health information, which some term the electronic health record or EHR, includes information derived from care provided by multiple sites or systems of care within a community, region, or state.⁴

CORRESPONDING AUTHOR: Richard C. Wasserman, MD, MPH, University of Vermont, N310 Courtyard at Given, 89 Beaumont Ave., Burlington, VT 05405, PHONE/FAX: 802-656-5455 / 802-656-4545, richard.wasserman@uvm.edu.

DISCLOSURES: The author has no conflicts of interest to disclose.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

EMRs have been used in a few locations for several decades,⁵ and adoption in pediatric primary care settings has been increasing steadily.^{6–8} Given the purported benefits to clinicians as well as the incentives and disincentives of the Electronic Health Record Incentive Program of the Centers for Medicare and Medicaid Services,⁹ it seems likely that nearly all U.S. pediatricians will be using EMRs to provide care in the coming decade.

Some have touted health information technology as a path to improving the health care system.¹⁰ Others have expressed skepticism about the ease with which this can be accomplished.¹¹ The ideal of a “Learning Healthcare System,” through which one could garner useful information and thereby learn from every patient encounter^{12, 13} in something approaching real time, presents both a challenge and an enormous opportunity. Computer-generated administrative data from submitted claims long have been used for epidemiologic and health services research^{14, 15} and quality assessment.¹⁶ However, claims take significant time to be processed and are related to what the clinician chose to charge for, as opposed to documenting the clinically important activities actually took place during a medical encounter. Analysis of the clinical data available in EMRs represents an enormous advance beyond claims data analysis, as the EMR offers specific information about clinical findings and even the pediatricians’ thought process. The concept of using the EMR to create research databases is not new,¹⁷ but the substantial capital investments required to install and maintain EMRs, combined with the expertise needed to transform EMR data to a form usable for research, have initially confined such research to staff model HMOs and their research institutes¹⁸ and to large academic medical centers. Because U.S. primary care pediatricians appear to have been slow in adopting EMRs¹⁹ and possibly because cost-benefit considerations discourage devoting resources to extracting pediatric primary care EMR data for research purposes, studies using pediatric EMR data have been limited.

The objective of this paper is to (1) inspect several core issues regarding using routinely collected EMR data for pediatric clinical research and (2) suggest possible ways for improving the process. The term *clinical research* is here used as defined by Feinstein as “systematic plans for discovering facts or principles” in the clinical realm “in a group or groups of people.”²⁰ Such research would include epidemiological investigations to determine associations between exposure and disease, health services research involving patient outcomes, and observational comparative effectiveness studies to weight the effectiveness of one therapy against another. The discussion purposefully excludes other very important types of inquiry on the pediatric research continuum, such as quality improvement activities, that could also make use of EMR data, but that are not intended to produce generalizable new knowledge. The discussion focuses on EMRs as they are used in pediatric primary care, but also is relevant to EMR data from other pediatric clinical settings. Vis-à-vis the title, the author is not an informatician, epidemiologist, or expert in epistemology (defined as the study of the nature, sources, and validity of knowledge),²¹ but a clinical researcher. The information presented here is therefore basic and intended for clinical researchers seeking a rudimentary understanding of how EMRs can be used in pediatric clinical research. For more comprehensive information about biomedical informatics, the reader should explore recent textbooks^{22–24} and journals (see Table).

Electronic Medical Records and Clinical Research

Datasets Available for Pediatric Research

As has been noted elsewhere, a proliferation of existing datasets currently are available to pediatric researchers.²⁵ They include the Kids’ Inpatient Database (KID),²⁶ National Health Interview Survey (NHIS),²⁷ the National Health and Nutrition Examination Survey (NHANES),²⁸ National Ambulatory Medical Care Survey, (NAMCS),²⁹ Youth Risk Behavior Surveillance System (YRBSS),³⁰ Medical Expenditure Panel Survey (MEPS),³¹

and others. These data sources have some distinct advantages for research, benefitting from multi-stage probability sampling of national populations and the use of survey items with known psychometric properties. They are however, more or less removed from the actual care of patients, involve selected group of subjects who've agreed to participate in research, and although they may allow for some longitudinal analysis, are less than ideal for outcomes research. One might argue that the best data for research on clinical outcomes are data gathered in the day-to-day care of patients.

The nature of EMR data

As opposed to research data derived from surveys, EMR data are the electronic clinical records of medical encounters. These data are in most ways highly comparable to those in paper records, typically organized in tabs with elements such as problem, medication and allergy lists, and visits organized according to history, physical examination, assessment, and plan. Data in primary care computer systems are stored as either narrative or structured data. Narrative data are word-processed free text as typically found in paper records. Narratives are essential to the medical endeavor. The patient medical history is, by definition, a narrative, as is the oral case presentation. Narratives are dearly valued by clinicians in properly describing patients' problems. Structured data, in contrast, are the somewhat less descriptive data that correspond to codes and are entered by "clicking" on choices presented in lists, forms, or templates.³² Structured data are helpful in establishing consistency and minimizing misunderstanding precisely because of the many free text options available to clinicians for naming and/or abbreviating similar conditions (e.g., "asthma," "reactive airway disease," "RAD"), symptoms (e.g., "fever," "elevated temperature," "high temp") and signs (e.g., "crackles," "rales," "crepitations"). Structured data are much easier to extract from EMRs than narrative data. A third category of data are images (e.g., PDFs of radiological studies or scanned documents and reports), which are extremely challenging with respect to extraction.

Potential utility of EMR data for research

Information derived from EMR data can be of enormous use for researchers in generating new knowledge. For example, the specific values of EMR data on child birth date, visit date, sex, height, and weight can be used to provide information on patient BMI and, in turn, on the prevalence of pediatric obesity in a population. Or, the values of HIV assays and CD4+ T cell counts in the EMR can be used with decision rules to determine the prevalence of pediatric AIDS in a population of children at risk. In these examples, the data elements themselves require some interpretation or inference to provide useful information, with data extraction as a necessary first step. Other EMR data, however, directly provide useful information without interpretation or inference. For example, an EMR diagnosis of "lobar pneumonia" tells the researcher that the clinician judged the child to have lobar pneumonia. With information either derived or directly obtained from EMR data, the researcher can thus identify patient samples for cohort or case-control studies or potential subjects for randomized controlled trials with efficiency and precision. For example, a cohort of patients placed on atypical antipsychotic medication could be assembled and followed forward in time to assess long term impact on BMI. Or, patients with Henoch-Schönlein purpura and controls matched for age, sex, and zip code could be identified to investigate the association between recent viral infection and disease. The possibilities for such epidemiological studies are legion. In addition, many of the variables needed for epidemiological study can be obtained from EMR data in anonymized fashion without having to query patients directly. As such, EMR data represent an electronic goldmine for researchers. As with an actual goldmine, however, potential challenges and hazards await the miner.

Accuracy of EMR data

In considering the accuracy of EMR data, it is first worth noting that paper records have been found to have only moderate correlation with what *actually has occurred* in the medical encounter. In one study in which the physicians were unaware that they were being studied and which employed the reports of scripted standardized patients³³ as the gold standard for what had occurred during medical encounters, agreements between documentation in the paper record and the reports of the standardized patients were modest – highest for therapies prescribed and laboratory tests ordered (68% and 64%) and lowest for history obtained and physical examinations performed (29% and 31%). A second study using standardized patients with a similar blinded design produced similar results.³⁴ There is reason to believe that if the pediatrician prescribes medication via electronic prescribing, orders diagnostic tests, or arranges referrals using the EMR, then accuracy would be improved, as the chance of the medications, diagnostic tests or referrals being undocumented would be largely eliminated. On the other hand, there is no reason to suppose that EMR history or physical exam data would necessarily be any more accurate. In fact, if defaults (i.e. settings assigned automatically by the EMR that must be overridden by the clinician) are set to “normal” in physical examination templates, EMRs could actually worsen accuracy, with the EMR indicating an exam with normal findings when elements of the examination either had not been performed or were abnormal but not noted as such. As with paper records, EMR data depend on what the clinician chooses to record in the act of caring for the patient. In many ways, the large scale retrospective extraction of clinical data from EMRs could easily be termed “chart review on steroids.” The simple fact that the data are electronic does not confer upon them a higher epistemological status, as the EMR data have been collected for clinical and billing purposes, and not to satisfy research objectives.

Challenges of extracting data from the EMR itself

EMR designers use data processing architectures that maximize real-time transaction processing to create highly operational systems and that differs from the data architecture required for research analyses.³⁵ The database architecture established for real time transaction processing facilitates rapid and detail-oriented clinical searches and updates on individual patients. To allow for efficient research analyses, data must first be warehoused and different database architecture established to allow for data analyses across populations of patients over time. Once warehoused, online analytical processing can be applied to facilitate the multidimensional and aggregated queries that are applied to the data in research.³⁶ Although EMR vendors often support a set of data extraction tools that work reasonably well to create reports and summaries with small numbers of variables, these tools begin to fail when the number of variables rises. Thus researchers need to employ an extract, transform, and load strategy (known in the trade as “ETL”)³⁷ that pulls data from the EMR, transforms (e.g., standardizes) the data for research needs, and loads it into the data warehouse.³⁸ Analysis can then be readily done on warehoused data using standard statistical tools.

Research across health care settings and the multiplicity of EMRs

According to the American Academy of Pediatrics (AAP) Council on Clinical Information Technology website, there are nearly 96 products in current use in primary care pediatrics.³⁹ These many products are not interoperable, as they are written in many different computer languages and have different data storage architectures. Thus, the large variety of EMR products presents a formidable challenge to research. However, the challenge is not insurmountable. The HMO Research Network⁴⁰ and Distributed Ambulatory Research in Therapeutics Network (DARTNet)⁴¹ are examples of two large research efforts extracting data from multiple EMRs in such a way that the data can be analyzed from these diverse sources. Interestingly, the approach used by these groups does not require a centralized

database. This model, which is called a distributed or federated database, provides mechanisms for combining data from individual sites and for coordinating data sharing activities among autonomous components. A discussion of how this is accomplished is beyond the scope of this article, and the reader is referred to a recent summary on the topic.⁴² Of note, such efforts require substantial investments in information technology and informatics resources.

The challenge that multiple EMR products present to research with EMRs

It is likely that EMR vendors did not consider data extraction for research purposes as a high priority when designing their software products. Storage of particular data elements often varies from product to product. However, even if the pediatricians in different settings used the same EMR product, obstacles would remain to data extraction from the different settings. This is because the data definitions and site of EMR storage for many clinical data elements are not necessarily specified by the EMR vendors, and even when specified, can be changed. An EMR product typically allows the clinicians and or informatics staff in charge of the local implementation considerable flexibility in making such decisions. As such, the definition of data items and the storage of those items within the same EMR product might differ from practice to practice or from institution to institution according to local design. For example, exposure to second hand smoke is a major pediatric health issue. In a single EMR product implementation, researchers seeking information on second hand smoke exposure might find it in free text on the problem list, in the social history, or in a visit about asthma, as a structured item in any of those locations, or as ICD-9-CM code E869.4 on a billing claim. This situation really is no different from the documentation vagaries that challenged the use of paper records for research, but illustrates how EMRs do not necessarily improve how research data are captured and organized.

New technologies to facilitate EMR data extraction for research purposes

The unstructured data in the free text narrative parts of the EMR provide a rich potential source of information (e.g., markers of socioeconomic status, information about daycare attendance, details of diet). Natural language processing tools show considerable promise in converting free text into structured data,⁴³ but these tools are far from perfected.⁴⁴ In a recent unpublished analysis of free text for patient encounters for otitis media *on a single day*, a research group at the Children's Hospital of Philadelphia encountered 278 different ways in 465 EMR notes to express the fact that patients had temperature > 102.0 F (e.g. "fever x 3 days max t 102.3 fever in the past 24 hours") and 123 different ways to express ear pain in 213 patients (e.g., "since last night ears hurt"). Nevertheless, progress is being made, with continued development of standards to allow the interoperability of data queries, as well as the precise logic to actually perform the analyses. For example, the Integrating Biology and the Bedside (i2b2) suite of software contains an optional software module that can be used to manipulate text reports and extract specific terms from them.⁴⁵

Another approach is to design EMRs specifically with data extraction in mind. The details of how this can be accomplished are beyond the scope of this paper, but two principles can be mentioned. One strategy, already touched upon, is to maximize structured data entry and minimize free text. A second is to build into the EMR rules-based paradigms allowing one to infer higher level concepts, such as presence of a disease from lower level concepts, such as a highly specific laboratory result and/or a list of drug combinations given only to patients with specific diseases. A trivial example might be the paradigm that a positive PPD test and treatment with isoniazid and rifampin automatically classify a patient as having tuberculosis. Researchers at the Regenstrief Institute have created an open source system for creating such EMRs, OpenMRS,⁴⁶ which show great promise in this regard, as it is designed specifically to improve data extraction. OpenMRS is not a ready-to-use EMR, but instead a series of

“building blocks” that allow designers to build their own EMRs through an application programming interface (API) – a set of specified commands, functions, and protocols which can be used in building software. With respect to data extraction from different care sites, this approach allows for the automated combination of EMR data from patients in multiple clinical settings using OpenMRS (e.g., primary care and specialty care settings with different available laboratory testing or diagnostic protocols) by acknowledging the different local ways of establishing the diagnosis. As such, higher-level extraction such as the definition of a group of patients with a particular diagnosis could be accomplished from the different sites, in spite of the sites using differing clinical protocols. Although this approach presents its own methodological challenges (e.g., as to whether the patients from the different sites are truly equivalent as to diagnosis), it might be of use for groups of research-oriented practices or clinics wishing to set up EMRs with a specific focus on data extraction for particular research questions. It cannot, however, be employed with clinical sites using existing EMRs.

For existing EMRs, centralized Web-based tools are being developed to extract data from the EMR using standard, non-product-specific approaches. These tools, known as Web services, are a promising development. A Web service API well known to most of us is the one used when we enter credit card information to purchase an item on line. The authorization of that purchase typically is made via a Web service, which checks with our credit card company to make sure that we’re legitimate users of the card and have not exceeded our credit limit. The Web service, which is invisible to the user, then relays that information back to the on line vendor, which did not have to create this capacity on its own. Defining APIs specifically for Web services that can interact with existing vendor EMRs and extract data from them is an essential next step for EMR-based research. For example, a Web service could accept data from multiple EMRs on children’s sex, height, weight, blood pressure, date of birth, and date of visit to identify a group of children classified as hypertensive according to NHLBI-defined standards.⁴⁷ The establishment of strong security policies and procedures will be among the many challenges to implementation across health systems of Web services that interact with EMR products to facilitate research.

Other challenges of EMR data for the researcher

EMR data present the researcher with a variety of methodological challenges. One is the issue of case identification. EMR data allow researchers to move beyond the ICD billing codes for case identification, but then create an overabundance of options. With childhood asthma, one could consider the EMR permanent problem diagnosis list (probably more specific, but less sensitive than an ICD code), pulmonary function data (useful only for persistent asthma), prescriptions for bronchodilators (highly sensitive, but not specific), or prescriptions for inhaled corticosteroids (specific but not sensitive), or a combination of these. To best identify cases, the researcher may need therefore to apply a complex case identification data rubric on warehoused EMR data. A second issue, when studying acute illness, involves specifying individual episodes of disease. In this case, the unit of analysis desired for the research dataset may not align well with the underlying clinical data. For example, a researcher might want to study individual episodes of acute otitis media or asthma. The EMR holds patient and encounter level data, as well as test order, test result, and prescription data. Bringing prescription data to the encounter level can become complex. For example, for a study of treatment breakthrough in otitis media, “find new otitis media encounters where the patient was on antibiotics” becomes a task of looking back to prior encounters and, based on order information, prescription end date, etc., determining time periods when the patient might have been on antibiotics. A third challenge, from an epidemiological perspective, is the lack of standards for defining denominator populations from EMR data. Although such standards exist for patients in closed panel HMOs, where

the site of health care is linked closely to the patient's insurance coverage, they are ambiguously defined outside of those settings, and make the comparison of rates across clinical sites extremely problematic. The three challenges cited here, as well as many others not discussed, are not at all unique to EMR data. They exist for any studies based on medical record data. The irony, however, is that (1) it may be as or more challenging to employ informatics techniques to define cases of disease, episodes of illness, and denominator populations than to use human coders to perform the same tasks and (2) the computer-based process ultimately needs to be validated by a human decision-maker who reviews a sample of records by hand.

Improving data quality is a challenge that may be more readily met in EMRs than in paper records. One obvious solution alluded to previously, is to encourage entry of additional data elements in structured or coded fashion. Of course, many data elements such as diagnoses, procedures, and pharmaceuticals are already associated with specific codes, but the overwhelming preponderance of clinically relevant data typically are entered as narrative. Clinicians accustomed to recording data in narrative fashion, however, may resist entering data in structured fashion.⁴⁸ In addition, in particular cases, such as intensification of medication for hypertension, the two types of data have proven complementary, suggesting that one will not necessarily replace the other.⁴⁹

A workshop conducted in 2006 by the European Federation for Medical Informatics Primary Care Informatics Working Group generated a list of recommendations for researchers collecting and processing routinely collected EMR data.⁵⁰ Two of the group's ten recommendations address the circumstances under which the data are collected. One of the potential advantages of EMR data is the ability to collate observations from very large numbers of practice sites. In this light, it is worth considering the recommendation to "record the characteristics of the practices involved in the study and how they might vary from 'usual' practice." Data on disease incidence and prevalence could easily vary according to whether the practice site was a primary care or specialty clinic, or whether the practice was located in an urban inner city or rural area. As in any research, the issues of standardized case definition and generalizability must be considered before lumping EMR data from diverse practice sources.

Another working group recommendation was to "describe the context of data recording." Context is easily overlooked, but is in fact fundamental to the nature of information recorded. The fact that data exist in the EMR does not lend them any particular epistemological status. Berg and Goorman have written that Information should be conceptualized as always entangled with the context of its production, and that disentangling the information to make it useful in a different context requires work.⁵¹ For example, let us take something as apparently straightforward as EMR data on height or weight. Putting aside the issues of the units (e.g., centimeters versus inches) and granularity (e.g., to the nearest centimeter versus nearest millimeter) in which the data are recorded, there is also the measuring process itself. Due to differences in the training of medical personnel and the equipment available, a height obtained on a patient in the context of a general pediatric practice is not likely to be identical to the height obtained on the same patient in a pediatric endocrinology clinic. Due to differences in clinical routines, a weight obtained on an infant in an over-air-conditioned office, with the medical assistant allowing underclothes to remain on the infant, will not be the same as the weight obtained on the same infant in an overheated office where the infant is fully naked. The reader can use her imagination to further expand on this theme. The differences in the clinical contexts in which data are collected (type of practitioner, intent of practitioner, past history of the patient, location of the practice, time of day, day of the week, etc.) all could affect the meaning of particular EMR data elements. This does not mean that the EMR data are unusable for research, but

that the researcher needs to be thoughtful about their use. For example, if a researcher is tracking growth over long periods of time, where the precision of single measurements of height or weight data are not critical to the research question, then there is less reason for concern. If, on the other hand, the researcher is seeking to establish new standards for height and weight, then the researcher needs to be very careful in using the data. Even more caution is called for when the EMR data involve discrete codes at a level necessarily involving more judgment and subjectivity on the part of the clinician. We already have mentioned ICD diagnoses codes. When relying on EMR diagnosis ICD codes, the researcher must consider several factors. These could include the billing context in which the clinician chose the coded diagnosis (e.g., clinician might choose a diagnosis that allowed for higher billing) or the ease with which the clinician is able to find and select the diagnosis in the system (e.g., choosing “anemia” versus “transient erythroblastopenia of childhood”). In a thought-provoking ethnographic study of general practitioner EMR coding in Great Britain and Denmark, Winthereik describes the need to take into account how the workload involved in coding and the local uses to which the coded data are put affect the accuracy of symptom and diagnosis codes. He states that producing accurate data is not by definition impossible, but that the data are only “accurate in relation to the context of use.”⁵² van der Lei, in his 1991 editorial entitled “Use and Abuse of Computer-Stored Medical Records,” stated a more radical view in his first law of medical informatics and its collateral: “Data shall only be used for the purpose for which they were collected,” and “If no purpose was defined prior to collection of the data, then the data should not be used.”⁵³ This law and its collateral suggest a way out of the problem of research and EMR data context.

EMRs, Epidemiology, and Epistemology: Concluding Thoughts

Epidemiology can be defined as the study of the distribution and determinants of disease in human populations;⁵⁴ epistemology, as mentioned previously, is the study of the nature, sources, and validity of knowledge. As noted above, the use of EMRs for epidemiological studies presents epistemological challenges. Should van der Lei’s first law of medical informatics – “Data shall only be used for the purpose for which they were collected” – preclude researchers from using EMR data, collected for the purposes of clinical care and billing, to generate new knowledge about the distribution and determinants of disease? Clearly, the answer is “No,” but van der Lei’s warning should be taken to heart. While the potential for efficiently generating new knowledge from the large numbers of clinical encounters documented in EMRs certainly exists, the epistemological issues should be recognized and addressed. Data not gathered specifically for research purposes *may* be incomplete and unreliable. This does not make those data unusable, but it does constrain the uses to which the data can be put and the inferences that can be drawn from them. Less-than-complete data of less-than-certain accuracy are not uncommon in clinical research. With care, data quality checks can be combined with selected epidemiological and statistical tools to address many of the problems presented by such data.

A better long term strategy, however, might be to undertake the important work of improving the validity and reliability of clinical EMR data in order to make them more useful for research. Ideally, our goal should be to shape at least some of the data elements so that research is, with a nod to van der Lei, one of the purposes for which the data were collected. Improving the quality of EMR data for research purposes is likely to impose a burden on practicing pediatricians, because research quality data demands structured data entry, which is often more time consuming. For this reason, this can be accomplished only with the input of the clinicians themselves, according to a few basic principles. First, we should focus (and perhaps limit) attention to areas of the EMR that present the biggest opportunities for generating new knowledge or improving care. Thus, we might choose not to devote efforts to portions of the EMR that deal with acute and self-limited clinical

problems (although these too offer interesting opportunities for study), but instead to aspects of the EMR that deal with high-prevalence chronic conditions or with preventive efforts that apply to large portions of the population. Second, we should create dialogues with clinicians about why certain kinds of information should or need to be entered consistently and completely, so that agreement can be reached as to why it is in clinicians' and patients' best interests to do so. Clinicians can be motivated to make changes that they view as beneficial. Third, we must engage clinicians in the design and testing of new structured formats for data entry. There should be no attempt to impose structure on EMR data that clinicians feel can only be effectively documented through narrative. In adhering to this principle, the new structured formats will work better, and clinicians will also feel that they "own" newly designed formats. Fourth, we must agree not to increase the amount of time it takes for pediatric clinicians to document care for the purposes of research, *unless the clinicians come to that conclusion themselves*. Increasing the work needed to take excellent care of patients is not in anyone's interest. Finally, we must report the results of the research derived from EMR data back to the clinicians as quickly as possible. Doing so will demonstrate the value of the partnership between clinicians, informaticians, and researchers that is necessary to reap maximum research benefit from EMR data.

Where can we turn to see working partnerships between clinicians, informaticians, and researchers? With their powerful blend of research expertise and informed clinician input, pediatric clinical research networks,⁵⁵ and especially practice-based research networks,⁵⁶ provide pediatricians with the best opportunity to collaborate with informaticians to improve the validity and reliability of clinical EMR data to make them more useful for research. An especially innovative group in this regard is the Pediatric Research Consortium (PeRC)⁵⁷ of the Children's Hospital of Philadelphia. Established in 2002, this group of researchers and clinicians from 30+ primary care practices benefits from using a single application of a single EMR product for clinical care. PeRC is able to bring practicing pediatricians together with epidemiologists, clinical informaticians, and administrators to, when necessary, shape the EMR in such a way as to facilitate clinical research projects. Such a process is iterative and must take place over time and with sufficient resources. Nevertheless, the PeRC example shows that working partnerships between clinicians, informaticians, and researchers can be realized successfully.

More recently, several national pediatric clinical research networks recently have begun projects aimed at improving EMR data extraction for comparative effectiveness research. Pediatric Research in Office Settings (PROS), the practice-based research network of the American Academy of Pediatrics, was awarded a cooperative agreement by the Health Resources and Services Administration Maternal and Child Health Bureau (HRSA-MCHB) to establish an EMR-based subnetwork among its member practices to conduct comparative effectiveness research on ADHD stimulant medication treatment.⁵⁸ The Pediatric Emergency Care Applied Research Network (PECARN) also has been awarded a cooperative agreement by HRSA-MCHB to use EMRs at several children's hospitals to conduct comparative effectiveness research on traumatic brain injury.⁵⁹ In the subspecialty world, the Improve Care Now inflammatory bowel disease network has received a grant from the Agency for Healthcare Research and Quality to build enhanced registries linked to hospital EMRs to conduct quality improvement and comparative effectiveness research.⁶⁰ In each of these examples, pediatric clinicians will work hand-in-hand with informaticians to make the clinical data entered in the EMR useful for research purposes. The successes achieved by research networks will provide a roadmap for future efforts.

In summary, although EMRs are primarily a tool for clinical documentation, the routinely collected data within them offer researchers a superb platform from which to enhance future pediatric clinical research. The clinical data provided by EMRs represent a substantial step

up from administrative claims, but for the reasons described above, present their own methodological challenges. The full potential of EMR data for pediatric clinical research will only be achieved when research becomes one of the explicit purposes for which pediatricians document patient encounters. Thus the technological advance of the EMR promises to put the pediatric clinician back at the center of twenty-first century pediatric clinical research.

Acknowledgments

The author thanks the dozen-plus individuals from multiple institutions who identified resources, answered questions, and critiqued drafts of the manuscript. This work was supported by Health Resources and Services Administration Maternal and Child Health Bureau awards #UA6MC15585 and UB5MC20286, National Eye Institute grant #R13EY019972, the American Academy of Pediatrics, and the Children's Hospital of Philadelphia.

References

1. Institute of Medicine (U.S.). Committee on Improving the Patient Record . The computer-based patient record: an essential technology for health care. Washington, D.C: National Academy Press; 1997. Revised ed
2. Reiser SJ. The Clinical Record in Medicine Part 1: Learning from Cases*. *Annals of Internal Medicine*. May 15; 1991 114(10):902–907. [PubMed: 2014953]
3. Wang SJ, Middleton B, Prosser LA, et al. A cost-benefit analysis of electronic medical records in primary care. *The American Journal of Medicine*. 2003; 114(5):397–403. [PubMed: 12714130]
4. Garets, D.; Davis, M. Electronic Medical Records vs. Electronic Health Records: Yes, There Is a Difference. 2006 [Accessed June 21, 2010]. www.himssanalytics.org/docs/wp_emr_ehr.pdf
5. McDonald CJ, Overhage JM, Tierney WM, et al. The Regenstrief Medical Record System: a quarter century experience. *International Journal of Medical Informatics*. 1999; 54:225–253. [PubMed: 10405881]
6. Burt C, Sisk J. Which physicians and practices are using electronic medical records? *Health Affairs*. 2005; 24:1334–1343. [PubMed: 16162581]
7. Kemper A, Uren R, Clark S. Adoption of electronic health records in primary care pediatric practices. *Pediatrics*. 2006; 118:e20–24. [PubMed: 16818534]
8. DesRoches C, Campbell E, Rao S, et al. Electronic health records in ambulatory care--a national survey of physicians. *NEJM*. 2008; 359:50–60. [PubMed: 18565855]
9. Centers for Medicare & Medicaid Services. Medicare and Medicaid Programs; Electronic Health Record Incentive Program. *Federal Register*. 2010:1844.
10. Blumenthal D. Stimulating the Adoption of Health Information Technology. *N Engl J Med*. April 9; 2009 360(15):1477–1479. [PubMed: 19321856]
11. Weiner M, Embi P. Toward reuse of clinical data for research and quality improvement: the end of the beginning? *Annals of Internal Medicine*. September 1.2009 151:359–360. [PubMed: 19638404]
12. Olsen, L.; Aisner, DJMM. Learning Healthcare System: Workshop Summary (IOM Roundtable on Evidence-Based Medicine). Paper presented at: IOM Roundtable on Evidence-Based Medicine; 2007.
13. Etheredge LM. A Rapid-Learning Health System. *Health Aff*. March 1; 2007 26(2):w107–118.
14. Wennberg JE, Roos N, Sola L, Schori A, Jaffe R. Use of Claims Data Systems to Evaluate Health Care Outcomes: Mortality and Reoperation Following Prostatectomy. *JAMA*. February 20; 1987 257(7):933–936. [PubMed: 3543419]
15. Quam L, Ellis L, Venus P, Jon C, Taylor C, Leatherman S. Using claims data for epidemiologic research: the concordance of claims-based criteria with the medical record and patient survey for identifying a hypertensive population. *Medical Care*. 1993; 31:498–507. [PubMed: 8501997]
16. Brook RH, McGlynn EA, Cleary PD. Measuring Quality of Care- Part Two of Six. *N Engl J Med*. September 26; 1996 335(13):966–970. [PubMed: 8782507]

17. Collen MF. Clinical research databases -- a historical review. *Journal of Medical Systems*. 1990; 14:323–344. [PubMed: 2132040]
18. Selby JV. Linking Automated Databases for Research in Managed Care Settings. *Annals of Internal Medicine*. October 15; 1997 127(8 Part 2):719–724. [PubMed: 9382386]
19. Menachemi N, Ettel D, Brooks R, Simpson L. Charting the use of electronic health records and other information technologies among child health providers. *BMC Pediatrics*. 2006; 6(1):21. [PubMed: 16869972]
20. Feinstein, AR. *Clinical Epidemiology: The Architecture of Clinical Research*. Philadelphia: W.B. Saunders Company; 1985.
21. *Philosophy Dictionary*. Definition of epistemology. <http://www.philosophy-dictionary.org/epistemology>
22. Chen, H.; Fuller, SS.; Friedman, C.; Hersh, W. *Medical Informatics: Knowledge Management and Data Mining in Biomedicine*. New York: 2005.
23. Shortliffe, E.; HeaC, JJ., editors. *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*. 3. New York: Springer-Verlag; 2006.
24. Lehmann, CU.; Kim, GR.; Johnson, KB., editors. *Pediatric Informatics: Computer Applications in Child Health*. New York: Springer Dordrecht; 2009.
25. Christakis DA, Johnston BD, Connell FA. Methodologic Issues in Pediatric Outcomes Research. *Ambulatory Pediatrics*. 2001; 1(1):59–62. [PubMed: 11888373]
26. Agency for Healthcare Research and Quality. Overview of the Kids' Inpatient Database (KID). 2011 [Accessed January 20, 2011]. <http://www.hcup-us.ahrq.gov/kidoverview.jsp>
27. Centers for Disease Control and Prevention. National Health Interview Survey. 2011 [Accessed January 20, 2011]. <http://www.cdc.gov/nchs/nhis.htm>
28. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey. 2011 [Accessed January 20, 2011]. <http://www.cdc.gov/nchs/nhanes.htm>
29. Centers for Disease Control and Prevention. Ambulatory Health Care Data. 2011 [Accessed January 20, 2011]. <http://www.cdc.gov/nchs/ahcd.htm>
30. National Center for Chronic Disease Prevention and Health Promotion. Youth Risk Behavior Surveillance System. 2011 [Accessed January 20, 2011]. <http://www.cdc.gov/HealthyYouth/yrbs/index.htm>
31. Agency for Healthcare Research and Quality. Medical Expenditure Panel Survey. 2011 [Accessed January 20, 2011]. <http://www.meps.ahrq.gov/mepsweb/>
32. de Lusignan S, van Weel C. The use of routinely collected computer data for research in primary care: opportunities and challenges. *Family Practice*. 2006; 23:253–263. [PubMed: 16368704]
33. Rethans JJ, Martin E, Metsemakers J. To what extent do clinical notes by general practitioners reflect actual medical performance? A study using simulated patients. *Br J Gen Pract*. 1994; 44:153–156. [PubMed: 8185988]
34. Jeff L, John WP, Timothy RD, Martin L, Peter G. How well does chart abstraction measure quality? A prospective comparison of standardized patients with the medical record. *The American Journal of Medicine*. 2000; 108(8):642–649. [PubMed: 10856412]
35. McDonald CJ, Overhage JM, Dexter P, Takesue BY, Dwyer DM. A Framework for Capturing Clinical Data Sets from Computerized Sources. *Annals of Internal Medicine*. October 15; 1997 127(8 Part 2):675–682. [PubMed: 9382379]
36. Murphy, S. Data warehousing for clinical research. In: Liu, LÖ.; Tamer, M., editors. *Encyclopedia of database systems*. New York: Springer; 2009. p. 679-684.
37. Lazarus R, Klompas M, Campion FX, et al. Electronic Support for Public Health: Validated Case Finding and Reporting for Notifiable Diseases Using Electronic Medical Data. *Journal of the American Medical Informatics Association*. 2009; 16(1):18–24. [PubMed: 18952940]
38. Nadler JJ, Downing GJ. Liberating Health Data for Clinical Research Applications. *Science Translational Medicine*. February 10; 2010 2(18):18cm16–18cm16.
39. AAP Council on Clinical Information Technology. EMR Review Project. [Accessed July 2, 2010]. 2010 <http://www.aapc.org/emr/readreviews.php>

40. Richard P, Robert D, Jonathan F, et al. Multicenter epidemiologic and health services research on therapeutics in the HMO Research Network Center for Education and Research on therapeutics. *Pharmacoepidemiology and Drug Safety*. 2001; 10(5):373–377. [PubMed: 11802579]
41. Pace WD, Cifuentes M, Valuck RJ, Staton EW, Brandt EC, West DR. An Electronic Practice-Based Network for Observational Comparative Effectiveness Research. *Annals of Internal Medicine*. September 1; 2009 151(5):338–340. [PubMed: 19638402]
42. Brown JSP, Holmes JHP, Shah KBA, Hall KM, Lazarus RMMPH, Platt RMDM. Distributed Health Data Networks: A Practical and Preferred Approach to Multi-Institutional Evaluations of Comparative Effectiveness, Safety, and Quality of Care. *Medical Care*. 48(6 Supplement 1):S45–S51. [PubMed: 20473204]
43. Ware H, Mullett CJ, Jagannathan V. Natural language processing framework to assess clinical conditions. *Journal of the American Medical Informatics Association*. Aug; 2009 16(4):585–589. [PubMed: 19390100]
44. Jagannathan V, Mullett C, Arbogast J, et al. Assessment of commercial NLP engines for medication information extraction from dictated clinical notes. *International Journal of Medical Informatics*. 2009; 78(4):284–291. [PubMed: 18838293]
45. i2b2 National Center for Biomedical Computing. i2b2 Software Version 1.5.1. 2010 [Accessed August 2, 2010]. <https://www.i2b2.org/software/index.html>
46. Hamlin BW, Biondich PGB, Wolfe BA, et al. Cooking Up An Open Source EMR For Developing Countries: OpenMRS – A Recipe For Successful Collaboration. *AMIA Annu Symp Proc*. 2006; 2006:529–533.
47. Weinberg ST. Web services can aid development of pediatric-friendly EHRs. *AAP News*. November 1.2009 30(11):27.
48. Lovis C, Baud R, Planché P. Power of expression in the electronic patient record: structured data or narrative text? *International Journal of Medical Informatics*. 2000; 58–59:101–110.
49. Turchin A, Shubina M, Breydo E, Pendergrass M, Einbinder J. Comparison of information content of structured and narrative text data sources on the example of medication intensification. *Journal of the American Medical Informatics Association*. Jun.2009 16:362–370. [PubMed: 19261947]
50. de Lusignan S, Metsemakers J, Houwink P, Gunnarsdottir V, JvdL. Routinely collected general practice data: goldmines for research? A report of the European Federation for Medical Informatics Primary Care Informatics Working Group (EFMI PCIWG) from MIE2006, Maastricht, The Netherlands. *Inform Prim Care*. 2006; 14:203–209. [PubMed: 17288707]
51. Berg M, Goorman E. The contextual nature of medical information. *International Journal of Medical Informatics*. 1999; 56(1–3):51–60. [PubMed: 10659934]
52. Winthereik BR. We fill in our working understanding”: on codes, classifications and the production of accurate data. *Methods Inf Med*. 2003; 42:489–496. [PubMed: 14534655]
53. van der Lei J. Use and abuse of computer-stored medical records. *Meth Inform Med*. 1991; 30:79–80. [PubMed: 1857252]
54. McMahon, B.; Pugh, T. *Epidemiology: Principles and Methods*. Boston: Little, Brown and Company; 1970.
55. Slora EJ, Harris DL, Bocian AB, Wasserman RC. Pediatric Clinical Research Networks: Current Status, Common Challenges, and Potential Solutions. *Pediatrics*. October 1; 2010 126(4):740–745. [PubMed: 20837583]
56. Lindbloom EJ, Ewigman BG, Hickner JM. Practice-Based Research Networks: The Laboratories of Primary Care Research. *Medical Care*. 2004; 42(4):III-45–III-49. [PubMed: 15026664]
57. Children’s Hospital of Philadelphia. Pediatric Research Consortium (PeRC). 2011 [Accessed January 20, 2011]. <http://www.research.chop.edu/cbmi/perc/percoverview.php>
58. Maternal and Child Health Bureau Research Program. Pediatric Primary Care Electronic Health Record (EHR) Network for Comparative Effectiveness Research (CER). 2010 [Accessed January 12, 2011]. http://mchb.hrsa.gov/research/project_info.asp?ID=166
59. Pediatric Emergency Care Applied Research Network. PECARN Current Research. 2011 [Accessed January 28, 2011]. <http://www.pecarn.org/currentResearch/index.html>

60. ImproveCareNow. Enhanced Registries Grant: Building Modular Chronic Disease Registries for QI and Comparative Effectiveness Research. 2010 [Accessed January 12, 2011]. <http://www.improvecarenow.org/news/13-latest/77-enhanced-registries-grant->

Table

Biomedical informatics resources Textbooks

Textbooks
Chen H, Fuller SS, Friedman C, Hersh W. <i>Medical Informatics: Knowledge Management and Data Mining in Biomedicine</i> . New York 2005
Shortliffe EHeaC, J.J. (assoc. ed.). <i>Biomedical Informatics: Computer Applications in Health Care and Biomedicine</i> . 3rd Edition ed. New York: Springer-Verlag; 2006
Lehmann CU, Kim GR, Johnson KB, eds. <i>Pediatric Informatics: Computer Applications in Child Health</i> . New York: Springer Dordrecht; 2009
Journals
Journal of the American Medical Informatics Association, http://jamia.bmj.com/
Journal of Biomedical Informatics, http://www.elsevier.com/wps/find/journaldescription.cws_home/622857/description#description
BMC Medical Informatics and Decision Making, http://www.biomedcentral.com/bmcmedinformdecismak/
Informatics in Primary Care, http://www.radcliffeoxford.com/journals/J12_Informatics_in_Primary_Care/default.htm
International Journal of Medical Informatics, http://www.elsevier.com/wps/find/journaldescription.cws_home/506040/description#description
Methods of Information in Medicine, http://www.schattauer.de/en/magazine/subjectareas/journals-a-z/methods.html
AMIA Annual Symposium Proceedings, https://www.amia.org/mbrcenter/pubs/proceedings.asp