

Telemedicine Evaluation

RASHID BASHSHUR, Ph.D., GARY SHANNON, Ph.D., and HASAN SAPCI, M.D.

ABSTRACT

This paper examines the context and status of evaluation research in telemedicine, and it proposes a two-pronged strategy for addressing the critical policy and programmatic concerns in this field. It explains the evolution of evaluation research in the United States, and it describes a comprehensive typology and requirements for valid evaluation. Major impediments for definitive evaluation are discussed, together with a summary of major trends in empirical studies. Two concurrent strategies are proposed for producing definitive findings and for assessing the available empirical evidence. These consist of large-scale experimental studies and theoretical and empirical triangulation for assessing the available empirical evidence.

INTRODUCTION

THE INITIAL IMPETUS behind the development of telemedicine was its intuitive appeal as an effective substitute for in-person medical care and as an efficient tool in developing integrated systems of care. This appeal has been supported by substantial experiential data in various settings and clinical applications and, more importantly, by a growing body of empirical knowledge that, though mostly positive, has yet to reach the level of conclusive and definitive evidence needed for universal acceptance. At this stage in the evolution of telemedicine evaluation, it is possible to reach equally justifiable yet differing conclusions as to the significance and even the justification of evaluation.

On the one hand, one can claim that the ultimate adoption and diffusion of telemedicine as an integral component of the medical care system will depend on the hard evidence produced by rigorous scientific studies that evaluate its benefits and costs. From another point

of view, one can also claim legitimately that attempts to evaluate telemedicine as a well-defined and discrete modality of care are, in the end, futile due to the fact that it is difficult, if not impossible, to differentiate it from related concepts. Moreover, the underlying technology of telemedicine has developed a life of its own, which continues to evolve on a fast pace. Telemedicine applications also continue to proliferate and now encompass all clinical areas in addition to public health and medical and health education. Therefore, it is futile to define a specific technology/human "fix," and call it telemedicine.

Indeed, since the field (including technology and applications) is in constant flux, the most appropriate evaluation should be aimed at investigating the benefits and costs of alternative modalities and various dynamic combinations and configurations of technology, human resources, and health applications.

We chose to combine elements of both arguments in this paper to move our thinking forward on telemedicine evaluation as well as the

field of telemedicine. At the same time, we acknowledge the need to be creative and flexible in evaluating this field to meet the requirements of various stakeholders, including policymakers, program developers, providers, and clients. But, in the end, the goal of evaluation should be to produce objective and credible evidence regarding the merits and problems of this field.

In this paper, we discuss basic issues and requirements for the scientific evaluation of telemedicine; review the status of evaluation research in telemedicine from both methodological and substantive perspectives; and propose appropriate strategies and methodologies that can be used in future evaluation studies and in interpreting the available empirical evidence. More specifically, we describe in general terms the origin and goals of program evaluation research, the objectives of health-care evaluation, a basic typology of evaluation methods, the status of evaluation research in telemedicine, and how we may interpret the empirical evidence produced to date. As a detailed illustration of how we can use the available evidence, we provide a theoretical model/scheme for evaluating the impact of telemedicine on access to care. This illustration is limited to telemedicine's effects on access as other presentations in this issue focus on outcomes and cost.

BASIC ISSUES IN EVALUATION

The evaluation of health-care programs combines scientific requirements and political realities, which are often incompatible. Scientific requirements pertain to the need for robust research designs, reliable and valid measurement, as well as rigorous methods for data collection and data analysis. Political realities stem from the priorities of public policy and funding agencies and the process of allocating research funds.

Health-care evaluation may focus on a specific intervention or device or it may encompass an entire health program or initiative. When encompassing a program, evaluation rests on the reasonable premise that allocating public resources for large-scale health pro-

grams ideally is based on their expected costs and benefits. Indeed, prudent policy must be informed by factual information—where such evidence exists—on costs and benefits of alternative programs and initiatives aimed at achieving the same goals.

The evaluation of specific medical interventions or medical devices, although based on a similar set of assumptions as those for programs, has a more limited focus. Specifically, the safety and effectiveness of new medical interventions and devices must be demonstrated before they can be approved for regular professional or consumer use. However, both conceptually and factually, telemedicine/telehealth/ehealth combines elements of both health programs and related medical or health interventions. Individual researchers, however, are at liberty to (and frequently do) focus on either programs or interventions or, less frequently, on some combination of the two. In any case, there currently is near-universal consensus that public policy on investment in telemedicine/telehealth/ehealth programs should be based on scientific evidence of their benefits and costs, as compared with alternatives. Hence, it is time for the telemedicine community to reach an informed consensus on an optimal methodology to collect empirical evidence and interpret findings in an objective and credible manner.

The discussion here focuses on the broader problem of program evaluation rather than the assessment of safety and effectiveness of specific medical or health interventions or devices. (The latter problem is addressed briefly later.) The issues of interest here include the origin and goals of program evaluation in general, a typology of evaluation objectives (or classification of what is evaluated), evaluation methods (or approaches), and the inherent methodological and logistical difficulties in evaluating telemedicine programs.

ORIGIN AND GOALS OF PROGRAM EVALUATION

The genesis of program evaluation research and the involvement of social scientists in public policy can be traced back to the 1930s when President Franklin Roosevelt's New Deal poli-

cies established a number of large-scale social welfare and public works programs aimed at ending the Great Depression, creating social equity, and invigorating the economy. Indeed, the development of the welfare-oriented state in the 1930s in the United States introduced social science involvement into public policy. Some of the best empirical social research during that decade was made possible by the Works Progress Administration and the National Youth Administration, which provided funding for research assistants and unemployed academics. Nonetheless, as a method of inquiry, evaluation research did not introduce any new or different research designs or analytic techniques. Hence, it is fair to characterize evaluation research as lacking a unique or distinctive methodology. In fact, evaluators use any or all of standard methodologies, including clinical trials, sample surveys, and even the dreaded focus group, which ignores nearly all requirements for rigorous scientific research.

Most of the early work on the evaluation of the New Deal was done by the Roosevelt Brain Trust, consisting only of economists—no sociologists, psychologists, or anthropologists. Some 30 years later, the Great Society programs of the 1960s represented an expansion of the New Deal, reflecting the political climate of the time. These programs were aimed at reducing, if not eliminating, the significant economic, social, educational, and health disparities in this country. In retrospect, the War on Poverty of the 1960s was established without any empirical assessment of New Deal programs, or any other federally sponsored social or public work program for that matter. Indeed, we know little, if anything, about the actual impact of specific New Deal programs in relation to their costs and benefits, except that unemployment rates actually declined and the Great Depression came to an end.

In contrast, we do have data to make a reasonable assessment of the War on Poverty in terms of employment trends, health and vital statistics, enrollment in educational programs, and the incidence of crime. However, none of these assessments would rise to the scientific level of establishing cause-and-effect relationships between specific interventions and outcomes. Therefore, we do not know with cer-

tainty the specific effects of initiatives in health, civil rights, early education, and other sectors, nor do we know the specific cost-benefit ratios that may have accrued from the substantial national investment in these programs. What was true then and is still true today is that the politics of social program implementation did not mesh well with the scientific requirements for producing accurate information. At the time, the overriding concern was to reduce as soon as possible basic socioeconomic disparities in society that brought about major social unrest. Programs were implemented without a careful analysis of various options or approaches that might be pursued for reaching this goal. Similarly, telemedicine was introduced and has been developed to address overriding public and political concerns with inequities in access to care, issues pertaining to the inequitable and uneven distribution of quality of care, and, of course, the ever-increasing cost of health care.

OBJECTIVES OF EVALUATION IN HEALTH CARE

As explained earlier, there are two types of evaluation research in health care. The first is concerned with testing the efficacy, effectiveness, and safety of specific medical interventions, devices, or medications. This type of evaluation research is typically used by the Food and Drug Administration to grant approval for professional or consumer use. The second type is the assessment of health program performance and achievement in terms of stated goals and objectives. Both types are complex, and both require the use of rigorous scientific methods. However, the methodology for testing the safety and effectiveness of specific devices or interventions is more direct, simpler, and less problematic than program evaluation, and it is more amenable to conventional scientific requirements. This is because the experimental variable is discrete or well defined, its effects are determinable, experimental allocation of cases is possible, and, with notable exceptions, the outcomes of interest are usually observable within the life of the clinical trial.

Health program evaluation is more complex by virtue of the fact that the experimental vari-

able(s) are not usually discrete or specific, experimental allocation is difficult at best, outcomes are often complex and their attribution difficult to make, and, midstream/midcourse changes in the program are often encountered. Furthermore, health programs have a variety of effects: direct and indirect, intended and unintended, and immediate and delayed. It is near impossible to capture all these effects within the life of a sponsored project, which typically lasts no more than 3 years. Nonetheless, program evaluation is essential, and it serves as an appropriate tool for public policy.

Health program evaluation can serve several policy objectives: (1) to determine success or failure of programs in reaching explicit policy objectives; (2) to ascertain if there is a less costly alternative to achieve the same objectives; (3) to determine if the program has undesirable unintended effects; and (4) to make informed policy decisions on continuation, termination, or change in the program, and whether to increase, maintain, or decrease funding for it.

Because its findings are based on scientific evidence, evaluation research has a significant effect on "normalizing" new technology in the academic community and for promoting its acceptance among practitioners and policymakers. Such evaluation can establish a factual foundation for the prevailing wisdom. The results are credible because scientific methodology consists of a set of mediating and standardized practices based on the logic of the scientific method. Indeed, the evidence produced by rigorous science provides the "seal of approval" or "gold standard" for professionals and the public at large. But, unlike basic research that tests specific theory or explicit hypotheses derived from theory, evaluation research is based largely on a conceptualization and measurement of operating systems in the real world, and it is aimed at determining their success or failure in achieving explicit goals and objectives as well as their unintended effects.

EVALUATION TYPOLOGY

Program evaluation categories are based largely on measurement feasibility and on the developmental stage of program implementa-

tion. A comprehensive typology of evaluation would include: (1) evaluability assessment, (2) documentation evaluation, (3) process or formative evaluation, and (4) summative or outcome evaluation.

Evaluability assessment: This can be conducted when starting or even planning large-scale programs with a serious intent on having the program evaluated systematically subsequent to its implementation. This type of assessment can be used effectively to frame the research questions, to determine the research design requirements, to develop the measurement tools and the data collection schemes, and to anticipate the analytical and statistical methods needed to manipulate the data. Its importance derives from the ability to clarify the goals and objectives of the program at the outset; to make explicit the intended effects, both short-term and long-term; to identify the variables of interest and how they will be measured; and, finally to specify the analytical tools for interpreting the findings when they become available.

Documentation evaluation: This is usually optional, and it consists of a narrative description of the actual implementation of the program. To be useful, documentation must include a true description of the steps, procedures, and protocols that were used in the implementation, as well as a reliable assessment of missteps, pitfalls, and problems encountered during implementation, and how the latter were addressed. Obviously, the important rationale for documentation evaluation is to provide potential new entrants to the field with important information on successful and unsuccessful approaches, pitfalls to be avoided, and possible ways to deal with problems that may arise.

Formative or process evaluation: This begins once a program is in place. Data can be gathered on its effects on the process of care. This is referred to as formative or process evaluation.

In many instances, large-scale programs are based on philosophical principles or policy imperatives that call for evaluation. For example, telemedicine is intended to bring about funda-

mental changes in the care process, changes that eliminate not only geographic barriers but also social and economic disparities in access to health care. Telemedicine is expected to provide a less costly alternative to in-person care by streamlining the care process, reducing the need for travel, and by providing effective substitutions in terms of site of care and provider mix. It is also expected to assist in developing integrated systems of care as well as in reducing physician isolation from mainstream medicine, thereby improving the overall quality of care. These expectations have prompted the Congress of the United States, like public and private agencies in other countries, to endorse telemedicine and provide financial support for telemedicine projects, in sharp contrast to the more cautious and contemplative approach by other branches of government, such as the Center for Medicare and Medicaid Services (formerly the Health Care Financing Administration), for fear of potentially devastating effects of paying for pent-up demand for care, especially in under-served areas and under-served populations.

Despite the inherent interest in outcome evaluation (described below), the vast majority of evaluation research focuses, by necessity, on process variables. Formative or process evaluation can capture information rich in behavioral, attitudinal, and cognitive changes that are likely to occur in the short term and can also be attributed to the program both methodologically and logically. Moreover, because these changes can occur in the short term and can be logically linked to the intervention, validity threats are reduced. This makes it more feasible to rule out certain rival hypotheses or explanations for observed trends in the data. Process variables are significant when there is a clear and logical connection between them as precursors and the ultimate outcomes in the sense that, if they occur, the outcome of interest is likely to occur also.

Summative or outcome evaluation: This can provide definitive evidence regarding the intended effects of a program, when it is based on rigorous scientific methodology. Moreover, the clearer, the fewer, and the more explicit or discrete the objectives of the program, the more

definitive can be the findings from outcome evaluation. That is why positive results of an outcome evaluation can be used as the ultimate justification of programs.

Because striking effects can be produced by the introduction of new hardware and software, there is often acute interest in “technological fixes” that are easy to install and likely to produce these effects, especially in the short term. Unfortunately, the introduction of technological fixes is also associated with high expectations, or expectations set at levels that are difficult to achieve quickly. The irony here is that moderate success may appear to fall short of high expectations. Moreover, technological fixes often have unforeseen and unintended effects, both positive and negative. These are usually difficult to account for and to measure.

Insofar as telemedicine is concerned, the most useful model in summative evaluation for rational decision or policy making is cost-benefit analysis. As suggested by Berki, “The fundamental effects flowing from the introduction of telemedicine will be changes in the processes of producing outputs as well as changes in the outputs themselves.”¹ Because of these changes, the alternative method, cost-effectiveness analysis, would be suboptimal because it would be limited to ascertaining the least costly alternative to achieve a given objective; cost-benefit analysis attempts to maximize the benefits for a given level of expenditure. Alas, cost-benefit analysis has the inherent limitation of converting all benefits and costs into monetary values, including such intangibles as the monetary value of duration of life, quality of life, and convenience. However, if its assumptions and metric are accepted, cost-benefit analysis can provide a rational basis for answering questions about critical choices among alternatives that may have varied and profound effects, including unanticipated effects. This topic is discussed in more detail in T. Reardon’s paper on economics in this issue.

Summative evaluation, as one might expect, is particularly important to policy and decision makers who are inherently interested in the ultimate benefits of programs to justify public expenditures and resource allocation decisions at the institutional, regional, national, and international levels.

STATUS OF TELEMEDICINE EVALUATION

The second generation of telemedicine activity, which began in the early 1990s in the United States and several other countries, received substantial funding. But, by virtue of the political process and perhaps by necessity, the bulk of the funding was allocated to program development rather than research. Of course, programs had to be planned, designed, organized, staffed, and launched. Infrastructure had to be installed, tested, and refined. Providers had to be recruited and trained, and patients had to consent and participate. And, finally operational procedures, protocols, and logistics had to be developed and routinized. In other words, the initial requirements for establishing telemedicine programs were and remain quite extensive and time consuming. Nonetheless, at the same time these programs were getting organized and launched, most funding agencies—as perhaps they should—required them to conduct an evaluation of their impact within the span of the funding cycle, often dealing with such complex issues as cost effectiveness and clinical outcomes.

The political considerations in federal funding for telemedicine programs in the United States may have inadvertently preempted the opportunity to design and execute large-scale, randomized controlled trials in this field. Purposely or not, we are left with a large number of programs widely distributed around the country, but none of them have the resources or long-term support needed for robust clinical trials. Moreover, the projects that attempted to collect uniform data sets across several sites have yet to publish results beyond descriptive trends. In this field, and perhaps others like it, there has always been a tension between research and development rather than a complementary relationship.

Despite, or perhaps due to, a history of several decades of dramatic change in the underlying technology, we have yet to agree on a uniform and precise definition of telemedicine. This has been exacerbated by a liberal trend in the imprecise use of such terminology as telehealth and e-health without a clear consensus on the delineation of boundaries between them

and telemedicine per se. A uniform agreement on the content and delineation of boundaries of telemedicine and other related terms would seem to be a prerequisite for valid evaluation of this field. Without such agreement, evaluation will remain largely project specific and, therefore, of limited generality.

Today, what we call telemedicine, telehealth, or e-health encompasses any existing or proposed configuration of technology, organization, and human resources. It also includes single or multiple clinical, educational, and public health applications. In this regard, no single or standard referent exists as to what telemedicine does or does not represent. From one perspective, this lack of a precise definition compromises our ability to evaluate the true concept in full “fidelity.”² Accordingly, whatever definitional imperfections exist in the configuration of a particular system under investigation will likely mask or mitigate the true effects of the telemedicine concept in its optimal form.

INTERPRETING CURRENT EMPIRICAL EVIDENCE

To start out, we present a synopsis of the conclusions reached by other investigators who have conducted literature reviews (a review of reviews) on the evaluation of this field. A number of published comprehensive and painstaking literature reviews of the work done in this field are available and are referenced later. The results of these reviews will be used as the logical foundation for developing alternative strategies for future investigation and analysis of evidence. Finally, we conclude with a discussion of options for future evaluative research and analysis of the available empirical evidence.

In brief, a literature review focusing on empirical research in telemedicine to date reveals that the research is not based on a precise and uniform set of parameters. Hence, this body of knowledge, although substantial in volume, is segmented and inconclusive. It is segmented in the sense that much of it focuses on specific applications rather than integrated systems of care related to patient care as well as provider and client education. Consequently, much of

this information has little to do with the systemic effects of telemedicine.

Furthermore, even well-designed experimental studies can be misleading. For example, diagnostic accuracy has been typically determined on the basis of observer agreement between in-person care (the gold standard) and care provided by telemedicine. As Koran³ pointed out, however, physicians tend to disagree with each other “once in ten cases, and often . . . one in five cases, whether they were eliciting physical signs, interpreting roentgenograms, electrocardiograms, or electroencephalograms, making a diagnosis, recommending treatment, or evaluating the quality of care.” Hence, reliability based on “agreement” should not be equated with accuracy.

It is interesting to note that, in an attempt to control interobserver variability, some researchers in this field have limited the number of observers to one or two who make two observations (telemedicine and in person) on the same subjects. This creates an even larger problem than the one it purports to resolve, namely the lack of independence in the data. Without independence, it is inappropriate to conduct any statistical analysis or to test any hypotheses.

After years and scores of attempts to develop and evaluate telemedicine initiatives and programs, the complaint continues to be found in the literature that the quality of evaluation studies is poor. As early as 1980, Bashshur⁴ lamented the fact that in the late 1960s and early 1970s, there had not been a broad-ranging, systematic effort to address all the basic and pertinent issues in the assessment of telemedicine and, further, “policy decisions regarding the funding of telemedicine . . . utilized a mixture of anecdotal and inconclusive empirical findings.” In the mid-1990s, the National Library of Medicine recognized a national need for robust telemedicine evaluation methodology and called for the Institute of Medicine to develop it.^{5,6} The report discussed general issues and suggested general guidelines but not an explicit framework and concrete guidelines for telemedicine evaluation.

DeChant *et al.*⁷ suggested a “staged approach” to deal with “the great variety in telemedicine applications and . . . new infor-

mation systems for health care delivery (that) pose(d) challenges to traditional methods of technology assessment.” A systematic review of over one thousand articles evaluating telemedicine cost-effectiveness revealed “only a few” controlled studies, and the vast majority did not produce the requisite scientific evidence.⁸ A review and critique of literature pertaining to evaluation of telehealth “solutions” called for consistency in the nature of evaluation activities “—if achievable—” and also suggested that this would certainly not occur in the short term.⁹ The DeChant report concluded that “most telehealth applications are either not well evaluated, or are evaluated in an ad hoc manner.” A review of clinical outcomes in telemedicine initiatives concluded that in only a small number of studies does evidence of its benefits exist, and there is need for further randomized controlled trials.¹⁰

The 2001 Report to Congress on Telemedicine indicated a lack of statistical evidence in most of the studies. It expressed “hope” for more statistically robust studies in the near future.¹¹ In 2002, an editorial in the *Journal of the American Medical Informatics Association* suggested that “the (telehealth) literature does not contain an adequate evaluation of telemedicine despite years of application of the technology and several calls to action.”¹² A literature review of telemedicine cost-effectiveness and patient satisfaction also concluded that, for the former, “there is no good evidence that telemedicine is [or is not] a cost effective means of delivering health care” (brackets added) and, for the latter, methodological deficiencies (small sample sizes, context, and study designs) limited the generality of the findings.^{13,14} Another review reported the scarcity of good-quality studies providing the evidence for the benefits of telemedicine.^{8,15} A review of clinical and educational telepsychiatric applications did not incorporate cost studies because the quality of data in the literature was “suboptimal and little . . . has been collected in a systematic, controlled prospective fashion.”¹⁶ And, “more short- and long-term quantitative and qualitative research is warranted on clinical outcomes, predictors of satisfaction, costs, and educational outcomes.” Drake raises the basic question about the future in his literature

review of telemedicine valuation in 2003.¹⁷ He concludes, "Even when mature evaluation techniques are extrapolated to telemedicine, after the studies are published, the most important question appears to remain: Is telemedicine technology worth advancing, and how as a society do we decide?"

These studies and conclusions represent recent systematic reviews of telemedicine evaluation studies. We may draw the general conclusion, therefore, that with few exceptions the research in this field has yet to produce an adequate body of empirical findings that rises to the level of conclusive evidence as traditionally defined. And, therefore, in a strict sense, we cannot ascertain with a reasonable degree of certainty the precise or specific effects of telemedicine on access, cost, and quality or the interaction between these effects, despite the importance of these issues from a public policy standpoint.

Nevertheless, the bulk of the research evidence to date has demonstrated the feasibility of telemedicine in almost all clinical and diagnostic applications. More than a decade ago, Grigsby and associates¹⁸ reported that teleradiology is "effective across all specialties." Indeed, the question of feasibility of telemedicine in all other clinical applications has been put to rest for quite some time, albeit at different levels of confidence for different applications. Interestingly, clinical applications have been classified by level of maturity on the basis of several performance measures and attributes, including volume and quality of research findings, demonstrable technical feasibility, diagnostic accuracy, specificity, clinical outcome, and cost effectiveness.¹⁹ According to this classification, teleradiology and telepathology occupy the first tier, and are labeled as "mature" applications. The second tier, or "maturing" applications, consists of telepsychiatry, teledermatology, telecardiology, and teleophthalmology. All other applications are in the third tier, and labeled as "emerging" applications. Of interest here is the fact that all research findings pertaining to clinical effectiveness across a variety of applications and at various levels of maturation are linked to the specific technology that was used in the application, and all technical failures and limitations are totally at-

tributable to specific sets of technologies that were used and have been improved since. Hence, no inherent limitations have been reported.

Finally, on careful reflection, several observations can be made about the evaluation literature.

- None of the systems or programs that were evaluated in the United States or other countries had fully exploited the technological capabilities at their disposal. Indeed, for a variety of reasons, all programs have substantially underused the capabilities of the available technology. Hence, any economic assessment that takes capital cost into account but does not incorporate sensitivity analysis under various assumptions of resource capacity, utilization volume, and level of maturation, learning curve, or steady-state operation is likely to underestimate the potential return on investment. Because capital cost is fixed, it is obvious that the more frequently and the more efficiently the program is used, the more favorable would be the return on investment. Moreover, given conditions of under-use, the logical question continues to be whether simpler and lower-cost technology (essentially, lower investment) could be more fully, and therefore, more cost-effectively used. But even this argument may have flaws because scale is important, and some minimal level of investment and operational size may be necessary to establish a viable program.
- In addition, nearly all telemedicine projects have missed significant opportunities to use the enormous capabilities of available information technology to create truly integrated health-care delivery systems. The underlying technology provides not only ready connectivity between various parts of the delivery system but, perhaps more importantly, ready access to information from various sources within and outside the institution for various components of patient care, protocol-driven disease management programs, and clinical decision support. The effective use of these rich capabilities and their integration into a single organization or system of care should reduce so-called errors, en-

hance efficiency, minimize unnecessary duplication and waste, and streamline the care process.

Nearly all projects had narrowly defined clinical or educational objectives or functions and specific target populations. This left unattended important questions pertaining to the full potential of the information technology at their disposal, similar to having a desktop computer, but using it only for electronic mail. While it makes perfect sense to use the computer for e-mail, one should not lose sight of the other numerous capabilities and the missed opportunities when they are not fully exploited. The ramifications for cost-benefit ratios are substantial.

- Given the observed limitations in the scope of the applications to date, it is not surprising that we have yet to assess adequately the broader systemic effects of telemedicine on use of service, referral and admission rates, access to care, quality, and cost of care in a way that allows true generalization about these effects. We have yet to achieve an in-depth understanding of the tradeoffs and interactions between these various effects, as discussed elsewhere in this paper.

EVALUATION RESEARCH: PROPOSED APPROACHES

The proposed approaches for valid evaluation of telemedicine are discussed here in terms of (1) prerequisites for optimal evaluation; (2) research design requirements; and (3) analytic tools for causal inference.

Prerequisites for optimal evaluation

Evaluation of telemedicine programs has been significantly hampered by the absence of optimally operating systems that have progressed sufficiently along the learning curve to permit a valid evaluation under optimal or at least steady-state conditions. Evaluating programs under these circumstances runs the risk of simply reflecting imperfections in design or implementation rather than inherent flaws in the basic concept or a necessary failure in the intrinsic capability of the intervention. Furthermore, even when a program is adminis-

tered in optimal fashion, one must separate the contextual effects from the experimental effects. To address this problem, Campbell²⁰ proposed the "climax model," which calls for evaluating programs only after they achieve a steady state of operation and contain all the intended core elements and capabilities of the program. The use of this model would assure that programs have sufficient fidelity in terms of strength, volume, and stability as prerequisites for assessing their true effects.

At the same time, it is clear that a telemedicine program constitutes an innovation "bundle" rather than a single categorical intervention. When the concept of "innovation bundle" was introduced by Rogers and Shoemaker,²¹ a precise definition was not offered. But it stands for a set of attributes, not all of which may be innovative. In telemedicine, for example, certain traditional clinical and communication practices are likely to be maintained and "mixed in" with the new mode of delivery. The broader medical care environment where these systems are installed is not likely to alter all established practices and routines simply by virtue of adding telemedicine to its repertoire of services. In other words, neither providers nor clients would alter all their traditional behaviors simply by virtue of having telemedicine available to them.

Nonetheless, an innovation bundle is expected to produce effects for the entire bundle that are likely to be incremental in nature. Similarly, the telemedicine bundle is likely to produce a variety of effects that are associated with various subsets of variables. These effects may represent tradeoffs rather than be unidirectional and additive. For example, telemedicine presents clear transportation/communication tradeoffs, which may increase access and enhance quality while reducing opportunity cost.

But we have yet to explain or understand what happens to cost when accessibility is improved. For example, would an improvement in access to care lead to a commensurate increase in demand, or even pent-up demand, and increase the risk of what economists call the "moral hazard," or the tendency to use medical services more because of decreased barriers to use? Would the availability of telemedicine lead to an increase in "provider-

induced" demand on the part of those seeking to enhance their revenues from the use of the technology? Would telemedicine initiatives displace the scarce medical resources available in a highly constrained rural environment, and thereby inadvertently diminish rather than enhance the available resources in those areas? Would telemedicine be likely to produce the paradoxical effect of promoting quality of care in under-served rural areas only to reduce the need for telemedicine over time? In other words, would its success lead to its ultimate demise? These and other important questions have yet to be addressed in definitive and comprehensive ways. Indeed, we have yet to understand whether success or failure in a specific clinical application or subsets of applications can be generalized to other applications or to the entire system of care. For instance, does success in teleradiology indicate success in other clinical applications or of entire programs that offer a variety of clinical services? Or, alternatively are telemedicine systems more than the sum of their parts?

Hence, the evaluation process has to be sufficiently refined and detailed to enable a valid analysis of the specific effects of the "newer" components as well as combinations of new and old ones, and it must be able to isolate telemedicine effects independently from contextual or other effects. This will not be easy, but it is particularly important in assessing telemedicine's effects on cost, quality, and access and the interactions among them.

While the calls for health care reform in the United States and elsewhere ebb and flow, reflecting the prevailing political climate, there is near-universal and omnipresent agreement on three related national goals: (1) Access to care should be improved, if not assured, to all people in need of care; (2) cost should be contained as much as possible without compromising people's health or safety, and, (3) quality should be maintained in all places, if not improved. With these goals in mind, it is understandable that telemedicine would remain on the agenda for health care reform for the foreseeable future.

Proponents have argued that telemedicine could accomplish all three goals effectively and efficiently, not only for rural and other

isolated populations, but also for the homebound and the institutionalized, as well as urban and poorly served populations. However, some critics have expressed concern that telemedicine may produce adverse effects by decreasing accessibility (by displacing rural providers); increasing total cost (by unleashing consumer pent-up demand and potential provider-induced demand); and compromising quality (by relying on a technology that may be a poor substitute for in-person care). Interestingly, when juxtaposed, the hypotheses of the advocates and the critics are enantiomorphic, that is, they mirror one another, thereby demonstrating their centrality and the need to subject them to rigorous scientific testing.

Research design

Two issues must be made clear in considering the choice of appropriate research designs for telemedicine program evaluation. First, the controlled experiment is the ideal method for evaluation research because it calls for the random assignment of test subjects to experimental and control groups, blind administration of the experimental variable, and pre- and post-intervention measurement. No doubt, the controlled experiment is the most powerful tool for achieving objective and valid results for policy making and program implementation. Weaker methods yield equivocal results that could be manipulated by program managers and decision makers. Hence, all things being equal, the controlled experiment is the preferred method for (1) testing explicit hypotheses, (2) ruling out rival hypotheses or other explanations for the observed phenomena, and (3) establishing cause-effect relationships between the specified variables. When one or more requisites of a controlled experiment cannot be obtained in a given situation, quasi-experimental designs and observational studies can be used, which ideally would mimic the controlled experiment to the extent feasible. This is true for simple medical interventions, such as new devices, new medications, or new procedures, as well as for medical programs, such as community health centers, regional medical programs, and, of course, telemedicine.

In brief, the essential feature of a controlled experiment or randomized controlled trial is

the active participation of the experimenter in randomizing subjects, administering the treatment or program intervention, and measuring differences between the experimental and control groups before and after the intervention. As we discuss later, randomized controlled trials may not be feasible in this field.

Second, the absence of knowledge or methodological sophistication is not hampering the progress of research in this field. Such knowledge exists, and the requisite methodological tools are in hand. We do not need to invent new methodology. For example, in complex evaluations, factorial designs enable the detection of differential effects of various combinations of variables, including characteristics of individuals using the service (old versus young, male versus female); various levels in severity of illness, and provider mix (generalist versus specialist). Subsequently, the design enables researchers to concentrate on features that either singly or in combination produce the greatest effects. In the real world, however, it is impracticable and, in some situations, impossible to conduct controlled experiments to evaluate telemedicine programs as total entities. Among other things, the requirements of scientific inquiry on the one hand and the political, legal, and ethical imperatives on the other are not consistent, as discussed earlier. Laws and regulations governing individual autonomy through informed consent and the proscription of harmful interventions are appropriate in a civilized society, although some of these regulations can create formidable difficulties in research.

Human subjects have the right to refuse participation in any study that observes them directly, collects information about them, or uses personal information already available. Furthermore, they have the right to withdraw their consent at any time during or after the conclusion of a study, and, in certain instances, to jump experimental allocation. And when the intervention may be deemed beneficial, it may be unethical to withhold it from those in need. It is not uncommon for researchers to have to nullify or diminish the effects of the experimental variable by promising remediation for ill effects as a result of participating in the study. For example, in the RAND Insurance

Study, participants received remuneration for higher out-of-pocket cost that resulted from their assignment to health insurance plans with a high co-payment.²²

It is important to add that these issues (withdrawal of patients, patient crossover, and the ethical issues when the interventions are found to be harmful or beneficial before the planned end of the study) occur frequently in clinical trials. Techniques such as application of the "intention-to-treat" principle, survival analysis techniques, and independent study safety committees are available.²³

These issues aside, random assignment of patients to experimental and control groups is not possible when the entire population of telemedicine users is extremely limited. Indeed, small sample size is one of the major obstacles in conducting randomized controlled trials as well as sample surveys to evaluate the impact of telemedicine programs. Further, situational or contextual factors cannot be held constant to rule out rival hypotheses given the empirical diversity of telemedicine programs produced by local variations in administration, variations in clinical applications and in organizational and staffing configurations, variations in the clinical competence of providers, and the continuing evolution in the underlying technology.

There are several other impediments to the use of experimental designs in telemedicine research. First, we have yet to agree on the precise scope of the program to be evaluated and to differentiate between the overall goals of the program and those of the specific applications within it. Telemedicine programs can range from single specialty service within a single delivery system, such as teleradiology, to a multispecialty, multisite network that offers a full range of medical and diagnostic services. Despite their intuitive appeal, full-service programs are difficult to implement. Second, as mentioned before, the prevailing laws and regulations covering the use of human subjects in research often challenge randomization rules, which blindly assigns subjects to an experimental or control group, do not permit subjects to "jump" or transfer from one experimental allocation to another, and do not allow administrators or project managers to make midcourse

changes in the content of the program—especially in terms of key variables that may alter the outcomes. Such changes include technological configurations (in terms of hardware and software), human resources and staffing, and the organizational structure of the program.

Beyond these serious difficulties, the evaluation of telemedicine faces some unique problems in terms of the nature of the experimental variable, or the unit of analysis. The usual units of analysis, as well as units of observation, that have been reported in the telemedicine literature typically consist of individual medical visits or encounters. The more appropriate units of analysis in telemedicine research may consist of episodes of care or illness episodes to be used in comparing telemedicine with in-person care. Indeed, more useful and valid information about the effects of telemedicine on utilization and cost would be captured when using episodes of care or episodes of illness rather than single visits or encounters. This is because episodes can capture more information for assessing the contributions of telemedicine. Additionally, episodes can be compared between the two modalities of care, telemedicine and in-person, while controlling for severity of illness. They can provide comparative information on the number and duration of visits, intensity of care, substitutions (between sites and providers), timeliness of care, waiting time, and outcomes. Their use would enable us to ascertain whether or not telemedicine has a significant effect on the process and outcome of care. For example, among other things, it would reveal whether or not the use of telemedicine resulted in more timely diagnosis, fewer visits, fewer and shorter hospitalization, shorter illness episodes, and ultimately a more successful health outcome.

As many observers have pointed out, the telemedicine concept continues to evolve in scope, application, and underlying technology. Hence, it is futile to evaluate it as a fixed entity. It can represent different things to different people at different times and in different settings. Nonetheless, for policy purposes, evaluation may have to focus on the general attributes of the practice of telemedicine, if such parameters exist, or on telemedicine as a whole

if it is considered a single entity, to assess its overall merit and how it fits into the health system.

The gains or losses that might accrue from telemedicine are likely to vary by the intended target or user, the nature of the application, and the specific technological configuration that is used. Moreover, the benefits and costs of telemedicine should be viewed from the varied perspectives of providers, clients, and society at large. Each perspective has concerns that must be identified and addressed in the design of the study. The specific applications also must be taken into account, such as clinical, educational, and managerial programs. Finally, the study must consider the specific technological design and configuration in use, for instance, whether it is a synchronous or asynchronous connection, bandwidth, and peripheral devices. This multifaceted approach can be represented by a three-dimensional model adapted from Donabedian,²⁴ as shown in Figure 1.

The first dimension in the model is made of the various applications that encompass the different specialty areas of medical practice as well as medical education and public health. If we take the nomenclature and purview of telehealth or e-health seriously, we must include all areas of public health, including disease surveillance, epidemiology, health education and health behavior, and health administration.

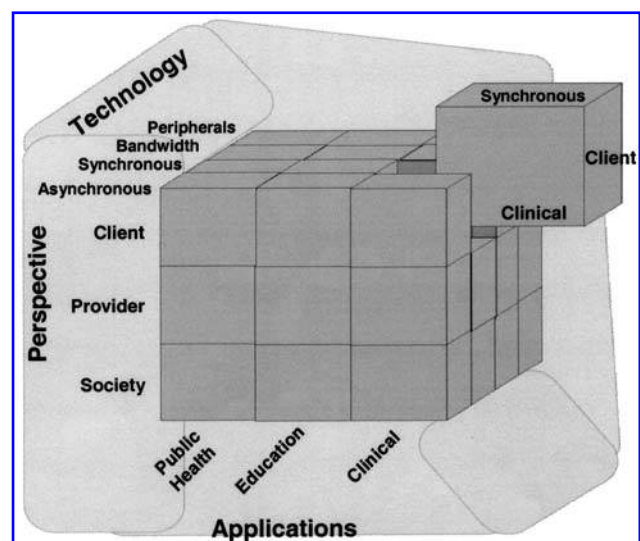


FIG. 1. Three-dimensional model for telemedicine evaluation.

The second dimension has to do with the specific perspectives from which an evaluation can be conducted, specifically those of clients, providers, and society at large. Obviously, despite some commonalities among them, the concerns of clients (or consumers) might differ from those of providers or of society at large. Each perspective must be measured separately. For example, on the question of cost, clients may be interested in out-of-pocket cost, whereas providers may be interested in revenue generation and return on investment. Societal interest may relate to total cost, equitable distribution of health-care resources, and meeting unmet needs of the population. Thus, even when concerned with the same basic issues, providers, clients, and society may approach them differently because their stakes are different.

The third dimension represents various technological systems and configurations in use, including synchronous online systems, asynchronous store-and-forward systems, various modalities of transmission (wired, wireless), bandwidth, and peripheral devices for diagnosis and treatment.

This model provides a systematic taxonomy or framework for interpreting and integrating research findings from different empirical studies, while leaving researchers to pursue their particular interests in addressing a variety of research questions in their respective domains. Ultimately, the findings from individual studies must be combined in a coherent set. This proposed model allows for such a compilation of the empirical evidence in a flexible and ongoing manner while at the same time providing the requisite structure for accumulating evidence from various studies performed by different investigators at various times. The distribution of "completed" compartments can also serve to identify research gaps that need to be filled.

The individual compartments or cubes within the three-dimensional matrix can be viewed singly or in various combinations. For instance, one researcher may choose to investigate the quality of diagnostic images and their cost-effectiveness in teleradiology, using store-and-forward technology. Of course, this would be from a provider perspective. Another researcher may assess the systemic effects of a

telemedicine program on access, cost, and quality of care from a societal perspective. The "extracted" cube in Figure 1 represents a clinical application from the client's perspective using synchronous technology. For example, this research might involve a remote diabetic patient sending in blood sugar readings to an endocrinologist and receiving information on what type of meals to prepare. In another scenario, the patient would submit blood sugar readings, exercise regimen, and other related information to a Web site that would respond with the appropriate dietary information.

The findings from each study or set of studies on the same topic may be used to fill in one compartment or cube in the model, which can be subsequently added to findings from other studies for a cumulative total that combines all other studies in the other compartments until the picture is complete. In the end, this cumulative process would begin to answer some of the pressing questions in the field. It would also serve to identify gaps in research. However, this would not be a simple or straightforward process, since a meaningful matrix of findings has to be based on compatible studies that meet common methodological standards.

Typology of research questions

As illustrated in Table 1, there are essentially two types of research questions that are appropriate for telemedicine evaluation. The first derives from the biomedical or clinical field, and it encompasses issues of clinical effectiveness, efficacy, and safety. More specifically, biomedical research seeks to ascertain the accuracy, precision, reliability, and sensitivity as well as safety (or side effects) of specific technological components in providing diagnostic or treatment information. Insofar as telemedicine is concerned, the basic goal of biomedical research is to determine the extent to which specified technological components meet or exceed clinical standards of performance as compared with in-person observation and measurement. The second type or set of research questions derives from health services research. Here, the questions are typically posed after the conclusion of biomedical or clinical research, and the technology is certified as safe

TABLE 1. TYPES OF RESEARCH IN TELEMEDICINE, RESEARCH QUESTIONS, AND MEASURES

	<i>Research questions</i>	<i>Measures</i>
Biomedical research	Safety, effectiveness	Precision, accuracy specificity, reliability
Health services Health care	Access, quality, cost	Utilization, referral, convenience, opportunity cost, etc.

and effective. Indeed, in a perfectly logical world, health services research would begin when biomedical research ends.

Whereas biomedical research is indispensable and must be continued, it seems to be faced with overwhelming odds against proving anything definitive on a permanent basis, in relation to the clinical effectiveness of telemedicine technology. This is due to continued technological developments that reduce or eliminate observed imperfections in the technology, as the field continues in its press to find effective substitutes for the five senses. The technological horizon keeps expanding. In turn, these technological improvements have a direct impact on clinical effectiveness as well as on cost-benefit ratios, and the tradeoff between cost and quality. In brief, technological advances tend to obviate limitations in diagnostic or therapeutic effectiveness attributed to specific technologies.

Questions and hypotheses stemming from both biomedical and health services research can be readily translated into specific measures to be used in data collection and subsequent analysis. Some of these measures are simple and straightforward and some are not. All of them require validity and reliability testing to ensure that they actually measure what they purport to measure in a consistent fashion. Telemedicine literature already contains several standardized measures borrowed from social science research, which may be used by researchers in this field.

Telemedicine evaluation: revised strategy

Other things being equal, the strategy of choice in telemedicine evaluation would consist of a large-scale research program that uses

rigorous scientific methods to investigate the various dimensions of telemedicine as well as the interactions between them. However, this is not likely now or in the foreseeable future. This situation leads us to offer an alternative that can be pursued immediately; namely, to ascertain in a systematic manner the weight of the available evidence and new findings as they become available.

In this section, following a brief statement pertaining to the contextual framework of telemedicine development and its expected role in improving access to care, we focus first on the notion of cause-and-effect and its measurement. Finally, we propose and discuss a new strategy that might be appropriate for assessing the impact of telemedicine on access, quality, and cost of health care. But, for illustrative purposes and because issues of cost and quality are considered in separate papers, we limit our consideration here to the association between telemedicine and access to care.

CONTEXTUAL BACKGROUND

Concern for the inequitable geographic distribution of medical care was expressed as early as the mid-19th century when the (then) nascent American Medical Association charged state medical societies to assess the geographic availability of physician care. Several states reported that physicians were fewer and less likely to have adequate medical training in rural areas when compared with those in urban areas. Development and diffusion of advanced technologies, such as the automobile and telephone in the late 19th and early 20th centuries, reshaped the argument regarding ge-

ographic access to care. But, the emergence of specialized medicine and office/clinic/hospital-based medical care in the mid-20th century exacerbated the geographical disparity in the distribution of medical-care resources between rural and urban areas, as well as between developed and developing countries.

Historically in the United States, attempts to redress the inequitable geographic distribution of medical care, such as the Sears Roebuck Foundation's Community Medical Assistance Program,²⁵ the Hospital Survey and Construction (Hill-Burton) Act,²⁶ and the National Health Service Corps created as part of the U.S. Public Health Service under the Emergency Health Personnel Act,²⁷ were among the leading initiatives aimed at redressing the prevailing geographic disparities in availability of health-care resources. But all were met with mixed results and generally limited success, especially as they pertain to specialty care. Thus, the excitement that met the introduction of telemedicine initially in the early 1970s and now was based in large part on past failures to redistribute medical-care resources physically and the emerging virtual distribution that would provide under-served populations with ready access to medical care when needed. This virtual distribution of health resources is expected to diminish opportunity cost for clients or patients, and it should enhance quality of care and hence improve health status. Other factors contributing to the excitement include the potential for creating integrated medical systems and reduction in so-called medical errors.

CAUSAL INFERENCE

Any evaluation of the impact of telemedicine on access implicitly or explicitly assumes a cause-and-effect relationship. For philosophers, the notion of cause and effect is murky at best. Hume and others argued, for example, that we can never know with certainty whether two events are causally related. There is simply a constant temporal conjunction observed repeatedly, and, "therefore what we term 'causality' can be nothing other than mere constant conjunction of the idea of the cause with that

of the effect," and this is known from experience.²⁸ But we are living in the real world, not that of philosophers, and we need to know to our satisfaction whether there is a causal association between telemedicine and access and, more specifically, the direction and magnitude of that association.

Here, we may be informed by a similar, perhaps simpler, situation that occurred in the development of epidemiology; namely, the evolution of criteria for "causality" between a disease agent (virus, bacterium, parasite, etc.) and a disease. Initially, following the discoveries of Pasteur in the late 19th century, the so-called Henle-Koch "postulates" were used. These postulates stated that an agent is a cause of a disease if it is present in all the affected persons ("necessary" cause), it is absent in healthy subjects ("sufficient" cause), and it can be inoculated into an animal to induce the same diseases that it causes in humans. However, implementation of these criteria in epidemiology proved impracticable, if not impossible.

Later, Sir Austin Bradford Hill²⁹ proposed nine criteria of causation as minimal conditions to establish a causal relationship between two events. Nonetheless, he cautioned that "none of these nine viewpoints can bring indisputable evidence for or against a cause and effect hypothesis. . . . What they can do, with greater or less strength, is to help answer the fundamental question—is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?" He even went so far as to suggest that none can be considered as a *sine qua non* of causality." However, Rothman³⁰ points out if the "cause" does not precede the effect, that indeed the temporality criterion is indisputable evidence that the association is not causal.

While the criteria established by Hill (and elaborated by others) were developed as a research tool in epidemiology, they may be equally applicable in other social sciences seeking to establish causal relationships among social phenomena, especially when it is not feasible to conduct prospective large-scale, randomized controlled trials. It must be remembered, however, that Hill's criteria and, historically, discussions on causality, have proceeded

once a statistically significant relationship between a potential causal factor and a disease has been found. In at least one instance, this threshold was ignored for apparent political purposes by a government agency. This led to considerable criticism from the scientific community on the role of environmental carcinogens in a “cancer epidemic.”³¹

Armed with these criteria, epidemiologists have been able to investigate suspected single- and multifactor causal relationships and to advise the public and private health sectors on health policy decisions. Essentially, the process is one of “triangulation,” or the application and combination of several research methodologies in the study of the same phenomenon.

By combining multiple sources of data, theories, methods, and empirical materials, triangulation may overcome the weakness of research design or intrinsic biases that come from small studies with inconclusive findings. Researchers are able to triangulate, or draw significant conclusions in terms of the convergence of findings, from various studies and methodological approaches to the same topic. But, even here certain criteria have to be met.

Addressing the field of human–computer interaction, Mackay³² suggests that, to mature as a field, there must be a shared research context and the need for triangulation: “using different techniques to operationalize behavior while attempting to measure the same phenomenon . . . which greatly increase the generality and construct validity of the findings.” A study of air-traffic controllers, for example, included biological analysis of sleep patterns, laboratory experiments of different user interface strategies, computer simulations in the use of new tools by controllers, cognitive models of air traffic control’s activities, and ethnographic studies of controllers at work.³³

Triangulating among various forms and sources of data means that if an observation holds across a variety of contexts viewed from different research perspectives, then we are more confident of the integrity of the finding. It has also been suggested that through the triangulation of qualitative and quantitative data, a “more complete, holistic, and contextual portrayal of the research issue is captured.”^{34–36} A

number of triangulation types have been identified:

- Data triangulation: The use of multiple data sources to help understand a phenomenon.
- Methods triangulation: The use of multiple research methods to study a phenomenon.
- Investigator triangulation: The use of multiple investigators in collecting, analyzing, and interpreting the data.
- Theory triangulation: The use of multiple theories and perspectives to help interpret and explain the data.

Although one might be tempted to downplay or even disregard a conclusion from a single data source, a conclusion derived from triangulation from multiple sources is more credible and can be valuable and meaningful. Triangulation of information can allow even statistically limited or weak conclusions drawn from one source to be compared with conclusions drawn from other sources. In the following section, we attempt to demonstrate the use of theoretical triangulation as it might apply to assessing the impact of telemedicine on access.

EVALUATING TELEMEDICINE AND ACCESS: A NEW STRATEGY

Generally, access refers to the ability of patients to use appropriate health resources in a timely manner. The concept summarizes a set of specific dimensions that describe the “fit” between a patient and the health-care system.³⁷ The Institute of Medicine⁵ suggested that access can be enhanced by increased availability of health information.

The realization of the need and appeals for scientific evaluation of the impact of telemedicine on access is longstanding. Nevertheless, the appraisal of telemedicine access-evaluation reveals few studies that provide definitive conclusions, as is the case for evaluating cost, quality, and other aspects to telemedicine. In fact, valid reasons have been offered for the seeming inability to evaluate the association between telemedicine and access adequately. These include the lack of mature telemedicine programs due largely to the recency of the in-

novation, as well as limited funding duration in some instances. Also complicating matters is an inherent lack of stability and the changing nature of telemedicine, reflecting the inclusion of innovative technology as it becomes available and applications in new areas. Additionally, to date there have been few if any large-scale telemedicine experiments specifically designed and funded to conduct randomized controlled trials to determine the various effects of telemedicine on cost, quality, and access and interactions among these effects.

Given these conditions, telemedicine may not lend itself to traditional methods for demonstrating cause and effect. Consequently, it is time to re-evaluate the concept and strategy of evaluation in this field.

A triangulation process may be the appropriate tool for interpreting the extant evidence. The "triangulation process" refers to conclusions from a number of sources within each criterion of causal inference. In turn, the total results from these triangulations produce a valid basis for reaching closure on the intended effect. The components of the proposed triangulation process are illustrated in Figure 2.

The criteria in Figure 2, also listed below for discussion, are adopted from epidemiology

and adapted for the purpose of assessing telemedicine.

Temporality

This criterion may imply a concomitant "temporal" relationship with the duration of the telemedicine program; that is, the longer the program has been in existence, the greater will be the observed impact.

Gradient

This refers to the increase or decrease in access for persons using telemedicine versus those who use in-person care for the same conditions and the same severity of illness. Alternate measures include various aspects of opportunity cost, including travel costs, travel time and distance, lost wages, and convenience. It implies that: (1) the increase in access to medical care, however measured, follows the introduction/availability of telemedicine and is not the result of simultaneous reduction in travel costs through improvement in transportation or other reductions in opportunity costs; and (2) the "stronger" the telemedicine program (reputation of consulting center, diffusion of information regarding program, acceptance by physicians and administrators), the greater will be the impact on access.

Consistency

This pertains to the degree of consensus on the impact of telemedicine on access obtained from a wide variety of people, places, settings, and applications. With specific reference to access, it includes a degree of agreement related to our currently accepted understanding of distance-interaction processes. Hence, the rules of evidence are summarized in terms of empirical findings, the sequence of variables, the strength of the association, and the consistence with other studies. It has been suggested that when an application is effective consistently across a representative set of indicators, it is not necessary to evaluate all indicators.⁶ This same notion may apply to the study of access in that the weight of findings of increased access for a wide variety of applications in diverse geographical and institutional settings can provide

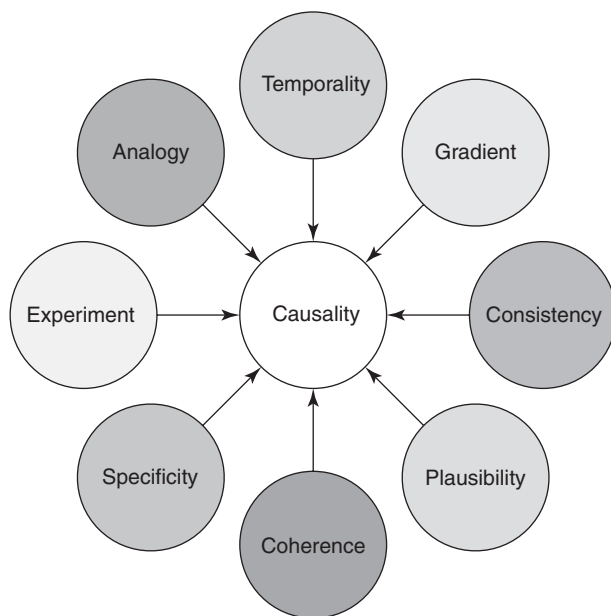


FIG. 2. Causality in telemedicine.

reliable evidence of the effects of telemedicine on access.

Plausibility

With regard to access, this notion is reflected in human behavior by the concept of the distance decay function (Fig. 3). Simply put, the function describes a decreasing probability of interaction with a place as the distance from that place increases. The use of a distance decay formulation helps identify potential patterns of service utilization by people. There is considerable literature, especially in the mental health field, where it is known as Jarvis' Law, that supports this proposition. Hence, we are on solid ground in terms of both the Principle of Least Action and the empirically defined law of distance interaction function as a basis for implementing telemedicine.

There is a considerable literature to support the notion that increased distance to medical care is associated with decreased utilization.³⁸⁻⁴¹ Furthermore, other things being equal, distance can be the determining factor as to when and if a person takes advantage of a medical care opportunity.^{39,42} These empirical observations support the notion of plausibility and directly reflect the basic principles and laws pertaining to human movement behavior.

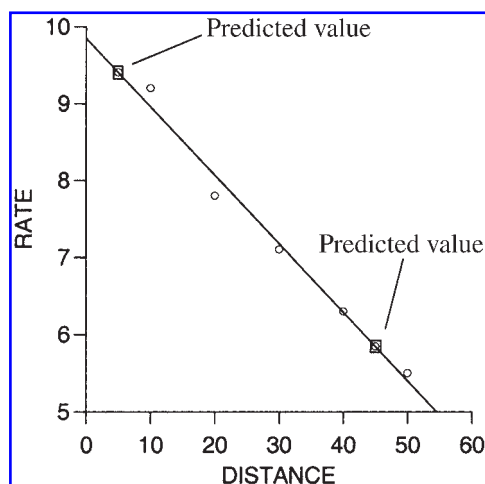


FIG. 3. Distance decay function.

Coherence

In the telemedicine-access equation, this means that the increase in accessibility is compatible with existing theory and knowledge. It presupposes a theoretical basis for the assumed relationship between telemedicine and increased access, which is derived from the ecology and ethology of human behavior.

Specificity

Ideally, the effect has only one cause. An outcome predicted by one primary factor adds credibility to a causal claim—a single cause is related to single effect. This would be attained when we are able to demonstrate that a single aspect of telemedicine is associated with a single effect on access.

Experiment

The demonstration that, under controlled conditions, changing the exposure causes a change in the outcome—in this case, the conditions suggested earlier pertaining to the need for large-scale, clinical trials.

Analogy

Typically, we are more willing to accept arguments that resemble others we have already accepted. In other words, a commonly accepted phenomenon in one area can be applied to another area. For telemedicine and access, this would again refer to the voluminous literature that supports the notion that placing a medical facility closer to its target population increases utilization by that population.

We have proposed here a revised strategy for evaluating the effects of telemedicine based on the available and forthcoming evidence, with special emphasis on the association between telemedicine and access to care, as an illustration on the use of this strategy. The impetus for this effort derives from past, present, and, most likely, future difficulties encountered in attempting to evaluate the multidimensional and continually evolving technology and applications of telemedicine. Moreover, it is important to note that telemedicine is not the only type of program that is difficult to evaluate using the controlled experiment. Suggestions pertaining

to the need for looking toward triangulation as one means of not only resolving the problem but also providing more feasible, holistic, and “surfer” analyses are found in other fields as well. We have argued here that triangulation may be a viable alternative evaluation strategy for telemedicine. Additionally, we have presented a preliminary and, perhaps, rather simplistic first attempt to illustrate the concept and how it might be implemented. To be sure, questions pertaining to the operationalization of the terms and development of appropriate measures remain to be answered. One central problem, should the strategy prove viable, is a mechanism to collect data from various sources and a central repository for results from the various types of research methodologies, both quantitative and qualitative, that will form the basis for implementing this strategy.

SUMMARY AND CONCLUSION

Despite several decades of growth and deployment of telemedicine programs, a review of the literature illustrates that, to date, the majority of evaluation studies does not rise to the level of producing definitive results on the benefits and costs of this field. Hence, claims of telemedicine program efficacy pertaining to improved access, equal or enhanced quality compared with traditional medical care, and reduced costs cannot be made with strong assurance. To review, reasons for this problem are many, including:

- A failure to use a precise and uniform definition of “telemedicine,” and the multidimensionality of the innovation bundle of telemedicine.
- The continuous improvement of the underlying technology of telemedicine, and the expansion of applications.
- Experimental problems, such as:
 - Lack of clarity in specifying the experimental variables;
 - Difficulty of experimental allocation and blinding and jumping experimental allocation;
 - Limited fidelity in program implementation, inadequate program maturity, and steady-state operation;
- Multiplicity of program affects, delayed effects, and unintended effects; and
- Lack of large-scale programs that would permit experimental studies.
- Insufficient funding for large-scale experimental studies.
- Failure to exploit the full potential of telemedicine technologies.

The inadequacies of scientific rigor aside, the overwhelming majority of published evaluations of telemedicine programs support the notion that telemedicine does improve access. It is a valid substitute for in-person care in many applications, and does, or has the potential to, reduce opportunity costs. And, in spite of the lack of substantive evaluation research, telemedicine programs continue to proliferate, perhaps not to the extent that proponents feel is appropriate, given the presumed potential. Proponents continue to call for support, while the field is criticized for not having definitive research evidence to justify its full adoption by mainstream medicine as an integral component of the health care armamentarium.

It appears that telemedicine program evaluation is faced with a quandary. Should we continue to pursue what appears to be a futile search for definitive results from small, limited studies that are scientifically rigorous? Or should we devise a strategy or strategies for assessing all the available evidence in systematic, coherent, and consistent fashion through triangulation? Should we wait for the credible evidence to emerge or capitalize on existing findings and still pursue the best feasible approach in such research?

In this paper, we have presented two strategies, which are not mutually exclusive. The first strategy is to fund large-scale experimental telemedicine programs and projects that can be designed and implemented to collect data sufficient to test specific dimensions and effects of the technology. Data from such studies can then be assessed with statistical confidence to draw probable and credible conclusions. We propose a three-dimensional model for evaluating telemedicine on the basis of the varied perspectives of the client, provider, and society at large, the specific clinical or other application, and the technological configuration. Moreover,

as findings are reported, results can be accumulated, using this three-dimensional telemedicine data matrix, to permit the triangulation of results and the translation of results into a summative research format. In turn, the cumulative research findings would form the basis for reaching conclusions about the benefits and costs of telemedicine.

We also propose and illustrated a second alternative using theoretical triangulation as one basis for assessing the impact of telemedicine on access to care. This strategy derives from the fact that in reality our search for the holy grail of scientifically evaluating telemedicine programs in the traditional manner must be revised. The new strategy is based on established theory together with the cumulative data from research studies that may be based on imperfect designs. The criteria for judging the results include: statistical significance, temporality, strength, consistency, coherence, plausibility, analogy, and experiment. We must integrate results from both quantitative and qualitative research designs.

REFERENCES

- Berki, S. Telemedicine, some economic implications. In Bashshur R, Armstrong P, Youssef Z, eds. *Telemedicine, Explorations in the use of telecommunications in health care*. Springfield, IL: Charles Thomas, 1975: 175–191, and reproduced courtesy of Charles C Thomas, Ltd., Springfield, Illinois.
- Sechrest LS, West G, Phillips MA, Redners R, Yeaton WH. Some neglected problems in evaluation research: strengths and integrity of treatments. In: Sechrest L et al. *Evaluation studies review annual*, Vol. 4, Beverly Hills, CA, 1979.
- Koran LM. The reliability of clinical methods, data and judgments. *N Engl J Med* 1975;293:695–701.
- Bashshur RL. A proposed model for evaluating telemedicine. In: Parker L, Olgren C, eds. *Teleconferencing and Interactive Media*. Madison, WI: University of Wisconsin-Extension Center for Interactive Programs, 1980;211–219.
- Institute of Medicine. *Telemedicine: a guide to assessing telecommunications for health care*. Washington DC: The National Academies Press, 1996. (Also available online at: <http://www.nap.edu/books/0309055319/html/index.html>).
- The Lewin Group. *Assessment of Approaches to Evaluating Telemedicine Final Report*. Prepared for Office of the Assistant Secretary for Planning and Evaluation, Department of Health and Human Services, 2000. (Available on-line at: <http://aspe.hhs.gov/health/reports/AAET/aaet.htm>).
- DeChant H, Tohme W, Mun S, et al. Health systems evaluation of telemedicine: a staged approach. *Telemed J* 1996;2:303–312.
- Roine R, Ohinmaa A, Hailey D. Assessing telemedicine: a systematic review of the literature. *Canad Med Assn J* 2001;165:765–771.
- Health Canada. *Evaluating Telehealth 'Solutions' A Review and Synthesis of the Telehealth Evaluation Literature*. Office of Health and the Information Highway, 2000. (Available at http://www.hc-sc.gc.ca/-bsi/pubs/2000_tele/tele_e.html).
- Hersh W, Helfand M, Wallace J, et al. Clinical outcomes resulting from telemedicine interventions: a systematic review. *BMC Medical Informatics and Decision Making* 2001;1:5 (Available at <http://www.biomedcentral.com/1472-6947/1/5>).
- U.S. Congress. 2001 Report to Congress on Telemedicine: Evaluation and Research. Office for the Advancement of Telehealth Publications, 2001. (Available at <http://telehealth.hrsa.gov/pubs/report2001/eval.htm>).
- Stead W. In: Hersh W, Patterson P, Kraemer D. Telehealth: The Need for Evaluation Redux. *J Am Med Informatics Assn* 2002;9:89–91.
- Whitten P, Mair F, Haycox A, et al. Systematic review of cost effectiveness of telemedicine interventions. *Br Med J* 2002;324:1434–1437.
- Mair F, Whitten P. Systematic review of studies of patient satisfaction with telemedicine. *Br Med J* 2000;329: 1517–1520.
- Hailey D, Roine R, Ohinmaa A. Systematic review of evidence for the benefits of telemedicine. *J Telemed Telecare* 2002;8(Suppl. 1):S1:1–7.
- Hilty D, Marks S, Urness D, Yellowlees P, Nesbitt T. Clinical and educational telepsychiatry applications: a review. *Canad J Psychiatry* 2004;49:12–23.
- Drake DE. Evaluating telemedicine: A literature review. *eHealth Int* 2003.
- Grigsby J, Kaehny M, Schlenker R, et al. *Analysis of Expansion of Access to Care Through Use of Telemedicine and Mobile Health Services Report 1: Literature Review and Analytic Framework*. Denver Center for Health Policy Research, 1993.
- Krupinski E, Nypaver M, Poropatich R, et al. Clinical applications in telemedicine/telehealth. In: Bashshur R, Mandil S, Shannon G, eds. *State of the Art :Telemedicine/Telehealth* 2002;8:13–34.
- Campbell D. Quasi-experimental design. In: Samuel Overman E, ed. *Methodology and Epistemology for Social Sciences: Selected Papers* Chicago: University of Chicago Press, 1988:168–190.
- Rogers E, Shoemaker F. *Communication of Innovations: A Cross-Cultural Approach*. New York: Free Press, 1971.
- Welch B, Hay U, Miller D, et al. The Rand Health Insurance Study: a summary critique. *Med Care* 1987; 25:148–156.
- Gordis L. Randomized trials: some further issues. In: *Epidemiology*, 2nd ed. Saunders: Philadelphia, 2000: 110–128.

24. Donabedian A. *Explorations in Quality Assessment and Monitoring: Volume I: The Definition of Quality and Approaches to its Assessment* Ann Arbor, Michigan: Health Administration Press, 1980.
25. Kane R, Warnick R, Proctor PH, Olsen DM, Gourley D. Mail-order medicine; an analysis of the Sears Roebuck Foundation's Community Medical Assistance Program. *J Am Med Assn* 1975;232:1023–1027.
26. Legislative History on the Passage of the Hospital Survey and Construction Act, Public Law 725, 79th Congress, July 1948. (Available at http://rmp.nlm.nih.gov/RM/A/A/E/H?_rmaah.pdf).
27. U.S. Senate. Senate Report 104-093-*Health Professions Education Consolidation and Reauthorization Act of 1995*. Section II Background and Need for Legislation, 1999.
28. Fowler T. The formality of reality: Xavier Zubiri's critique of Hume's analysis of causality. *Xavier Zubiri Rev* 1998;1:57–66.
29. Hill BA. The environment and disease: association or causation? *Proc Soc Med* 1965;58:295–300.
30. Rothman K. Causation and causal inference. In: Schottenfeld D, Fraumeni J, eds. *Cancer Epidemiology and Prevention*, 2nd edition. New York: Oxford University Press. 1996:3–11.
31. Milloy S, Gough M. The Environmental cancer epidemic that never was. *Regulation: The CATO Review of Business and Government*, 2004. (Available at <http://www.cato.org/pubs/regulation/reg19n2g.html>).
32. Mackay W. Triangulation within and across HCI disciplines, 2004. (Available at: <http://www-ihm.lri.fr/~mackay/pdf/HCI98.Commentary.pdf>).
33. Mackay W, Fayard A. HCI, natural science and design: a framework for triangulation across disciplines. In: *Proc Design Interact Syst DIS '97*. ACM: Amsterdam, The Netherlands, 1997.
34. Jick T. Mixing qualitative and quantitative methods: triangulation in action. *Admin Sci Q* 1979;74:602–611.
35. Mitchell E. Multiple triangulation: a methodology for nursing science. *Adv Nurs Sci* 1986;8:18–26.
36. Handfield R, Ragatz G, Monczka R. A triangulated approach to empirical research: the global procurement and supply chain benchmarking initiative. *Decision Line* 1995;26:3.
37. Donabedian A. Accessibility. In: *Aspects of Medical Care in Organization* Cambridge: Harvard University Press, 1973:419–436.
38. Guagliardo M. Spatial accessibility of primary care: concepts, methods and challenges. *Intl J Health Geographics* 2004;3:3. (Available at <http://www.ij-healthgeographics.com/content/3/1/3>).
39. Gregory P, Malka E, Kostis J, et al. Impact of geographic proximity to cardiac revascularization services and service utilization. *Med Care* 2000;38:45–57.
40. Fortney J, Rost K, Zhang M, Warren J. The impact of geographic accessibility on the intensity and quality of depression treatment. *Med Care* 1999;37:884–893.
41. Shannon G, Bashshur R, Metzner C. The concept of distance as a factor in accessibility and utilization of health care. *Med Care Rev* 1969;26:143–161.
42. Shannon G, Bashshur R, Lovett J. Distance and the utilization of mental health services. *Milbank Memorial Fund Quarterly Health and Society* 1986;64:302–330.

Send reprint requests to:

Rashid L. Bashshur, Ph.D.

Director of Telemedicine

The University of Michigan Health System

8B03 N1B, Box 0402

300 North Ingalls

Ann Arbor, MI 48109-0402

E-mail: bashshur@med.umich.edu

This article has been cited by:

1. Micaela Brown , Nicola Shaw . 2008. Evaluation Practices of a Major Canadian Telehealth Provider: Lessons and Future Directions for the Field. *Telemicine and e-Health* 14:8, 769-774. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
2. N. Eminović, N.F. de Keizer, P.J.E. Bindels, A. Hasman. 2007. Maturity of teledermatology evaluation research: a systematic literature review. *British Journal of Dermatology* 156:3, 412-419. [[CrossRef](#)]