At its core, the practice of medicine is an information-intensive endeavor. Most of what physicians involve the collection, review, and management of information. Examples of such activities include obtaining and recording patient information, consulting colleagues, reading the scientific literature, planning diagnostic procedures, devising strategies for patient care, interpreting tests, and conducting research. The ever-increasing biomedical knowledge base that must be considered to deliver optimal patient care only adds to the challenges facing medicine today.

Successfully addressing these challenges to deliver the best health care possible requires not only the existence of valid and generalizable data sets derived from systematic basic, clinical, and epidemiological research efforts but also the ability to apply the knowledge derived from these research efforts at the point of care. It is easy to understand, therefore, why the field of biomedical informatics, a field that is concerned with collecting, managing, and optimally using information in health care and biomedicine, is critical to the current and future practice of medicine and the study of healthcare outcomes that result from such practice.1,2

Biomedical informatics approaches and related health information technology (health IT) platforms are key to enabling knowledge-driven healthcare and practice improvement initiatives based on a solid research foundation. Similar biomedical informatics approaches and resources are also critical to advancing outcomes research. Indeed, such technologies such as electronic health records (EHRs), clinical data repositories, and research-specific data management systems are already transforming the way we practice medicine and conduct research. This transformation is being further advanced by federally directed funding and research infrastructure development efforts.3,4

In the sections that follow, we provide an overview of how biomedical informatics and health IT processes and tools can affect the conduct of research and the delivery of evidence-based health care from our perspective. Given the current state of development in this area, we introduce a conceptual framework by which clinicians, researchers, and the healthcare community at large can optimally leverage biomedical informatics processes and tools to facilitate outcomes research and drive improvements in healthcare outcomes. We conclude by proposing a course of action intended to support greater collaboration between the clinical, research, and biomedical informatics communities to achieve the goals set forth.

**A Framework for Knowledge-Driven Outcomes Research**

**Outcomes Research as a Cyclical Activity**

Outcomes research “seeks to understand the end results of particular health care practices and interventions” on individual patients and populations.5 This often includes the evaluation of economic impacts linked to health outcomes such as cost-effectiveness, cost utility, and comparative effectiveness. Such research involves a range of data collection and aggregation methods, including drawing on primary studies and collecting data de novo for such research.6 Therefore, the conduct of outcomes research can be thought of as involving an information cycle that includes collecting data about healthcare practices and patient/population-level outcomes, analyzing those data, and reporting on the findings.

When considering the types of information relevant to the conduct of outcomes research and the practice of evidence-based medicine, we can identify 2 major types as relevant: patient-specific information, or information generated during the care of patients such as numerical values, free text, imaging information, etc, that might be found in an EHR, and knowledge-based information, or information derived from the scientific medical literature that is based on biomedical and healthcare research. Because the collection, storage, retrieval, and optimal use of these types of information constitute the material focus of biomedical informatics, it becomes clear why the intersection of biomedical informatics and outcomes research is critical to the advancement of medical science and practice.

Indeed, the efficient conduct of outcomes research requires access to robust clinical and population-level data sets, as well as existing research and knowledge data sets, to evaluate relevant metrics such as procedural complications, days of hospitalization, health status, and mortality. Moreover, the results of research can be used to develop system-level interventions designed to facilitate the improvement and/or...
optimization of healthcare policies and practice, often by leveraging clinical informatics platforms such as EHRs and clinical decision support systems.

### Biomedical Informatics Systems and Their Roles in Research

Biomedical informatics can be seen as a key enabler of the outcomes research cycle. One way to explore the role of biomedical informatics is in terms of the resources and platforms developed and used for contemporary clinical practice and research practice. Examples of such systems include EHR platforms, clinical data warehouses, and research information management platforms, all of which are becoming increasingly available in the healthcare environment.

At each stage in the research process, general-purpose or research-specific IT systems may be of utility. Payne et al described a model for the applicability of IT systems to steps in the clinical research process. Given that outcomes research is a subtype of clinical research, the same model can be applied to understand the role of IT systems for the conduct of outcomes research. Examples of general and research-specific information systems that are able to support the conduct of clinical research include the following:

- Literature search tools such as PubMed can be used to conduct background research necessary for hypothesis development and study preparation.
- EHRs can be used to collect clinical data on research participants in a structured form that can reduce redundant data entry and identify patients who are eligible for interventions.
- Data mining tools can be used to identify particular cohorts of potential subjects for studies or conduct retrospective analyses from existing databases.
- Decision support systems can be used to alert providers at the point of care that an individual may be eligible for a clinical trial.
- Computerized physician order entry systems, which collect data describing the therapies delivered to research participants, can be used in both participant tracking and study analyses.

In addition to the preceding general-purpose and clinical IT systems, research-specific IT systems have been developed that include the following:

- Simulation and visualization tools can streamline the preclinical research process (eg, disease models) and assist in the analysis of complex data sets.
- Protocol authoring tools can allow geographically distributed authors to collaborate on complex protocol documents.
- Participant screening tools can assist in the identification and registration of research participants.
- Research-specific Web portals provide researchers with a single point of access to research-specific documents and information for collaboration.
- Electronic data collection or capture tools can be used to collect research-specific data in a structured form and to reduce the need for redundant and potentially error-prone paper-based data collection techniques.
- Research-specific decision support systems provide study-specific guidance to researcher for tracking, for example, the participants’ status and protocol compliance.

Numerous reports have concluded that the use of such tools and platforms can lead to increased data and research quality. Furthermore, the use of informatics platforms across multisite studies has been shown to increase the efficiency and efficacy of such team-science endeavors. The ability to use IT in support of clinical research relies on the ability to collect, store, and analyze data. Examples of information systems that are becoming increasingly common include electronic data capture systems and integrated clinical trials management systems, which target multiple aspects of the clinical trials process, including electronic data capture, financial management, data quality assurance, and research participant tracking. Indeed, the use of such IT for clinical research is growing rapidly, and it is projected that nearly 50% of all studies will soon use information management technologies.

### Informatics Methods for Data Exchange, Integration, and Use

A common problem for the outcomes researcher is the aggregation of data from varied and disparate information resources. Therefore, resources that bring together clinical data from otherwise inaccessible and nonintegrated medical record, laboratory, hospital-based, and other healthcare enterprise systems throughout a given region to facilitate more efficient and effective care processes are becoming an important part of the research IT solution set.

The growing availability of such integrative data sets via the development of regional health information exchange organizations is an area of particular relevance to this discussion. Although health information exchange organizations are in various stages of development across the country and some have struggled to establish themselves, there are early examples of successful health information exchange models that have enabled the use of population-level data for public benefit. Examples include the use of health information exchange organizations to significantly improve reporting of common diseases over standard means, to allow surveillance for disease outbreaks, and to facilitate population-level epidemiological studies that would otherwise have been difficult, if not impossible.

Fundamental to the success and utility of such efforts, whether at the regional or even at the individual institutional level where multiple data sources are common, are the informatics underpinnings that enable the aggregation and integration of data from disparate sources into data repositories. One key element of such informatics methods underpinning the exchange of data are transactional standards. A commonly used example is Health Level 7 version 2, a health data messaging interchange standard used to transfer information between systems that resources like health information exchange organizations often use to exchange information between source and destination systems. It is worth noting, however, that because such standards traditionally define the mechanism of exchanging data but do not define
the semantic annotations necessary to ensure a shared understanding of the meaning of such data across platforms or organizations, their adoption alone is not sufficient to enable the seamless exchange of complex data sets. Work is ongoing to address this issue in new versions of such standards such as Health Level 7 version 3.44

Another key element is the methodologies and approaches used to maintain integrative data repositories of data in resources such as data warehouses. A data warehouse is a type of database or data repository that is designed to have certain characteristics that enable its use for purposes such as research including the following definitional factors:45–47:

- Subject oriented: Data elements being collected and managed via the data warehouse correspond to real-world entities and often have a set of hierarchical and/or semantic interrelationships.
- Time variant: Because data and data sources change over time, the natural history of such information in the data warehouse is stored and can be retrieved by end users.
- Nonvolatile: No data are deleted or expunged from the data warehouse, allowing it to serve as an authoritative, longitudinal repository of targeted data types.
- Integrated: Data stored in the data warehouse are consistent in scope and comprehensiveness, and appropriate linkages between data sets that have hierarchical, semantic, geographic, or temporal interdependencies are maintained regardless of the specific data modeling or management approach used in the implementation of the warehouse.

There are many potential uses for a data warehouse,48,49 but it is the ability to perform longitudinal or episodic queries based on ≥1 criteria of interest in support of research activities (eg, study planning, retrospective data analyses) and the delivery of task- or role-specific data marts to support context-specific access to data sets by other applications or direct query and analysis by authorized end-users that make these resources so useful in the biomedical research domain.

Indeed, within the biomedical domain, numerous reports provide contexts in which data warehouses have been used:

- The retrieval of patient cohorts for either clinical trial feasibility analyses or active participant recruitment.50
- The application of data mining and statistical analysis tools to large-scale data extracts to identify or test hypotheses relative to relationships between demographic, phenotypic, and biomolecular parameters.51
- The identification of trends or phenomena surrounding events of interest such as infections, adverse events, or complications associated with clinical interventions, as well as evidence-based guideline compliance.53
- The support of business intelligence applications that enable operations research or optimization, including the provision of real-time performance indicators and dashboards.45,54
- The execution of complex, integrative reports for oversight, regulatory, and financial monitoring purposes.45,54

Along with all of these benefits, there are also challenges associated with the use of data warehouses, including

- overcoming regulatory and data privacy and confidentiality concerns; ensuring the provenance and quality of data being included in the data warehouse; implementing and supporting sufficiently robust and timely interfaces between production systems and data warehouse–specific extraction, transformation, and load processes; and providing timely access to limited or deidentified data sets for retrospective research or research planning purposes. However, despite such potential limitations and challenges, the use of data warehouse platforms in biomedical settings has been and continues to be associated with major increases in productivity and efficiency surrounding many application scenarios.

Another factor motivating the development of resources that enable connectivity and information exchange within and between major research centers is the increased National Institutes of Health funding in recent years for research informatics infrastructure. Examples include the Cancer Biomedical Informatics Grid, Informatics for Integrating Biology and the Bedside, and Clinical and Translational Science Award initiatives.40,55,56 These efforts have led to the development of additional resources and approaches for advancing information integration and accessibility to advance all kinds of research.

As mentioned previously, patient- or population-level data are a key type of information needed to advance evidence-based medicine. The other is knowledge-based information such as that contained in the scientific literature. Here, too, informatics resources such as Medline/PubMed, the Cochrane Database, and other data repositories of medical knowledge, including guideline repositories, have become critical components of the evidence-driven medicine solution. In addition to relying on manual review of such resources on an ad hoc basis in the course of hypothesis generation or to answer clinical questions, these repositories have the potential to influence care directly through biomedical informatics approaches.

One such example involves the combination of knowledge-based information with patient-level data to drive healthcare support via rules engines, often as part of EHRs. Such clinical decision support systems assess individual rules and determine their applicability in a specific case or situation.57 At the most basic level, such a rules engine comprises 2 components, a knowledge base, which consists of rules represented in a computational format, and an inference or execution engine, which can reason about incoming data on the basis of the contents of the knowledge base to generate some form of output such as an instruction set or alert. An additional critical component of a rules engine is the knowledge engineering facility, which provides the capability to curate the contents of the knowledge base on an automated, semiautomated, or manual basis.

Biomedical informaticians have contributed various models for representing clinical decision support rules throughout the years. Examples such as the Arden syntax, which can be used to write what are called medical logic modules, form the basis of clinical alerting systems.58 Work has also been done on frameworks and systems to represent the knowledge of clinical guidelines. Examples such as the guideline interchange format, EON, and its successor, SAGE, allow computation on guidelines for the purposes of point-of-care
decision support and have been used in multiple studies.\textsuperscript{59,60} Indeed, products of the EON project have been deployed in a production-level advisory system for the management of hypertensive patients in Veterans Affairs hospitals called ATHENA.\textsuperscript{61}

**Developing an Integrative Model for Informatics-Supported Outcomes Research**

As the preceding descriptions make evident, the biomedical informatics community has generated a broad variety of techniques, platforms, and theoretical models capable of supporting the myriad information management and analysis activities essential to the conduct of outcomes research and ultimately the practice of evidence-based medicine. However, such physical and methodological resources are often underused, usually because of a combination of human factors, sociotechnical issues, and technology-based barriers.\textsuperscript{3,4} In analogous research domains such as clinical research and biomolecular translational science, the development of conceptual frameworks that illustrate the interrelationships of critical components of the sociotechnical system has been shown to help in advancing the development of informatics solutions to similar issues.\textsuperscript{19,62} Therefore, we propose the basis for such a model to advance the integration of biomedical informatics and outcomes research. The proposed model builds on the Agency for Healthcare Research and Quality description of outcomes research, which states, “Outcomes research seeks to understand the end results of particular health care practices and interventions . . . . For clinicians and patients, outcomes research provides evidence about benefits, risks, and results of treatments so they can make more informed decisions.”\textsuperscript{9}

Building on this description, we first acknowledge the relevant methodological approaches and operational or research products:

1. Healthcare policies and practices that serve to influence behaviors and clinical outcomes.
2. Direct and surrogate measurements of the preceding behaviors and outcomes at multiple levels of dimensionality and granularity, including patient data generated during routine clinical care, population-based data sets such as those associated with public health interventions and studies, and research data generated during targeted clinical studies.
3. Synthesized results of analytical operations applied to the preceding data sets to elucidate the benefits, risks, and results of the policies and practices in item 1.
4. Informatics platforms and components such as clinical decision support systems, guideline delivery systems and evidence dissemination applications (eg, literature databases, guideline repositories), all of which are informed by or leverage the synthesized results associated with item 3.

Spanning the preceding 4 methodological and operational/research products is a set of enabling biomedical informatics practice areas that are concerned with the design, application, and evaluation of tools and techniques for the following:

1. Data capture and storage, as exemplified by EHRs, personal health records, data warehouses, and clinical trial management systems, which collectively allow for the population of data sets based on the outcomes/impact of healthcare practices and policies.
2. Data integration and exchange, as exemplified by mechanisms used to aggregate disparate and heterogeneous data sets that are generated via data capture and storage platforms to enable a full spectrum of hypothesis discovery and testing activities.
3. Knowledge representation and reasoning, as exemplified by inferencing applications used to deliver appropriate clinical decision support or practice guidelines on the basis of the best available evidence and patient- or context-specific variables.

Taken together, the combination of the 4 methodological and research/operational products and the corresponding and enabling informatics practice areas creates a framework to facilitate and support the improvement or optimization of healthcare practice and policies on the basis of the systems-level understanding of outcomes, risks, and benefits that they afford researchers, practitioners, and policy makers (the Figure). It is important to note that this model includes an iterative feedback cycle to ensure that experiences with interventions inform the creation and optimization of healthcare practices and policies, including those related to the development of clinical and research informatics solutions.

**Proposed Courses of Action**

As reviewed in the preceding sections, the conduct of outcomes research and the corresponding provision of evidence-based medicine constitute a complex and information-intensive endeavor. Central to such activities is a systems-level approach to information and knowledge generation, collection, analysis, and dissemination. We believe that the development of a community-accepted conceptual model that advances a collective understanding of the complex interplay between the constituent operational, research, and informatics components is both necessary and advantageous to those working in these domains. In the preceding section, we outlined the basis for such a model in the hopes of catalyzing the further discussion and research required to formalize such a construct. However, to realize the benefit of such a model and the processes by which it will be fully conceived and validated, a number of important factors must be addressed by our clinical and research communities:

- Efforts must be undertaken at the local, state, and national levels to promote and support the development of formal and informal collaboratories that span the clinical, outcomes research, and biomedical informatics communities. Currently, the formation of such collaborative research alliances is limited by insufficient or inadequately distributed resources, policy-based barriers to the conduct of team-science activities (including tenure and promotion criteria in academic settings that often do not reward interdisciplinary science), and the absence of appropriate venues for the training and career development of multidisciplinary outcomes researchers who span the preceding practice and scholarly domains.
Significant rationalization of regulatory frameworks is needed to enable the secondary use of clinical data in an efficient, timely, and secure manner while still ensuring patient privacy and confidentiality. The current regulatory environment is rife with often contradictory and uncoordinated policies and regulations relative to the use of patient data, even in a deidentified format, for secondary research purposes. Such a confusing environment makes it difficult, if not impossible, to conduct large-scale outcomes research programs without expensive and resource-intensive prospective consent processes, even when the use of existing data for such research activities in a retrospective manner is exceptionally low risk in terms of patient privacy and confidentiality. Furthermore, these same regulatory frameworks also contribute to the complexity surrounding contemporary time- and personnel-intensive Institutional Review Board processes at most if not all academic health centers and can significantly delay research conduct.

Greater integration between clinical care and clinical research should be fostered. Currently, organizational and regulatory policies and procedures situate the conduct of clinical research so that it is usually distinct from conventional clinical care. However, in many settings, clinical care and research are often interrelated, even in the traditional sense. When one begins to view clinical care and the related collection and use of healthcare data as part of the research cycle described above, one begins to recognize the value of considering the benefits of integrating research and clinical care in the same environment. Indeed, the differentiation and formal decoupling of these activities only serve to complicate workflow, impede efficient information exchange, and reduce patient access to cutting-edge treatment modalities. Ultimately, to realize the benefits of outcomes research, especially in the modern informatics era, we must embrace an approach to clinical medicine in which all patients are given the opportunity to be involved in some aspect of the research cycle. Without such a model, we will be hard pressed to realize the powerful benefits of our informatics techniques and platforms and our ability to generate the evidence that outcomes research is intended to enable.

Summary

The conduct of outcomes research is an information-intensive endeavor and therefore benefits from the application of biomedical informatics approaches, resources, and platforms. It is our contention that a tighter integration of biomedical informatics, clinical care, healthcare policy, and outcomes research can advance improvements in healthcare research and practice. As the preceding overview of the current state of knowledge in these domains illustrates, such integration will require at the most basic level the development of team-science approaches to outcomes research that include clinicians, researchers, and biomedical informaticians. However, the formation and support of such teams will not be possible without addressing some of the barriers and requirements summarized here. Building on the recent and ongoing advances in biomedical informatics and addressing the issues raised above create a tremendous opportunity to link research, healthcare delivery, and policy in such a way as to have a direct and demonstrable impact on the health and quality of life of the public.
Sources of Funding
Dr Embi’s contributions were supported in part by grants from the National Institutes of Health/National Library of Medicine (R01-LM009533) and National Institutes of Health/National Center for Research Resources (UL1-RR026314). Dr Payne’s contributions were supported in part by grants from the National Institutes of Health/National Cancer Institute (P01-CA081534, R01-CA134232) and National Institutes of Health/National Center for Research Resources (U54-RR02384).

Disclosures
None.

References
48. Lawrence D. All hype? Have business intelligence tools actually changed, or have they just been repackaged and renamed? Healthc Inform. 2007; 24:16, 18, 20.


Biomedical Informatics and Outcomes Research: Enabling Knowledge-Driven Health Care
Peter J. Embi, Stanley E. Kaufman and Philip R.O. Payne

*Circulation.* 2009;120:2393-2399
doi: 10.1161/CIRCULATIONAHA.108.795526
*Circulation* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2009 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/120/23/2393

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Circulation* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to *Circulation* is online at:
http://circ.ahajournals.org//subscriptions/