Future Directions in Evaluation Research: People, Organizational, and Social Issues

B. Kaplan¹, N. T. Shaw²

¹Yale University School of Medicine, New Haven, CT, USA
²THINK (The Health Informatics Network), Centre for Healthcare Innovation and Improvement, BC Research Institute for Children’s and Women’s Health, Vancouver, BC, Canada

Summary
Objective: To review evaluation literature concerning people, organizational, and social issues and provide recommendations for future research.
Method: Analyze this research and make recommendations.
Results and Conclusions: Evaluation research is key in identifying how people, organizational, and social issues - all crucial to system design, development, implementation, and use - interplay with informatics projects. Building on a long history of contributions and using a variety of methods, researchers continue developing evaluation theories and methods while producing significant interesting studies. We recommend that future research:

1) Address concerns of the many individuals involved in or affected by informatics applications.
2) Conduct studies in different type and size sites, and with different scopes of systems and different groups of users. Do multi-site or multi-system comparative studies.
3) Incorporate evaluation into all phases of a project.
4) Study failures, partial successes, and changes in project definition or outcome.
5) Employ evaluation approaches that take account of the shifting nature of health care and project environments, and do formative evaluations.
6) Incorporate people, social, organizational, and cultural, and concomitant ethical issues into the mainstream of medical informatics.
7) Diversify research approaches and continue to develop new approaches.
8) Conduct investigations at different levels of analysis.
9) Integrate findings from different applications and contextual settings, different areas of health care, studies in other disciplines, and also work that is not published in traditional research outlets.
10) Develop and test theory to inform both further evaluation research and informatics practice.

Keywords
Evaluation; technology assessment, medical informatics, telemedicine, organizational culture; attitudes towards computers; implementation; barriers; people, organizational, social, issues; sociotechnical; ethical issues; qualitative methods; ethnographic methods; multi-method; human-computer interaction.

With increased pressures for implementing information and communication technologies (ICT), spending on information technology in health care is expected to rise [1]. Introducing these technologies involves change, in at least three respects. Changes occur due to the way introducing new systems is handled. They also occur because of concomitant changes in communication flows and work patterns [2]. Changes occur, too, because of organizational repercussions on the way a new system functions and is used [3]. This complex interplay of changes involves a variety of people, organizational, and social issues, including human-computer interaction, sociotechnical, cultural, and ethical concerns.

Evaluation [4-7] can serve multiple purposes during these change processes. Evaluation can help in identifying where change may need fine tuning or major adjustment, thereby preventing harm and minimizing disruption. It can help in providing evidence for decision-making and extending knowledge [8]. Evaluation can explore how changes at different levels of analysis interrelate, so that changes in work roles or communication patterns between clinicians, for example, may affect how a hospital functions. In these ways, evaluation can enhance decision making for managing change in itself, or even better, for managing change as a means for strategic organizational development [5-7]. Evaluation and change management [9-11], then, address similar concerns and involve related theories [5, 12-14].

Evaluation researchers employ numerous approaches, including: randomized controlled trials, experimental designs, simulation, usability testing, cognitive studies, record and playback techniques, network analysis, ethnography, economic and organizational impacts, action research, content analysis, data mining, studies based on actor-network or structuration theory, ethnomethodology, balanced score cards, soft systems and participatory design methodologies, surveys, qualitative methods of data collection and interpretive analyses of it, technology assessment, benchmarking, SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis, and social interactionism [5-7, 14-23].

Many evaluations focus on impacts. People, organizational, and social issues are central to understanding those impacts, and hence, to evaluation. The UK’s National Health Service recognized this in advising that evaluations include business, user (i.e., organizational and patient), and technical impact [15]. In this paper, we review streams in evaluation research (sometimes called “assessment”) as related to people, organizational, and social issues, and we suggest directions for further research.

Foundation Studies
Early evaluation studies of medical information applications are represented in [4]. These early papers, many by researchers who still are active in the field, report insights that remain relevant and serve as a foundation on which an expanding body of knowledge has been based. One, an evaluation of the PROMIS system at the University of Vermont [24, 25], was a multi-method study by external evaluators. This early study is exemplary both methodologically and for the people, organizational, and social issues insights it produced. The study identified how the implementation of a new medical record and clinical guidance...
system was related to issues concerning professional roles and status; change management; user involvement; organizational communication; and relationships between the structure and organization of the medical record, philosophy of health care delivery, and clinical work—all still important in electronic record and clinical decision support systems. Other early evaluations focusing on these kinds of issues include a series of studies at what was then called the Rockland Research Institute in New York (e.g., [26, 27]), and another series at Methodist Hospital in Indiana (e.g., [28, 29]).

The new hospital information system at El Camino Hospital in California also was extensively evaluated by independent researchers during the 1970s [30]. Originally developed by Lockheed Missiles and Space, this system became the popular Technicon, then TDS, and, later, the Eclypsis system.

During the 1970s, analyses also began to appear of lessons learned and prescriptions for success [31]. Management issues, user acceptance, and diffusion and adoption of information systems have been discussed in the medical informatics literature at least since the early 1980s [32]. Authors linked diffusion and communication studies, evaluation research, and change management since early in the development of the field (e.g., [33–40]). For example, studies of the diffusion of a hospital information system showed that physicians’ professional networks influenced adoption and could be used to encourage system use [28, 29]. Others analyzed medical informatics applications according to innovation characteristics known to affect adoption [36, 41].

**Current Evaluation Research**

In a recent literature review, Kaplan summarized people, organizational, and social issues identified in evaluation, and discussed the fit of information and communication technologies with various aspects related to these concerns [42]. These include how information and communication technologies fit other contextual issues surrounding their development, implementation, and use. Researchers have addressed work flow and routines [14, 43–52]; clinicians’ level of expertise [48]; values and professional norms [53–55]; institutional setting, history, and structure [37, 50, 56]; communication patterns [37–39, 52, 55]; organizational culture, status relationships, control relationships, division of labor, work roles, and professional responsibility [25, 49, 56, 57]; cognitive processes [58]; congruence with existing organizational business models and strategic partners [59]; compatibility with clinical-patient encounter and consultation patterns [50, 52, 60]; and the extent to which models embodied in a system are shared by its users [3, 14, 55, 61, 62]. Authors have also addressed (in various combinations) fit between information technology and how individuals define their work, user characteristics and preferences (e.g., information needs), the clinical operating model under which a system is used, and the organization into which it is introduced [37, 43, 55, 63–68]. Others have focused on interrelationships among key components of an organization, such as organizational structure, strategy, management, people’s skills, and technology [69]; and compatibility of goals, professional values, needs, and cultures of different groups within an organization, including developers, clinicians, administrators, and patients [69–76]. Some have discussed transferring to another country a system designed for use under a different country’s health care system [50, 53, 77, 78]. Some have investigated incorporating user requirements and involvement [79, 80] and patient preferences [81]. In addition, studies have been done on ways in which informatics applications embody values, norms, representations of work and work routines; assumptions about usability, information content and style of presentation, and links between medical knowledge and clinical practice; and how these assumptions influence system design [17, 50, 52, 55, 62, 73, 82–86].

The significance of these research contributions related to people, organizational, and social issues in medical informatics is evidenced by the International Medical Informatics Association (IMIA), American Medical Informatics Association (AMIA), European Federation for Medical Informatics (EFMI), and European Science Foundation all supporting working groups in this area. These working groups have been building on years of activities and research by organizing conference sessions and publications. However, researchers recognize that these concerns have become more challenging because technological and institutional changes in health care contribute to making complex organizational, social, and personal arrangements even more complex. A recent White Paper reflected this when proposing a research agenda for key people and organizational issues [87].

**Barriers to Evaluation**

Evaluation is inherently political: what happens when a new technology is introduced is both affected by organizational and implementation processes, as well as affecting them. Evaluation, too, is political in nature because it concerns needs, values, and interests of different stakeholders [88]. Further, evaluation is increasingly important because of its role in government regulation and policy concerning technologies in health care [89, 90]. It may provide the political rationale for adoption or discontinuation of projects on apparently disinterested scientific grounds [90].

Some, therefore, resist evaluation for fear of potential disruptiveness of the investigation or its findings [8]. Some fear the blame that may fall on those who point out difficulties [91]. Others question what can and cannot be evaluated [92]. Some recog-
nize that evaluation methods must fit managerial and scientific preconceptions of what constitutes worthwhile methods and measures of success [93]. Still others may view evaluation results as site specific. There are many barriers to evaluation, including methodological complexity of the undertaking, motivation, and ethical considerations [94, 95]. For all these reasons, some discount either conducting evaluation studies or accepting their results [96, 97].

Evaluation Frameworks

A variety of efforts has been undertaken to address these concerns while providing a structure for evaluations. For example, The National Health Service in the UK issued an evaluation framework with the intent to improve the quality of evaluations [15]. In the European Union, the first phase of ATIM, the Accompanying Measure on Assessment of Information Technologies In Medicine, was undertaken as an Accompanying Measure in the Programme for Telematics Systems in Areas of General Interest (DG XIII) in the area of AIM (Advanced Informatics in Medicine). Its goal was to develop consensus on both methods and criteria for assessment [96]. The IMIA Technology Assessment and Quality Development in Health Informatics Working Group (WG15) is directly concerned with these issues [8], as are the aforementioned sister working groups.

In addition, a number of authors have suggested frameworks for conducting evaluation studies. Green and Moehr [98] reviewed six major Canadian frameworks alone, noting that documented empirical or theoretical support is lacking. Others draw on a variety of theories, suggest different methods, and address a variety of concerns. For example:

- Grant, Plante, and LeBanc [102] argue for standardization in evaluation approach in order to avoid partial evaluation approaches and hence avoid remoteness from actual use. They propose their TEAM (Total Evaluation and Acceptance Methodology) methodology, based on the three dimensions of role, time, and structure, as a basis for future standardization.
- Kaplan’s 4Cs framework [12, 13, 44, 103] – focusing on communication, control, care, and context – and her set of evaluation guidelines call for flexible multi-method longitudinal designs of formative and summative evaluations that incorporate a variety of concerns. She elsewhere [12] suggests a matrix with one dimension, stages of patient treatment [104], and as a second dimension, beneficiaries of the information resource [105]. Together with other colleagues, Kaplan et al. [87] propose a research agenda of key issues based on a matrix crossing organizational level of analysis with different social science disciplines.
- Kazanjian and Green [106] propose a “Comprehensive Health Technology Assessment” framework.
- Lauer, Joshi, and Browdy [107] illustrate how an equity implementation model can apply to evaluating user satisfaction with a patient scheduling system.
- Lehoux and Blume [89] suggest a sociopolitical technology assessment framework that addresses four sets of issues: 1) the potential actors, 2) implications for flow of materials and resources, 3) knowledge production and circulation, and 4) power relations.
- Shaw [110] identifies six aspects in her CHEATS framework for a comprehensive evaluation strategy: clinical, human and organizational, educational, administrative, technical, and social.
- Balanced score cards also have served as the basis for evaluation. Protti [23] draws on Kaplan and Norton’s Balanced Scorecard [111] and recommends using them to measure and evaluate information management and technology’s efficiency, effectiveness, transformative potential, and impacts from multiple perspectives and in assorted ways. Both Gordon [112] and Niss [113] also each advocate the use of a balanced score card.
- A multidisciplinary team evaluating electronic records at multiple sites in the UK [114] developed several frameworks to help guide their study. One of these frameworks, SAFE, was built upon that of Boloix and Robillard’s framework [115] for evaluating software.
- Others, including an evaluation textbook [7] and the UK’s National Health Service PROBE project [15, 114, 116], use, elaborate, or extend Donabedian’s well-known structure-process-outcome evaluation model [117, 118]. One such elaboration [119] combines it with the DeLone-McLean model of information system success [120]. Another [121, 122] combines the DeLone-McLean model with Fineberg’s model for efficacy of clinical diagnostic technology [123]. Others advocate using the DeLone-McLean model to study both system successes and failures [124].

Earlier authors categorized evaluation methodologies, suggesting which methods are appropriate for which evaluation questions (e.g. [125-128]). The IMIA Technology Assessment and Quality Development in Health Informatics Working Group (WG15) has been creating a framework for assessing the validity of a study [129]. Recently, there has been interest in examining evaluation itself, with some researchers studying the evaluation process and how to evaluate evaluations [90, 130-133].

There is disagreement over the usefulness of frameworks per se [114, 134]. While some argue the need for a unifying evaluation model [79, 102], we prefer a diversity in approach, even though we have developed frameworks ourselves. We believe that frameworks need to be applied with care. The power of any model is that it simplifies, necessarily omitting some things while highlighting others. No framework is likely to be comprehensive, appropriate to all projects, nor even to all stages of a project.
Evaluation Theory

Jones classifies evaluation approaches into four models: randomized controlled trials, scientific/quantitative/objectivist, project management, and qualitative/interpretive/subjectivist [130], while Friedman and Wyatt pose a basic dichotomous distinction between objectivist and subjectivist [7]. Fulop similarly divides approaches into positivism, interactionism and postmodernism, but also into deductive and inductive approaches [135]. Evaluation approaches also may be classified as a) undertaken either from an organizational or sociotechnical viewpoint [23]; b) being driven by different theories of change or resistance [13, 127]; or c) being grounded in other theories or different social science methods.

Different approaches and different evaluation questions are preferred by different groups (e.g., developers, managers, physicians, nurses, and secretaries) within a health care setting and according to whether individuals have a health care background [122, 136]. All these different preferences, theories, and methods have been subject to considerable debate about the appropriateness of which methods evaluation researchers should use. Most evaluations are based on positivist, rationalist, or rational choice theoretical perspectives and study designs [42]. Technology assessment, too, similarly is dominated by quantitative experimental designs (based on randomized controlled trials) and cost/benefit analysis. The focus is on cost, safety, and efficacy [89]. Both evaluation and technology assessment, therefore, often have not addressed social, ethical, political, organizational, and cultural considerations.

This orientation, too, produces atheoretical evaluations [137]. Recently there have been calls to move away from an emphasis on methods and toward more theory-based evaluation research, including calls for studies informed by social science theories [42, 87, 138]. The social sciences provide the basis for other approaches to evaluation and technology assessment that do address social, ethical, political, organizational, and cultural concerns.

Social and Behavioral Sciences

Evaluations informed by theoretical work in organizational theory, communications, and the social and behavioral sciences have been undertaken for some time [4, 40, 42, 139]. Lorenzi gives an overview of organizational theory influences [10, 140], and Kaplan of social science ones [42, 87]. Earlier examples of work based in the social sciences and communications include works cited in [31, 87] as well as [4, 5, 24, 25, 141-151].

Although there is a tradition of evaluations based on social and behavioral sciences, recently these and post-modernist approaches have become more common. A variety of theoretic approaches have been advanced, including ethnography [152], ethnomethodology [18], action research [137, 153], critical theoretical perspectives [154], feminist theories [54, 55], diffusion of innovation [49, 155-157] and social interactionism based on diffusion of innovation theory (e.g. [29, 37-39, 42, 103, 158-160]). Among other recent examples are a variety of studies drawing on a constructivist tradition emphasizing organizational, political, social, and cultural concerns: constructivist [161] and social constructionist approaches [47, 55, 162, 163]; actor-network theory [76, 164], a sociological approach that employs sociocultural analyses and sociotechnical design as well as drawing on actor-network theory and situated action/design [3, 17, 165-169]; activity theory [170]; a purely subjective approach [171]; structuration theory and concepts based on it [52, 143, 172], or combinations of constructionist theories and structuration theory [173]; and Zerubavel’s [174] work on sociotemporal order [145]. Kaplan argues that each of these theoretical threads is a form of social interactionist theory [42, 103]. She sees in social interactionism an explanation for the concept of “fit” summarized above, and also a theoretical base from which to derive evaluation frameworks, principles, and guidelines [12, 13, 42].

Although the social and behavioral sciences have underpinned much work on people, organizational, and social issues, highlighting their contribution and explicitly drawing upon them is continuing to lead to new directions in evaluation for people, organizational, and social issues.

New Interdisciplinary Directions

There is increasing recognition in evaluation research that many different disciplines can contribute to evaluation efforts and to addressing different questions in technology assessment [135]. Social scientists, computer scientists, and information scientists all are exploring the complexities of people, social, and organizational issues, while individuals from disciplines as diverse as anthropology, communications, computer science, history, library and information science, management information systems, science and technology studies, and sociology are examining implications of the idea that technologies are embedded in social contexts [2].

There is, therefore, a growing tendency to form multidisciplinary evaluation teams, and to incorporate social scientists, computer scientists, health informaticians, experts in organizational behavior, and economics into implementation and evaluation efforts [175]. Evaluations undertaken by such teams may address important new evaluation questions [89]. These evaluations tend to be robust and very comprehensive, producing greater insight than those of less diverse evaluation teams. However, multidisciplinary evaluation research designs and research teams present inherent difficulties that should not be underestimated. Their very value is that of providing a different perspective, but this, in itself can prove divisive, disheartening, or demoralizing, depending on whether the team views these perspectives in positive or negative ways. Experiences in the UK are demonstrating that teams need to be fluid and flexible as projects develop and progress. It is notable that four of the published UK-based multi-disciplinary teams

---

b An organizational viewpoint involves a rational, mechanistic, explicit, fixed view of an organization operating according to standard operating procedures. A sociotechnical viewpoint involves a view of work as being performed through collaboration among groups and individuals who each have their own goals and aspirations, and who may not act according to the formal rules and procedures [23].
(Harrison et al., May et al., Peel et al., and Purves et al.) have evolved and reformed in structure over time, which suggests that it is worth striving for the benefits obtained from a multidisciplinary approach. How to conduct such evaluations and manage interdisciplinary project teams is a developing area in itself [43, 114, 135, 176].

Working group and other activities are also promoting an interdisciplinary evaluation approach [129]. New forums and growing interaction between medical informatics evaluation researchers and other research communities are bringing together researchers from different traditions and creating opportunities for findings from evaluation studies of different application areas to enrich each other. These working groups also have been instrumental in creating frequent sessions at information systems conferences [180-182]. Researchers in the field of information systems are more and more publishing evaluation studies conducted in health care environments (e.g. [43-45, 54, 91, 100, 103, 116, 130, 137, 172, 180, 183-189]). As interest grows in information technology in health care, other information systems conferences recently have included tracks on the topic.

These multidisciplinary and interdisciplinary efforts help counteract the insularity that evaluation studies and researchers have had from each other. This insularity has resulted in an impoverished understanding of people, organizational, and social issues [42]. Such efforts towards drawing on a variety of disciplines (those we include as well as others) creates synergies between medical and health care informatics evaluation research and other fields, enriching both. As researchers act as conduits for work which crosses disciplinary boundaries.

### New Challenges

#### Changes in Technologies and in Health Care Delivery

New technologies and new ways of organizing health care delivery bring new evaluation challenges and questions. New applications of information and communication technologies may be used in settings quite different from those in which evaluation studies have typically been undertaken, potentially raising not only different evaluation questions but possibly requiring different research designs. We take telehealthcare and telemedicine as an example of the kinds of issues that arise.

The telehealth evaluation literature discusses small-scale demonstration projects and feasibility studies. While considered methodologically inadequate [190], this literature discusses a range of problems that relate to evaluation [191-205]. Among these are not only issues about the nature of evaluation and evidence when technologies have not yet become stable [134], but also about changes healthcare delivery is undergoing. As with previous medical technologies and information systems, as well as newer consumer informatics applications and health management uses, telehealth technologies may be used in ways that redefine how healthcare is practiced and delivered [52, 142, 184, 206-209]. Telehealth, too, makes more evident concerns over what “information” consists of in health care [52, 60, 207, 210], and highlights how computer-mediated clinical encounters differ from face-to-face ones [210-214].

Additionally, telehealth provokes ethical questions little considered in evaluation – such as empowerment; equity and equality of services and use of health care resources; trust among clinicians; the role of employers in health care management; monitoring and surveillance of patients at home; the therapeutic or dehumanizing effects of health care technologies; as well as ethical issues of system design and evaluation itself – suggesting that such concerns should be reflected in evaluations of other areas of information technology in health care as well [89, 93, 173, 209].

Further, it is clear that health care itself is changing from single-institution episodic encounters between practitioner and patient to being more patient-focused without regard to physical or geographic boundaries [209, 215, 216], redefining what constitutes a clinical encounter and giving rise to virtual health care organizations [211]. Together with the growth of home health care and Internet health information and services, this increasingly complex and changing environment may require evaluation studies themselves to change from a focus on individual technologies, individual institutions, and individual users, to the changing context of patient-centered care and integrated delivery systems, and networked technologies that support them [59].

Telehealth and telemedicine then are examples of how new technologies, like some older applications, raise new issues related to different meanings information and communication technologies have for different users and how people make sense of them [50, 55, 61, 66, 75, 85, 103, 143, 200, 217]. Studies of imaging technologies, like ultrasonography [151], CT scanning [143-146], or clinical imaging systems [218, 219] point to the need for newer studies to understand contemporary systems [131]. As other new technologies – such as robots and augmented reality [220] in surgery, visualization used for genetic alterations as well as for medical
research, and other means of virtual or distant care – become more common, evaluation researchers will need ways of studying them and the issues they raise.

**Mutual Transformation and Co-Evolution of Technologies, Organizations, and Evaluations**

Developing and implementing systems within complex organizations is an inherently unpredictable process. Evaluation involves a variety of dynamic technologies, projects, and organizational settings. Both system and process boundaries can shift and may be hard to determine, while the nature of both the technology and the context may change during project phases [89, 221]. As a result, the focus of evaluation itself may change throughout a study as various actors, reflecting political, professional, and commercial interests, adapt, modify and transform themselves, the technology, and the evaluation.

Because of these moving targets, it is increasingly apparent that causal relationships are not unidirectional or straightforward [42]. Researchers in several fields are moving away from studies based on technological or socio-economic deterministic models. Instead of seeing evaluation as a neutral technical process of applying specific methods, the process of producing evidence of efficacy and utility is recognized as resulting from a series of contingent decisions, even where the design of evaluation or research projects apparently is structured in a quite rigid and “objective” way [60, 161, 206, 222].

As a result, new questions are being raised concerning the nature and role of evaluation itself [90, 223]. A recent trend in evaluation takes a postmodern stance and turns reflexively to examine evaluation itself [222]. Researchers, rather than studying “impacts” of technology, favor viewing technology and society as part of a “seamless web” [2, 224] in which implementation is seen to involve a series of negotiated definitions and actions rather than as a fixed process with pre-defined successes and project uses [62, 150, 161]. Evaluation is seen as a fluid process that is but one component of extended social and technical networks.

These contingent processes present major challenges for evaluators who are dealing with a technology that is applied and deployed in the real world of health care provision, rather than in the laboratory of system developers [134]. Because numerous contextual aspects can change over the course of a project, co-evolving in a process of mutual transformation and thereby changing the context from the one when the evaluation originally was planned [3, 42, 46, 221], a conceptual model may help illuminate some of the complex dynamics of evaluation. It also might challenge basic assumptions of evaluation itself, viz., whether more information leads to better decisions, about what exactly is being evaluated in the face of so many moving targets, about what the point of evaluation is once a technology has become stabilized and its use routinized, and whether evaluation – or any other study – provides the “scientific”, evidence-based information that is sought. Recent efforts address such issues [89, 90, 223].

**Recommendations**

As the many studies cited here indicate, integral to system success is not only system functionality, but also a mix of organizational, behavioral, and social issues involving clinical context, cognitive factors, methods of development and dissemination [42, 63, 87, 95, 225-227], and who defines “success” and when the determination of “success” is made [124]. Numerous studies support the observations that: “Sociologic, cultural, and financial issues have as much to do with the success or failure of a system as do technological aspects” [228] because “information technologies are embedded within a complex social and organizational context” [229].

Thus evaluation needs to address more than how well a system works. Evaluation also needs to address how well a system works with particular users in a particular setting, and further, why it works that way there, and what “works” itself means. Evaluation needs to address how conceptions of whether a system works change with time and with who is making the judgment. This scope to evaluation will help answer such key questions as:

- Why are the outcomes that are studied as they are?
- What might be done to affect outcomes?
- What influences whether information and communication technologies will have the desired effects? and desired by whom?
- Why do individuals use or not use an informatics application?
- What from one study might be generalizable or transferable to other sites or applications?
- How can information and communications technologies be used so as to improve patient care?

To help do this, a number of areas need additional development. We, therefore, propose the following:

1. **Address concerns of the many individuals involved in or affected by informatics applications**

Many evaluations focus on practitioners, primarily physicians [42], and most studies do not clearly identify primary stakeholders [124], yet different groups have different reactions to a system [25, 230]. Different participants (e.g., hospital specialists, general practitioners, and patients) have different perceptions of the same application, and those perceptions, in turn, can differ from those of the researchers conducting a study [132]. This makes it all the more important to take account of the ways that different individuals and groups may interact, may variously define a project or information needs, and may differ in their views of what constitutes health and health care or a successful informatics project [75, 89, 154].

In addition, assumptions underlying both physician-centric health care and basing informatics applications on what experts (e.g., clinicians, system designers, vendors, insurance companies, legislators) think is best for others are increasingly being challenged. Evaluation research that considers direct effect on patient care or...
health outcomes for patients has been scarce [231-235]. Even in studies of patient care information systems, few evaluations focus on impact on care [124]. The notorious difficulty of measuring direct patient outcomes related to information technology leads to proxy measures. Nevertheless, effect on patient care is an important criterion by which health care professionals judge innovations [70, 236]. The fundamental goal of medical informatics, and hence all associated research, is to improve care. Although it is exceptionally difficult to evaluate whether care is improved, the attempt should be made. Health care, and ultimately, health, is a people, organizational, social, and cultural issue [237].

Recently, evaluators have been paying more attention to patients and health care consumers in evaluations. Both patient perceptions and patient outcomes deserve more study. More evaluations are needed to address concerns of the many individuals involved in or affected by informatics applications, including not only more study of nurses, administrators, pharmacists, patients, laboratory technologists, or personal caregivers, but also the interests of vendors, policy makers, third-party payors, etc.

2. Conduct studies in different types and size sites, and with different scopes of systems and different groups of users.

Do multi-site or multi-system comparative studies?

Most studies are small-scale, single site, and of stand-alone systems [114]. Many are in academic settings in industrialized countries. Variety in kinds of study site is needed, as are large-scale systems studies, multi-site studies, and comparative studies. Studies of different kinds of systems, studies at different kinds of sites, and comparative studies are important for making underlying assumptions more apparent and for illuminating contextual issues.

Studies need to be undertaken outside academic medical centers and satellite provider centers. Rural health care facilities, inner city clinics, emergency rescue operations, nursing homes and geriatric centers, remote facilities in developing areas or at military installations, home health management systems, or home use of web-based health information – some of many diverse health care settings – face problems not common among individuals at academic medical centers [238]. They may raise different health care delivery and evaluation issues.

Systems may not be easily transportable across geographic or cultural boundaries [87]. When systems are developed in one context and moved to another, assumptions about development, implementation, and use may be challenged [189]. Environmental differences, such as availability of reliable power, computing resources, a system of patient identifiers, and sufficient economic resources, let alone experience with electronic storage and retrieval systems, also might come into play, as they did for an international team successfully modifying an electronic medical record system developed at a US academic medical center for use in rural Kenya [53, 77, 78]. Reporting these issues and examples of how to adjust to local conditions can help those in similar situations help increase validity by generalizing across implementations, and point to what evaluation results may be transferable.

Comparative studies, too, are needed. Such studies may be designed to compare similar groups using the same technology at different sites, different groups using the same technology at one site, or various other combinations. The value of such research is illustrated by comparative studies of an electronic medical record [239], physician order entry [11, 49, 75, 156, 175, 240-245], pediatric office systems [217], CT scanning [37, 143], physicians’ use of images [218], Community Information Systems (CIS) within the UK’s National Health Service [246], electronic mail use vs. web use among the same provider population [238], a stroke guidance system evidence-based practice tool [247], and a drug distribution system [85]. Multi-site studies can lead to increased ability to generalize from case studies [246]. However, comparative studies are difficult [114], and extending studies not only across sites, but cross-culturally, remains a challenge [248].

Much may be learned from experiences in places with marked differences from where a system was first developed and from multi-site studies. Such research makes contextual issues more apparent, along with those assumptions on which medical informatics developments are based.

3. Incorporate evaluation into all phases of a project

Evaluation methods and questions depend upon both system development phase and purpose of the evaluation [5, 6, 14, 15, 102, 110, 249]. Recent software engineering methodologies use evaluation methods throughout the software life cycle [22]. Usability testing, iterative design, and prototyping, which blur life cycle phases, are premised on the importance of evaluating a system before finalizing its design. Evaluation may be done concurrently with various project phases. It should be an on-going process throughout the life of a project, and include a variety of approaches [95]. The role of evaluation in design serves as an example.

Using evaluation to influence system design enables building into a system an understanding of users’ goals, roles, tasks, and how they think about their work. To do this, situated action and participatory design approaches – often based on the influential work on situated action by anthropologist Lucy Suchman [250-252] – have been undertaken in efforts to link work design and software design, including attempts to model work according to users’ views [42-45, 253, 254]. These themes are apparent in a special issue of Artificial Intelligence in Medicine [255]. The authors draw on successfully Scandinavian participatory design approaches [256] and on the writings of Winograd and Flores [257], as well as on Suchman’s research. A stream of work undertaken by Timpka and colleagues promotes action design, a combination of action research, participatory design, and situated action (e.g., [82-84, 148, 258, 259]). An underlying principle is that “knowledge can never be decontextualized” because knowledge “is situated in particular social and physical systems” and “emerges in the context of interactions with other people and with the environment” [260]. Another undertaking uses cognitive approaches to identify major cognitive problems physicians will encounter when dealing with an interface [22, 261-264],
while another employs ethnographic techniques for usability testing [265].

Evaluation can be used to influence system design, development, and implementation. While results of post-hoc or summative assessments may influence future development, formative evaluation, which precedes or is concurrent with the processes of systems design, development, and implementation, can be a helpful way to incorporate people, social, organizational, ethical, legal, and economic considerations into all phases of a project [3, 6, 12, 13, 22, 62, 266-268].

4. Study failures, partial successes, and changes in project definition or outcome

“Success” and “failure” have different and changing definitions in different projects and among different stakeholders [3, 124]. Attention, nevertheless, is needed not only to successes, but also to failures, partial successes, and changes in project definition or outcome. Despite an accumulation of best practices research that has identified a series of success factors, many projects still fail [246]. Some 40% of information technology developments in a variety of sectors are either abandoned or fail [91], while fewer than 40% of large systems purchased from vendors meet their goals [269]. Similar numbers have been estimated for health care, and the number has unfortunately remained approximately the same for at least the last 25 years [14, 231, 270-272]. Some researchers have examined failures, removals, sabotage of systems, or how failures became successes or were otherwise redefined (e.g. [47, 91, 161, 162, 242, 273-280]); and some high-profile failures have been reported (e.g. [130, 246]) – the failure of the London Ambulance Service’s dispatch system, for example, has been studied extensively (e.g. [186, 274, 281, 282]). Nevertheless, over the years, publication bias has provided little opportunity to learn from studies in which technological interventions resulted in null, negative, or disappointing results [231, 283]. These failures need to be studied and reported.

The political nature of evaluation and the understandable reluctance to be an institution or project reported as an example of “failure” also contribute to the aversion to reporting project difficulties. Further, stakeholders have an interest in promoting a project despite evaluation results. Projects may simply be strategically redefined as “successes” [284]. In addition, projects often are defined in terms of the technology [3], a focus which emphasizes how technology is used and the benefits it could bring. This focus may skew notions or reports of “success”; in health care, much as with other computerization movements [72, 285]. The combination of publication bias, avoiding embarrassment, project promotion, and ideology contribute to uneasiness in reporting or hearing studies that challenge the prevailing ethos.

5. Develop evaluation approaches that take account of the shifting nature of health care and project environments, and do formative evaluations

Evaluation also is affected by not paying sufficient attention to failure in another way. A failed project is less likely to be evaluated after the fact. Focusing on, or assuming, success may reinforce a tendency towards summative evaluation, i.e., evaluation undertaken after a project is completed. Randomized controlled clinical trials and experimental research designs also increase the tendency towards summative evaluation in that studies of process can be seen to “taint” research involving outcomes of interest.

Summative evaluations often assume that systems, organizations, and the environment in which they operate are stable, so that outcomes may be compared to the situation before the intervention, and also compared with a priori expectations and goals, even though these measures may no longer apply. However, summative evaluation approaches do not work well when situations change radically; the assumptions on which they are based do not hold under such conditions [91]. Definitions of both project and evaluation success shift as they are negotiated over time among participants. Evaluation research designs need to take account of the flux in and the dynamic nature both of projects and of evaluations, and also to provide feedback to projects so they can adjust better to what is going on [286]. Nevertheless, evaluations tend to be summative [124]. Much is to be gained from formative evaluation, i.e., evaluation conducted during the course of a project. Formative evaluation has many advantages, including: a) allowing for changes and uncertainties in project technology and implementation; b) capturing the fluid nature of project and evaluation objectives and in evaluation design; c) enhancing organizational learning and buy-in; d) attracting management attention; e) identifying unexpected or emergent benefits; f) tracking a project over time longitudinally; g) adjusting study design and questions to the site involved and to new insights as they develop; h) allowing for varying and changing definitions among site and research participants of the systems project, the evaluation, or of what constitutes project goals or “success”; i) enabling evaluation at multiple sites each at different stages of implementation; j) tracking ways in which users “re-invent” the system or how a system and organization are mutually transformative, thereby enabling learning by monitoring the many experiments that naturally occur spontaneously as part of the processes of implementation and use; and k) providing the possibility of influencing system design or implementation [3, 12, 13, 44, 91, 102, 110, 114, 221].

6. Incorporate people, organizational, social, cultural and concomitant ethical issues into the mainstream of medical informatics

Change is needed to incorporate people, organizational, social, and cultural issues into the mainstream of medical informatics in general and into evaluation research in particular. Publication, practice, and curricula should reflect the importance of these issues. Medical informatics tends to “delete the social”, resulting in devaluing non-technical aspects of systems design and implementation [287]. Many evaluations employ study designs to focus on system and information quality and technical issues. They rarely consider social, ethical, political, organizational, or cultural issues; system development processes; or implementation processes [88, 89, 124, 288]. They pay little
attention to the roles of politics [289] or affect [290], to take two rather different examples, in organizational interactions, clinical encounters, and innovation. They little acknowledge how medical informatics privileges particular kinds of information and users, work roles and statuses, and organizational structures; embodies assumptions, such as about clinical work, hierarchy, information flow, responsibility for health and health care, and gender issues [143, 154, 178, 291]; and ignores relationships individuals form not only with each other but also with technologies [292]. More broadly, social concerns, value judgments, relationships with communities and families, the central role of health and disease in social relations and spiritual considerations, and the like, also are ignored. All these, as well as the nature of the assumptions embodied in system design, evaluation research, and medical informatics itself, clearly also have ethical dimensions that need addressing.

While it is increasingly recognized that information technology is embedded in a nexus of social, organizational, political, cultural, and personal concerns and norms, more work – and likely a culture change within medical informatics – is needed to examine these issues and to incorporate the findings into design and implementation. Indeed, evaluations and assessments based on randomized clinical trials and experimental designs are considered “gold standard” precisely because they attempt to eliminate all influences except the intervention itself and because they are thought to be more “scientific” and value-free [89, 293, 294].

7. Diversify research approaches and continue to develop new approaches

Experimental designs and randomized controlled trials dominate in medical informatics evaluation [42, 88, 295, 296] and are advocated as the best approaches, particularly if combined with economic analyses [293, 297]. They also are more aligned with management values [93]. For a number of reasons, including lack of attention to a variety of human, contextual, and cultural factors that affect system acceptance in actual use, these approaches have come under increasing criticism [16, 19-22, 42, 88, 89, 114, 221, 229, 295, 298-300]. Other approaches, when used under controlled conditions, also have been similarly criticized [42]. Some have called for making it a priority “to develop richer understanding of the effects of [system] benefits in health care and to develop new evaluation methods that help us to understand the process of implementing it” [229].

A school of thought has developed suggesting that neither randomized controlled trials, experimental designs, nor economic impacts are suitable in and of themselves for evaluating informatics applications. Such designs may pinpoint what changed, but they make it hard to assess why changes occurred. Additionally, these traditional designs prove difficult for following changes as they are developing, or in determining system design and implementation strategies that are well suited to particular institutional settings and societal considerations. Longer-term field studies and more interpretive approaches are better for investigating processes, multiple dimensions and directions of causality, and relationships among system constituents and actors [12, 13, 161, 301].

Quantitative study designs are considered the gold standard. A review of evaluations of patient information systems indicated few case studies, the only subjectivist approach the authors found [124], with similar tendency evidenced in evaluations of clinical decision support system [288]. Consequently, here a special plea is made for incorporating qualitative/interpretive/subjectivist methods, without prejudice to other approaches. Including such approaches in evaluation research reveals issues that otherwise would not surface [295]. As numerous well-regarded evaluation researchers have pointed out, qualitative methods especially are needed in order to study organizational, social, contextual, cultural, ethical, and people issues, particularly in the complex and dynamic environments in which health care information technology is used. They may be used to: a) explore how participants conceptualize, perceive, understand, and attribute meaning to systems and to their work; b) examine how organizational, social, and historical context influence, and are influenced by, system implementation and use; c) investigate causal procedures by identifying not only what changes occurred, but why; d) identify unexpected consequences; e) form the basis for formative evaluation; f) focus on routine, often tacit, ways people do things in practice, and identify recurring patterns of action and interaction; g) reveal the close relationship between work practices and organizational processes, even if on the surface they appear not to be rational; h) make more evident how technology embodies values and assumptions, and how individuals project values onto its design and use; i) point to different kinds of information and how various kinds and forms of information are conveyed and used in clinical practice; j) allow results to be incorporated throughout the life cycle, from planning through design into implementation; and k) increase utilization of evaluation results, in part because of a focus on what information is relevant to particular situations and individuals [3, 5-7, 16, 18, 22, 23, 42, 88, 114, 138, 149, 173, 175, 221, 299, 301]. For reasons like these, some consider it ethically imperative to incorporate qualitative methods into research design [173].

Related to an emphasis on experimental designs, perhaps, is a search for generalizable lessons or success factors. Generalizability is an important goal, and we argue below for the need for it. Nevertheless, while valuable for pointing to significant considerations in system design and implementation, lists of success factors lose their potency by being highly generalized. Notwithstanding the lists included in this article, how lessons or factors will play out in any given situation is what is most important in that situation. Consequently, insights gained from close study of different implementations are crucial for improving insights into whatever new situation currently is at question [3]. Here again, qualitative methods are tremendously valuable in providing the richness of such insights and suggesting their applicability in or how they might be transferrable to various settings and under various guises.

We also advocate more than mere tolerance for a variety of approaches. Innovation is needed in evaluation as well as in
8. Conduct investigations at different levels of analysis

Most evaluation studies focus on groups of individuals that fit familiar categories in health care: nurses, physicians, developers, etc. Researchers collect data on the individual level and aggregate it according to these categories. Important differences among these individuals, however, may be obscured by grouping them together in this way. For example, a study of laboratory technologists indicated that individual technologists could be categorized based on how they conceptualized their jobs, and also that individual laboratories within the institution could be so categorized [43]. Simply considering the category “laboratory technologist” would have lost these findings and revealed little of interest in how laboratory technologists responded to a new laboratory information system. Similarly, focusing on a health care team instead of on individuals helped elaborate the need to support the collaborative nature of work in environments such as intensive care units [178].

Concern with granularity and units of analysis increases in importance as health care changes and information technologies become more widespread. The availability of new technologies may change what individuals do, thereby changing their relationships with others. These changes can affect social, occupational, or organizational structures. For example, micro-level work role changes related to introducing new information technologies resulted in structural organizational changes when CT scanning and subsequent imaging technologies were introduced [144, 146].

Changes at an individual level, as in information and communication technologies to support home health care, also have societal repercussions with respect to health, public health, and the organization of health care. Health care information technologies are diffusing across geographic boundaries and health care itself is becoming less localized in centralized institutions. Individuals increasingly use these technologies to gain information about their health and health care and to manage it [209]. Community informatics, for example, uses the community as a level of analysis to study how communities pursue their goals in such areas as local economic developments, cultural affairs, civic activism, electronic democracy, self-help, advocacy, and cultural enhancement [182]. Formation of on-line communities and other forms of participation in health care can be a part of these activities, with consequent implications both for informatics applications and evaluation research, as well as for health and health care per se.

Like cross-cultural and comparative studies, research employing different degrees of granularity and different units and levels of analysis, and research investigating how changes ripple across them, may provide not only new insights, but also may challenge basic assumptions derived from traditionally focused studies.

9. Integrate findings from different applications and contextual settings, different areas of health care, studies in other disciplines, and also work that is not published in traditional research outlets

Medical informatics already has been enriched by scholars with degrees in a variety of disciplines. Evaluation researchers include specialists in organizational behavior, education, information systems, communications, history, biology, computer science, anthropology, engineering, sociology, library and information sciences, psychology, social studies of science, and, of course, numerous health care specialties – to name but some of those disciplines. Researchers who keep up with another field of study enable their, and consequently others’, medical informatics work to be informed by new theories, knowledge, and approaches. Different disciplines may ask new research questions in people, organizational, and social issues, and employ different ways of addressing them.

In addition, much valuable work is undertaken outside usual academic and research contexts. Mechanisms are needed to report and disseminate work that is not published in traditional research outlets. For example, insights gained through evaluations of governmental projects worldwide, vendor and consultant activities, and experiences in developing countries, need to be more widely known. Practitioner knowledge, gained through much experience, also would provide insights not easily obtained through other means.

Both evaluation methods and theory would benefit from continued efforts to integrate findings from different applications and contextual settings (including ones not in health care), and to bring together understanding developed through studies and other activities undertaken in different areas of health care as well as in other disciplines. This latter integration is beginning to occur in social informatics, producing generalizations that also are applicable in medical informatics [302]. Medical informatics, as a field, would be enriched by integrative activity across studies, disciplines, and practice. Moreover, findings from these other areas would both enhance and can be enhanced by related medical informatics evaluation research.

10. Develop and test theory to inform both further evaluation research and informatics practice

Evaluation studies tend to be undertheorized. Compared with other disciplines, there is more emphasis on context than on theory [137]. Although some significant streams of work start from an explicit theoretical base, and use it to organize and analyze results, little has been done to generate theory across studies. Testing or generating
theory need not diminish from contextual description and consideration. By integrating findings across different studies, theory can be generated and tested, and individual studies can build into cumulative bodies of generalizable and transferrable knowledge.

Many medical informatics studies are atheoretical, or evidence tacit theoretical assumptions rather than explicit discussion [42]. Much medical informatics work is characterized by considering technology, work, specific clinical practices (e.g., ordering laboratory tests), profession, and institution, as static and isolated constructs. This approach allows studies to be designed to investigate the impact of one factor, such as a system reporting drug-drug interactions, on another, such as reduction in medication errors. It ignores, however, the complex network of social interactions that surround the introduction and use of such a system by isolating the system as a static, individual item rather than as part of a nexus of practices, relationships, history, regulatory requirements, etc. and the meanings that are attributed to all these [146]. Evaluation researchers could theorize concepts and constructs so that they include dynamic social and institutional nexuses of which they are a part, as well as to integrate across studies to develop theory to inform both further research and also informatics practice. Evaluation researchers, too, need to turn their theoretical and empirical eyes reflexively on evaluation research itself.

Conclusions

Our review leads us to observe common threads that run through newer orientations towards evaluation. These themes are shared with trends in social informatics studies in other fields [303]. The underlying basis for attention to people, organizational, and social issues is that human and organizational concerns should be taken into account during system design, implementation, and use. International perspectives are converging to a broad and encompassing multi-method approach to evaluation throughout the life of a project, with studies conducted in actual clinical settings so as to allow for complex contextual issues to be addressed through a variety of theoretical lenses [42, 110].

Considerable work has been undertaken concerning appropriate evaluation paradigms. Newer evaluations build on the growing body of empirical and theoretical evaluation research that focuses on roles of different actors and the connections between them; on contextual, organizational, social, and cultural concerns; on meanings attributed to the experiences by the persons involved; and on the processes and interactions among different aspects of system design, implementation, and use. These trends share a holistic view and the conviction that analyses be set in a complex context of work group, professional community, cultural unit, or organization. Moreover, there is a general commitment to methodological pluralism and development of approaches for studying numerous areas, issues, and concerns that have not been well addressed.

We have reviewed this body of research and suggested directions for further development:

1. Address concerns of the many individuals involved in or affected by informatics applications.
2. Conduct studies in different type and size sites, and with different scopes of systems and different groups of users. Do multi-site or multi-system comparative studies.
3. Incorporate evaluation into all phases of a project.
4. Study failures, partial successes, and changes in project definition or outcome.
5. Employ evaluation approaches that take account of the shifting nature of health care and project environments, and do formative evaluations.
6. Incorporate person, professional, organizational, cultural, and concomitant ethical issues into the mainstream of medical informatics.
7. Diversify research approaches and continue to develop new approaches.
8. Conduct investigations at different levels of analysis.
9. Integrate findings from different applications and contextual settings, different areas of health care, studies in other disciplines, and also work that is not published in traditional research outlets.
10. Develop and test theory to inform both further evaluation research and informatics practice.

Acknowledgments

This work by N. Shaw was funded when she was a National Health Service Executive (North West Region) Research & Development Post-Doc
toral Training Fellow. The support of the UK Department of Health, formerly the National Health Service Executive (North West Region) Research & Development Department and Oldham Primary Care Trust is therefore gratefully acknowledged.

References


165. Berg M, Langenberg C, Berg I, Kwakkenbait J. Considerations for sociotechnical design: Ex-


212. Pinch TJ, Bijker WE. The social construction of facts and artifacts: Or how the sociology of science and the sociology of technology might benefit


286. Stead WW. Matching the level of evaluation to a project’s stage of development. JAMIA 1996; 3 (1): 92-94.