Assessing the Validity of Statistical Inferences in Public Health Research: An Evidence-Based, 'Best-Practices' Approach

Recommended Citation

Peace, Karl E.; Parrillo, Anthony V.; and Hardy, Charles J. (2008) "Assessing the Validity of Statistical Inferences in Public Health Research: An Evidence-Based, 'Best-Practices' Approach," *Journal of the Georgia Public Health Association*: Vol. 3 : No. 1 , Article 2.
DOI: 10.20429/jgpha.2008.030102
Available at: https://digitalcommons.georgiasouthern.edu/jgpha/vol3/iss1/2

Assessing the Validity of Statistical Inferences in Public Health Research: An Evidence-Based, 'Best Practices' Approach[i]

Karl E. Peace
Jiann-Ping Hsu College of Public Health
Georgia Southern University

[ii]Anthony V. Parrillo
Jiann-Ping Hsu College of Public Health
Center for Rural Health and Research
Georgia Southern University

Charles J. Hardy
Jiann-Ping Hsu College of Public Health
Georgia Southern University

ABSTRACT

Like many fields, public health has embraced the process of evidence-based practice to inform practice decisions and to guide policy development. Evidence-based practice is typically dependent upon generalizations made on the bases of the existing body of knowledge – assimilations of the research literature on a particular topic. The potential utility of scientific evidence for guiding policy and practice decisions is grounded in the validity of the research investigations upon which such decisions are made. However, the validity of inferences made from the extant public health research literature requires more than ascertaining the validity of the statistical methods alone; for each study, the validity of the *entire* research process must be critically analyzed to the greatest extent possible so that appropriate conclusions can be drawn, and that recommendations for development of sound public health policy and practice can be offered. A critical analysis of the research process should include the following: An a *priori* commitment to the research question; endpoints that are both appropriate for and consistent with the research question; an experimental design that is appropriate (i.e., that answers the research question[s]); study procedures that are conducted in a quality manner, that eliminate bias and ensure that the data accurately reflect the condition(s) under study; evidence that the integrity of the Type-I error – or false-positive risk – has been preserved; use of appropriate statistical methods (e.g. assumptions checked, dropouts appropriately handled, correct variance term) for the data analyzed; and accurate interpretation of the results of statistical tests conducted in the study (e.g., the robustness of conclusions relative to missing data, multiple endpoints, multiple analyses, conditions of study, generalization of results, etc.). This paper provides a framework for both researcher and practitioner so that each may assess this critical aspect of public health research.

INTRODUCTION

> "*Anybody who makes or influences decisions that can affect the health of populations deserves ready access to the best evidence on what works, for what purpose, and at what costs, in order to make good choices among policies and to consider alternative uses of resources.*"
>
> Johnathan E. Fielding, Chair
> Task Force on Community Preventive Services
> (in Zaza, Briss, and Harris, 2005, p. xi)

The need for evidence-based decision-making in public health has grown in prominence during recent years. In the mid-1990s, when the first efforts were undertaken to synthesize scientific information about the effectiveness of health promotion and disease prevention programs, few public health practitioners and policymakers were familiar with the notion of evidence-based practice. Now, more than a decade later, themes of evidence-based public health (EBPH) have become the focus of regional, national, and international public health meeting agendas, and the phrase "evidence of effectiveness" is a central theme of public health practice (Anderson, Brownson, Fullillove, Teutsch, Novick, Fielding, and Land, 2005). EBPH is grounded in the evidenced-based practice movement in the field of medicine. However, there are notable differences between the two disciplines of medicine and public health that require distinct approaches of the application of evidence-based practice (Allee, 2004). Specifically, the goal of evidenced-based medicine is "...the best possible management of health and disease in individual patients..." (Milos and Stachenko, 2003, p. SR2). The goal in EBPH is "...the best possible management of health and disease and their determinants at the *community* level." (Milos and Stachenko, 2003, p. SR2) As such, policy development and interventions to improve the health of our public require an understanding of the complexities of organizational structures, interactions, and a myriad of other dynamics that impact decision making at the local, state, regional, and national levels.

    In today's public health marketplace, we have been asked to build a practice that is the synthesis of scientific skills, enhanced communication, political acumen, and common sense, yet, how best to conduct evaluation research and/or to interpret research findings in the "best practices" world in which we now live has been left largely up to researchers and/or governmental agencies. Because of the overwhelming volume of research literature, it has been difficult to sort through and extract what is effective for one's unique public health practice needs. Though the *Guide to Clinical Preventive Services* (U.S. Preventive Services Task Force, 1996) and the *Guide to Community Preventive Services: What Works to Promote Health?* (Zaza, Briss, and Harris, 2005) promote evidence-based approaches to medical and public health practice, creating practical ways of sorting through the large body of research is especially important because much of the public health workforce is not always well-trained in how to critically analyze the research evidence.

    The purpose of this paper is to present a roadmap for public health practitioners to utilize when critically analyzing research to determine the "best-evidence" to drive policy and practice decisions (Allee, 2004). By "best evidence" we emphasize that it is the quality of the evidence, and not its quantity, that is our concern. Consistent with this premise, our aim is to assist practitioners in making decisions based on the "best" information available on a particular public health issue. The roadmap presented in this paper identifies issues that should have been considered in developing, conducting, and reporting randomized controlled clinical trials (RCCT) research so that valid statistical inferences can be made. By selecting RCCT, the typical approach practiced in clinical drug trials, we do not mean to "force-fit" a classic experimental design methodology into our roadmap for public health practitioners. We recognize that much of

the research in public health utilizes quasi-experimental designs; though not generalizable in the literal sense, such studies do not exempt themselves from the same rigorous standards that are used in true experimental designs (Campbell and Stanley, 1966). While not all research designs should be judged to be equal in value in the decision making process, all decision making should be based upon valid research – the degree to which the research truly measures what it intended to measure and how truthful the research results are (Golafshani, 2003). Thus, determining the validity of statistical inferences for public health research is a fundamental element in the critical assessment and evaluation of research in the EBPH process.

Determining the validity of research studies to inform the development, implementation and evaluation of public health interventions and policy development requires an understanding of two key concepts: evidence and critical appraisal. "Evidence can be defined as that 'which furnishes proof,' and critical appraisal can be defined as an evaluation process 'which determines the significance or worth of something by careful appraisal and study" (Allee, 2004, p. E-2). Reviewing the scientific evidence and linking this evidence to recommendations for public health practice requires one to critically evaluate the validity of each study being considered (through this process the practitioner will be able to determine that a public health intervention is effective, ineffective or harmful, or that the evidence is not sufficient to determine that the intervention is effective (Zaza, Briss, and Harris, 2005). All of these findings are important in determining the utility of a particular study in developing recommendations to public health practice and policy development – EBPH. In this paper we suggest that a critical factor in this process is the determination of the validity of the individual interventions (studies) that are selected for inclusion in the evaluation.

At its genesis, the validity of a research investigation has not been assessed. Validity in research is only achieved through careful attention to detail and insistence on quality in all phases of the research process. All aspects of a research investigation – a public health intervention, basic laboratory research, a clinical trial testing the efficacy of a drug, or medical research that is conducted at a university – should be documented. Doing so permits an audit of what *was to be* done, what *was* done, and how conducting the study may have led to *differences in the two* that might ultimately have affected conclusions or inferences drawn from the data. Planning the research investigation culminates in the development of a protocol, or the set of rules upon which the study will be governed (Peace, 1991c; 2005). The protocol starts with a well-defined question or objective that the study will seek to answer (Peace, 2006a). The data – or endpoints – needed to provide an answer are identified. The question(s) is (are) then formulated within a hypothesis testing framework. The number of participants required to address the question (statistical power) is then determined. Procedures for conducting the experimental investigation that will produce the required data are developed, and methods for collecting, computerizing and quality-assuring the data are specified. Finally, statistical methods for analyzing the data to address the question(s) is (are) decided upon and described.

Nearly six decades have elapsed since the Medical Research Council (MRC) undertook two controlled clinical trials of potentially curative drugs. The second study, conducted in 1947-1948, is widely accepted as the first randomized, controlled clinical trial – RCCT (MRC Streptomycin in Tuberculosis Trials Committee, 1948; Sutherland, 1998; Thomson, 1975). It is widely held that the randomization and experimental design aspects of the trial (instituting true controls) were the brain-children of Sir Austin Bradford Hill, Director of MRC's Statistical Research Unit (Armitage, 1992; Hill, 1990). The 1962 Kefauver-Harris Amendments to the Federal Food, Drug and Cosmetics Act of 1938 (Bren, 2007) represented a watershed event in the evolution of using evidence to support drug claims. This landmark legislation – often referred to tongue-in-cheek as the "Full Employment Act" of biostatisticians in the pharmaceutical industry – required that all drugs thereafter be proven effective in order to gain approval for marketing in the United States by the Food and Drug Administration (FDA).

President John F. Kennedy signs the Kefauver-Harris Amendment (October 10, 1962)

(Source: U.S. Food and Drug Administration, 2007)

In the time since the first RCCT was conducted, the FDA has been a major player in both promoting and advancing the need for better quality clinical investigations, as well as evolving evidentiary methodological standards to accomplish this goal. Over the years, tough standards, scrutiny, collaboration with science (e.g., National Research Council and National Academy of Sciences), and the passage of legislation addressing these issues, has greatly reduced the number of poor studies and ineffective products:

> *"…in the early days, there were no standards, no controlled trials, and no post-marketing surveillance. But we got better over time. And in the age of effectiveness, we also got better at safety."*
>
> Robert Temple, MD (in Bren, 2007)

As a result, much progress has been made in strengthening evidence to support claims deriving from clinical investigations; the double-blind (DB) RCCT is now considered the gold-standard for evidentiary medicine.

Strengthening the evidence to support claims deriving from clinical investigations is the result of first recognizing the need for improvement and second, the collective desire to improve quality in all aspects of such investigations (Peace, 1991a; 1992; 2006a). Improving the experimental design of the investigation is one aspect, and this includes ensuring an adequate number of participants (Peace, 1991b; 2005; 2006a; Thomas, 1977). Improving the quality of reporting the investigation (Bailar and Mosteller, 1988; Begg, Cho, Eastwood et al., 1996; Moher, Cook, Eastwood, Olkin, Rennie, and Stroup, 1999; Peace, 1984; Stroup, Berlin, Morton, et al., 2000) is a second important aspect of quality improvement in all aspects of clinical intervention trials.

The investigational new drug exemption/new drug application (IND/NDA) rewrite of the mid 1980s (see 21 CFR Parts 312, 314, 511, and 514, 1987) is an example of an effort that

recognized the need for better design and quality throughout drug research and development, particularly clinical development, coupled with the need for better summarization and presentation of results; this legislation introduced for the first time the dose comparison or clinical dose-response trial. One impact of the IND/NDA rewrite was to serve notice to the pharmaceutical industry that they had to do a better job at identifying dose regimens for drugs to be marketed. It is widely held that Dr. Bob Temple at the FDA was of the opinion that the doses of drugs on the market prior to the IND/NDA rewrite were generally too high. This position is understandable in the absence of regulation requiring evidence of clinical dose-response. This had considerable impact on the design of clinical development programs. The IND/NDA rewrite also had an enormous impact on how data are organized and presented in NDAs to expedite FDA review (U.S. Department of Health and Human Services, 1998).

Step 1: Evaluating the Research Plan

RESEARCH QUESTION AND ENDPOINTS

The research question of the investigation should be defined so that it is unambiguous. A thorough review of all relevant scientific literature will assist the researcher in defining the research problem. Deductive logic typically yields clear and testable research questions. For example, in an investigation about the antihypertensive efficacy of drug D in some defined population, the statement: "The objective of this investigation is to assess the efficacy of drug D," although providing general information as to the question – Is drug D efficacious? – is ambiguous. The statement: "The objective of this investigation is to assess whether drug D is superior to placebo P in the treatment of hypertensive patients with diastolic blood pressure (DBP) between 90 and 100 mmHG for six months" is better – the hypertensive population to be treated and what is meant by efficacious in a comparative sense are each specified. However, the data or endpoint(s) upon which antihypertensive efficacy will be based is (are) not specified. DBP is stated, but how will it be measured? Will a sphygmomanometer or a digital monitor be used? Will DBP be measured in the sitting, standing, or supine position? Further, what function of the DBP will be used? For example, will it be the change from baseline to the end of the treatment period, or whether the patient achieves a therapeutic goal of normatensive (DBP $\leq$ 80 mmHG) by the end of the treatment period? Not attending to – and controlling for – each of these variables results in a threat to validity.

THE HYPOTHESIS TESTING FRAMEWORK

Reformulating the question within a hypothesis-testing framework adds needed clarity. The question regarding the efficacy of drug D as an anti-hypertensive is the alternative hypothesis (H_a: $\mu_D - \mu_P = \delta > 0$) versus the null hypothesis (H_o: $\delta = 0$), where δ is the difference between the true effects of drug D and placebo P. It should be noted that H_a has to be directional in nature (one-sided) in order to capture the question of the efficacy of drug D (as compared to placebo P) (Peace, 1991c). Again, as in the development of the research question and endpoints, familiarity with the research literature provides guidance for the development of testable hypotheses. We are aware that within public health, not all research questions of interest require formulating as the alternative hypothesis. Confidence intervals provide an alternative inferential framework. Pilot studies may be conducted to provide familiarity with experimental methods and/or to conduct exploratory, hypothesis generating analyses to be pursued in future research. A key point to be made here is that there can never be a scientifically valid answer to a question deriving from data analyses without an *a priori* commitment to that question before collecting the data.

THE NUMBER OF PARTICIPANTS

The number of participants required to provide a valid inference must be determined prior to beginning the investigation (Peace, 2006b). It requires specifying the difference δ between interventions reflecting similarity or superiority, the magnitude of the Type-I error α, the statistical power $1 - \beta$, or degree of certainty required to detect δ, and an estimate of variability of the data or endpoint reflecting the question. The difference δ reflects the minimum difference between regimens in order to conclude the superiority of one regimen (if the question is superiority), and reflects the maximum difference between regimens to conclude similarity or non-inferiority (if similarity or non-inferiority is the question). Sample size determination is not a cookbook exercise and should not be taken lightly (Brasher and Brant, 2007; Lenth, 2001; Lwanga and Lemeshow, 1991; Peace, 2006b). The specification of δ (i.e., quantification of the research question) is the responsibility of the substantive area scientist (clinician or medical or public health scientist), and requires careful thinking and exploration by both the statistician and the substantive area scientist. A δ too large may lead to failure to answer the question due to too small a sample. A δ too small would increase costs of conducting the investigation and may not be accepted as clinically meaningful.

PROCEDURES FOR CONDUCTING THE INVESTIGATION

Procedures for conducting the investigation are crucial to the success of the investigation. All procedures or methods pertinent to: how participants are selected and treated; the data measurement process; elimination or reduction of bias; visit scheduling; patient and investigator expectations; handling of adverse events; problem management and so on, should be specified. Failure to minimize sources of variability (Peace, 1992) other than true inter- and intra-participant variability may lead to failure to reach the desired conclusion.

DATA COLLECTION, COMPUTERIZATION AND QUALITY ASSURANCE

Data collection, computerization, and quality assurance methods for the data should be specified prior to beginning the investigation. Fundamental to any valid interpretation or inference is the integrity of the data analyzed. One must have assurance that the data analyzed are the data collected. In addition, the validity and reliability of the instrumentation employed are key factors. Every effort should be made to select sound measures and to familiarize oneself with what data is being produced by the measurement tool and the measurement scales of the data. Moreover, prior to any data collection all participants must be fully informed as to the procedures, benefits and risks involved and provide voluntary consent (usually written consent is required by most Institutional Review Boards).

STATISTICAL METHODS

Statistical methods for analyzing the data must be identified and described prior to beginning the investigation and included in the statistical analysis plan (Peace, 2005). For an interpretation or an inference from a statistical analysis to be valid, there must be an *a priori* commitment to the question, analysis unit, measurement tools, and methods of analysis – subject to assumptions underlying the methods being satisfied by the data.

Step 2: Evaluating How the Investigation was Conducted

The investigation, per the scientific plan or protocol and procedures, must be conducted in a quality and unbiased manner. The study should be carefully monitored to ensure adherence to the protocol and any institutional or regulatory requirements. For multi-investigator studies, an additional goal of monitoring is to reduce inter-investigator heterogeneity (Peace, 1992).

Step 3: Evaluating the Statistical Analyses, Interpretation, and Inference

Once the data collected are computerized and quality assured, planned statistical analyses may begin. Assumptions underlying the validity of the analysis methods should be checked to see if they hold for the data being analyzed. For example, analysis of variance methods require the data to satisfy assumptions related to normality, independence, and homogeneity of variance (Keppel, 1982). Similarly, analyses of covariance methods further require the covariate to be independent of the intervention groups and that the regressions of the data on the covariate are parallel across the groups (Cohen and Cohen, 1983). If the assumptions do not hold, then transformations of the data that lead to the assumptions being satisfied or nonparametric methods may be performed.

The number of participants needed to be enrolled and completed in a research investigation is computed based on the Type I error, the estimate of the variability of the endpoint reflecting the question, the quantification of the question in terms of the magnitude of the difference (δ) to be detected, and the power to detect that difference. Rarely does an investigation finish with complete data on the planned number of participants. There are numerous reasons for this: drop outs due to adverse experiences or lack of efficacy, missed visits due to brief illnesses or a variety of logistical reasons, relocation, and so on. The reasons for missing data must be thoroughly investigated and if the data are missing at random, procedures exist that permit a valid statistical analysis. However, whether the inference is credible and generalizable will depend on the amount of missing data and the methodology for dealing with the missing data.

Crucial to the validity of an inference is the integrity of the Type-I error or false positive risk (probability of concluding an effect when no effect exists). On the simplest level, a valid analysis produces an estimate of comparative effect of the intervention(s) and a corresponding *p*-value. The estimate may be regarded as a real comparative effect, provided the *p*-value is small (e.g., ≤ 0.05) and is correctly determined. Parenthetically, decisions affecting the public's health should not be based on the size of the comparative effect in the absence of quantifying the risk that the effect is consistent with chance fluctuations. Analyses of multiple endpoints or multiple analyses of the same endpoint lead to chance findings when no effects exist.

Step 4: Evaluating the Reporting of the Results

Penultimately, when considering inferences from statistical analyses reported of public health intervention trials, the points raised in the commentary by Mayo-Wilson (2007) give one pause to consider the totality of conditions that must be met in order for the stated inferences to be valid. Not only must the assumptions underlying a statistical analysis methodology hold, but the entire process that ultimately produced the data that are analyzed should be examined in light of the question to be answered. This in large part depends on determining what *was to be* done, what *was* done, and how conducting the investigation may have led to *differences in the two* that may ultimately have affected conclusions or inferences drawn from the data. Finally, it is strongly recommended that limitations, delimitations, and assumptions of the research be acknowledged.

One goal of reporting the results of clinical investigations is to permit translating such investigations into practice. Attempts at translating investigations into practice have been reported in many areas, including AIDS (Turner, Newschaffer, Zhang, Fanning, and Hauck, 1999),

cancer risk (Jatoi and Proschan, 2006), Epoetin use (Cotter, Thamer, Narasimhan, Zhang, and Bullock, 2006), genital herpes (Hook and Leone, 2006), heart failure (Patel, White, and Deswal, 2007; Sharpe, 1998), hypertension (Goldstein, Coleman, Tu, et al., 2004), internal medicine (Julian, 2004), and obstetrics (Jesse, 2007). If the results of investigations are to be translated into practice, reporting the implementation aspects of the investigation must be improved (Mayo-Wilson, 2007).

Recognizing that clinical drug trials do not necessarily mimic clinical practice has spawned the relatively new area of translational research. Translational medicine is a branch of medical research that attempts to more directly connect basic research to patient care. Translational medicine is growing in importance in the healthcare industry, and is a term whose precise definition is in flux. In the case of drug discovery and development, translational medicine typically refers to the "translation" of basic research into real therapies for real patients. The emphasis focuses on the linkage between the laboratory and the patient's bedside, without a real disconnect. This is often called the "bench to bedside" definition (Wikipedia, 2007). From another perspective, translational medicine:

> *"…can have both a narrow as well as a more general definition. Perhaps the most specific definition is "bench to bedside" research wherein a basic laboratory discovery becomes applicable to the diagnosis, treatment or prevention of a specific disease…"*
>
> Phillip A. Pizzo, MD (2002)

For translational medicine to truly be an "applied" science, the protocol is brought forth by either a clinician-scientist who works at the interface between the research laboratory and patient care or by a team of basic and clinical science investigators.

In terms of their effectiveness, clinical interventions, along with programs in other areas of public health cannot fail to account for – and adapt to – the historical, legal, economic, political, and cultural aspects of the communities and populations they serve; such is the essence of translational research. In health promotion and disease prevention research, for example, such interventions can be centrally planned and outcome-focused (Nutbeam, 1996a), or they can engage the community in a responsive or reactive mode to adapt the program to the participatory input of local practitioners and residents (Nutbeam, 1996b); such "bottom up" community involvement is critical for findings to translate into success, since the problems of "best practices" research arise largely when a "top down" strategy is used, when a:

> *"…recommended or required best practice (usually tested in one or more particular localities) is imposed as policy from central authority upon the highly variable other settings in which they may not fit the particular circumstances."*
>
> (Green, 2001, p. 173; citing Hubbard and Ottoson, 1997)

Public health practitioners use applied research in the course of their work by first trying to prevent problems before they occur; this is what prevention is all about. And if a problem has already occurred, public health practitioners work hard to control the situation. If the problem affects many people, then surveillance systems will be developed and maintained. Thus, in theory, health problems are kept under control by monitoring them, as well as providing data to evaluate the effectiveness of the solutions; such strategies seek to prevent the problem from occurring again. While this approach appears to have been effective for acute situations, the evidence is less convincing for chronic situations that impact the health of our public. The *Healthy People* initiatives, grounded in science, built through public consensus, and designed to measure progress, provide evidence of such challenges.

Moreover, the development of sound public health policy should be based upon conclusive evidence of intervention effectiveness. Interpretation of the research literature can assist with the assimilation of the evidence for such action. It is important to note, that there are many policy areas where such evidence does not exist or the evidence does not take into account the context or local character. Thus, practitioners must be well versed in the evidence in order to move promising interventions to the policy level.

If innovative interventions are employed it is critical to evaluate their effectiveness so that the evidence base can be expanded. The process presented in this paper can assist both the researcher and the practitioner by providing a framework for evaluating existing public health research, as well as a model for conducting research to evaluate the effectiveness of interventions aimed at enhancing the health of our public. In each case, the end-product will be valid research projects that yield valid findings that contribute best practices results for our field.

CONCLUDING REMARKS

The demand for evidence-based practice in public health interventions and policy has reached increased importance (Anderson et al., 2005). This paper describes the importance of one aspect of the process used to summarize the evidence of effectiveness of public health interventions – the critical appraisal of the validity of individual research investigations considered in the development of the evidence for the public health intervention under consideration. While we understand that the "real world" of public health often demands that problems and challenges are tackled in the absence of definitive research evidence, we also realize that our world is changing, and that existing research that can be assimilated needs to be subjected to critical review before developing policy and practice recommendations. The Task Force on Community Preventive Services employed a standardized approach to the search of the scientific literature for evidence, the identification of valid evidence, and the translation of scientific-based evidence that would result in specific recommendations for public health practice (see Zaza, Briss, and Harris, 2005). Regardless of the setting in which it is conducted, every investigation requires design of a research protocol, conducting the investigation per that protocol, collecting the data, analyzing the data using valid statistical procedures, making valid inferential conclusions relative to the objectives of the protocol, and reporting results. When considering the validity of statistical inferences in public health research – intervention trials, or other experimental and quasi-experimental studies – one's attention may be restricted to whether the statistical methodology used to analyze the data is appropriate for the type of data, and whether the assumptions underlying the methodology hold for the data. This is essential if an inference from a statistical analysis is to be valid, but not necessarily sufficient; valid inferences are derived from well-planned, well-conducted, properly analyzed investigations (see Text Box 1). We add this final thought: *No analysis of data can salvage a poorly designed or poorly conducted investigation*. To put it simply, there is no statistical fix for poor research!

Collectively, in our experience supporting research and clinical development of pharmaceuticals and public health intervention, or conducting research in the behavioral and social sciences of health promotion, we have had the opportunity to analyze data from a variety of perspectives; in many cases, innovative analysis methods have been developed. While valid data analyses are both necessary and important, it is our belief – based on experience *and* observation – that advances in science, advances in the treatment of patients, and improvements in the public's health are the result of well-planned and well-conducted investigations more than they are the product of any esoteric analysis of data. As Dr. Lewis Thomas said so eloquently:

> *"From here on, as far ahead as one can see, medicine must be building as a central part of its scientific base a solid underpinning of biostatistical and epidemiological knowledge. Hunches and intuitive impressions are essential for getting the work started, but it is only through the quality of numbers at the end that the truth can be told."*
>
> Thomas, 1977, p. 675 (emphasis added)

While there are significant advantages to developing public health policy and practice recommendations on the basis of an assimilation of valid research investigations, there are several limitations to the approach. For example, in public health, even where there is existing research literature on interventions, "the evidence base always will be incomplete for some variation in intervention design and/or subpopulation of interest" (Anderson et al., 2005). Moreover, synthesizing research from sources other than randomized trials – that is, the systematic fusion of both the quantitative *and* the qualitative – is among the new and exciting frontiers of public health research. Currently assimilation of the scientific literature is heavily grounded in analysis of quantitative research. Much more attention needs to be given to the body of qualitative research so that appropriate and valid conclusions can be drawn.

Box 1: Questions to Consider in Determining the Validity of Inferences

---

Questions to Consider in Determining the Validity of Inferences

When considering whether an inferential conclusion from an investigation is valid, consider these questions:

1. Was there an *a priori* commitment to the question?
2. Was the endpoint appropriate for the question?
3. Was the experimental design appropriate for the condition being studied?
4. Was the investigation conducted in a quality manner to eliminate bias and to ensure accuracy of the data?
5. Were steps taken to preserve the integrity of the Type-I error?
6. Were the statistical methods for analyses valid (assumptions checked, dropouts appropriately handled, correct variance term, etc)?
7. Are the results of statistical analyses properly interpreted (the correct inferred population, impact of multiple endpoints or analyses, etc.)?
8. Were limitations, etc. addressed?

Research in public health is changing to include a move away from single investigators and/or disciplines to a transdisciplinary research model. As the research model changes, it is hoped that the link between practice and research will grow stronger. According to the report by the Institute of Medicine entitled *Who Will Keep the Public Healthy? Educating Public Health Professionals for the 21st Century* (Gebbie, Rosenstock, and Hernandez. 2003), "The study of interventions, will in turn, dictate the third sea-change in public health research: community participation." (p. 12) This will require development of new research models and expertise by the all of the partners in the research process – community members, researchers and practitioners.

While the need for public health interventions and policy decisions based on sound evidence is widely acknowledged, the underutilization of evidence-based research will not be solved by simply adopting the roadmap presented herein: We must change the way we do our business. The Canadian Institutes for Health Research (CIHR) have argued that the gap between "what is

known" and "what is done" in practice settings can be bridged by effective *knowledge translation* (KT) (NIDRR, 2005).   KT is "…the exchange, synthesis and ethically-sound applications of knowledge – within a complex system of interactions among researchers and users – to accelerate the capture of the benefits of research…through improved health, more effective services and products, and a strengthened health care system." (CIHR 2004, p. 4)   KT is based upon the principle of active exchange of information between researchers who create new knowledge and those who use it to enhance the public's health.   For success in public health settings, all stakeholders (i.e., researchers, practitioners, administrators, policy makers, and members of the community) must be brought together in all stages of the research process; such an approach is thought to guard against the ineffectiveness of the traditional knowledge transfer model, best described as having a unidirectional flow of knowledge from the researcher(s) to the practitioner(s).   Interestingly, KT places the "…emphasis on the quality of the research prior to dissemination and implementation of research evidence within a system." (NIDRR, 2005, p. 2)

While not without its limitations, we believe the approach presented in this paper will facilitate appropriate inclusion of public health research in implementing the "evidence of effectiveness" process that has become part of the dialogue in the field of public health (see Anderson et al., 2005).   Such an approach will allow us to take full advantage of our scientific knowledge in the development of sound public health practice and policy.   Moreover, it will assist researchers in expanding the science base underlying public health practice.   Evidence-based decision making in public health has the potential to improve the science, practice, and policy development of the broad domain of what constitutes the field of public health as long as the process is based upon valid research studies.

REFERENCES

21 CFR 312, 314, 511, and 514 (1987). *New Drug, Antibiotic, and Biologic Drug Product Regulations.* [Docket No. 82N-0394] 52 FR 8798.

Allee, N. (2004) Supporting decisions with best evidence. In Allee, N., Alpi, K., Cogdill, K.W., Selden, C., & Yougkin, M.   *Public Health Information and Data: A Training Manual.*   Bethesda, MD: National Library of Medicine. Available at: http://www.phpartners.org/pdf/phmanual.pdf.

Anderson, L.M, Brownson, R.C., Fullilove, M.T., Teutsch, S.M., Novick, L.F., Fielding, J., & Land, G.H. (2005). Evidence-based public health policy and practice: Promises and limits. *American Journal of Preventive Medicine, 28,* 226-230.

Armitage, P. (1992). Bradford Hill and the randomized controlled trial. *Pharmaceutical Medicine, 6,* 23 37.

Bailar, J. & Mosteller, F. (1988). Guidelines for statistical reporting in articles for medical journals: Amplifications and explanations. *Annals of Internal Medicine, 108,* 266-273.

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schulz, K.F., Simel, D., & Stroup, F. (1996). Improving the quality of reporting of randomized con-trolled trials: the CONSORT statement. *JAMA 276*(8), 637-639.

Brasher, P.M.A. & Brant, R.F. (2007). Sample size calculations in randomized trials: Common pitfalls. *Canadian Journal of Anesthesia, 54,* 103-106.

Bren, L. (2007). The advancement of controlled clinical trials. *FDA Consumer Magazine, 41*(2). Washington, DC: U.S. Food and Drug Administration. (March-April, 2007).   Available at: http://www.fda.gov/fdac/features/2007/207_trials.html.

Brownson, R.C., Baker, E.A., Leet, T.L., & Gillespie, K.N. (2003). *Evidence-Based Public Health.* New York, NY: Oxford University Press.

Campbell, D.T. & Stanley, J.C. (1966). *Experimental and Quasi-Experimental Designs for Research.* Chicago, IL: Rand McNally College Publishing Company.

Canadian Institutes of Health Research (CIHR). (2004). *Knowledge translation strategy 2004-2009: Innovation in action.* Ottawa, ON: Canadian Institutes of Health Research.

Cohen, J. & Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, Second Edition*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Cotter, D., Thamer, M., Narasimhan, K., Zhang, Y., & Bullock, K. (2006). Translating epoetin research into practice: The role of government and the use of scientific evidence. *Health Affairs 25*(5), 1249-1259.

Field, M.J., & Lohr, K. (Eds.) (1990). *Clinical Practice Guidelines: Directions for a New Program*. Washington, DC: National Academies Press.

Gebbie, K., Rosenstock, L., & Hernandez, L.M. (Eds.) (2003). *Who Will Keep the Public Healthy? Educating Public Health Professionals for the 21st Century*. Committee on Educating Public Health Professionals for the 21st Century, Board of Health Promotion and Disease Prevention, Institute of Medicine. Washington, DC: National Academies Press.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. The Qualitative Report, *8 (4)*, pp. 597-607. Available at: http://www.nova.edu/sss/QR/QR8-8/golafshani.pdf.

Goldstein, M.K., Coleman, R.W., Tu, S.W., Shankar, R.D., O'Connor, M.J., Musen, M.A., Martins, S.B., Lavori, P.W., Shlipak, M.G., Oddone, E., Advani, A.A., Gholami, P., & Hoffman, B.B. (2004). Translating research into practice: Organizational issues in implementing automated decision support for hypertension in three medical centers. *Journal of the American Medical Information Association, 11,* 368-376.

Green, L.W. (2001). From research to "best practices" in other settings and populations. *American Journal of Health Behavior, 25*(3), 165-178.

Hill, A.B. (1990). Memories of the British Streptomycin trial in tuberculosis. *Controlled Clinical Trials, 11,* 77-90.

Hook, E.W. & Leone, P. (2006). Time to translate new knowledge into practice: A call for a national genital herpes control program. *The Journal of Infectious Diseases, 194,* 6-7.

Hubbard, L. & Ottoson, J. (1997). When a bottom-up innovation meets itself as a top-down policy. *Science Communication, 19,* 41-55.

Jatoi, I. & Proschan, M.A. (2006). Clinical trial results applied to management of the individual cancer patient. *World Journal of Surgery 30*(7), 1184-1189.

Jesse, D.E. (2007). *Translating POP (Psychosocial Obstetrical Profile) research results into practice and policy*. The 18th International Nursing Research Congress Focusing on Evidence-Based Practice. Vienna, Austria: July 11-14, 2007.

Julian, D.G. (2004). Translation of clinical trials into clinical practice. *Journal of Internal Medicine, 255,* 309-316.

Keppel, G. (1982). *Design and Analysis: A Researcher's Handbook, Second Edition*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Lenth, R.V. (2001). Some practical guidelines for effective sample size determination. *The American Statistician, 55,* 187-194.

Lwanga, S. & Lemeshow, S. (1991). *Sample Size Determination in Health Studies: A Practical Manual*. Geneva, CH: World Health Organization (80 pp.).

Mayo-Wilson, E. (2007). Reporting implementation in randomized trials: Proposed additions to the Consolidated Standards of Reporting Trials statement. *American Journal of Public Health, 97*(4), 630-633.

Milos, J. and Stachenko, S. (2003). Evidence-based public health, community medicine, preventive care. *Medical Science Monitor, 9*(2), SR-2.

Moher D., Cook, D.J., Eastwood, S., Olkin, I., Rennie, D., & Stroup, D.F. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. *Lancet, 354,* 1896-1900.

MRC Streptomycin in Tuberculosis Trials Committee (1948). Streptomycin treatment of pulmonary tuberculosis. *British Medical Journal, ii, 769-783.*

National Center for the Dissemination of Disability Research (NCDDR) (2005). *What is Knowledge Translation?* Focus: A Technical Brief from the National Center for the Dissemination of Disa-bility Research. Available at: http://www.ncddr.org/kt/products/focus/focus10/Focus10.pdf.

Nutbeam, D. (1996a). Achieving "best practice" in health promotion: Improving the fit between research and practice. *Health Education Research, 11*, 317-326.

Nutbeam, D. (1996b). Improving the fit between research and practice in health promotion: Over-coming structural barriers. *Canadian Journal of Public Health, 87*(Supplement 2), 18-23.

Patel, P., White, D., & Deswal, A. (2007). Translation of clinical trials results into practice: Temporal patterns of beta-blocker utilization for heart failure at hospital discharge and during ambulatory follow-up. *American Heart Journal, 153*(4), 515-522.

Peace, K.E. (1984). Data listings and summaries should also reflect experimental structure. *Biometrics, 40*(1), 256.

Peace, K.E. (1991a). Shortening the time for clinical drug development. *Regulatory Affairs Professionals Journal, 3*, 3-22.

Peace, K.E. (1991b). *Sample size considerations of clinical trials pre-market approval.* Invited presentation given at the 27th Annual Meeting of the Drug Information Association, Washington, DC.

Peace, K.E. (1991c). One-sided or two-sided *p*-values: Which most appropriately address the question of drug efficacy? *Journal of Biopharmaceutical Statistics, 1*(I), 133-138.

Peace, K.E. (1992). The impact of investigator heterogeneity in clinical trials on detecting treatment differences. *Drug Information Journal, 26*, 463-469.

Peace, K.E. (2005). Statistical section of a clinical trial protocol. *The Philippine Statistician 54*(4), 1-8.

Peace, K.E. (2006a). Importance of the research question relative to analysis. *Philippine Statistical Association Newsletter 1*(1), 7-9.

Peace, K.E. (2006b). Sample size considerations of clinical trials pre-market approval. *The Philippine Statistician, 55*(2), 1-27.

Pizzo, P.A. (2002). Letter from the dean. *Stanford Medicine Magazine.* Palo Alto, CA: Stanford School of Medicine. Available at: http://mednews.stanford.edu/stanmed/2002fall/letter.html.

Sharpe, N. (1998). Translation of clinical trials results into practice. *European Heart Journal, 19*(Suppl L), L28-32.

Stroup, D.F., Berlin, J.A., Morton, S.C., Olkin, I., Williamson, G.D., Rennie, D., Moher, D., Becker, B.J., Sipe, T., & Thacker, S.B., for the Meta-Analysis of Observational Studies in Epidemiology (MOOSE) Group (2000). Meta-analysis of observational studies in epidemiology: A proposal for reporting. *JAMA, 283*(15), 2008-2012.

Sutherland, I. (1998). Medical Research Council streptomycin trial. In P. Armitage & T. Colton (Eds.), *Encyclopedia of Biostatistics.* Chichester, United Kingdom: John Wiley & Sons, Ltd. (pp. 2559-2266).

Thomas, L. (1977). Biostatistics in medicine. *Science, 198*, 675 (November).

Thomson, A.L. (1975). *Half a Century of Medical Research, Volume 2: The Programme of the Medical Research Council.* London, United Kingdom: HMSO (pp. 238-239).

Turner, B.J., Newschaffer, C.J., Zhang, D., Fanning, T., & Hauck, W.W. (1999). Translating clinical trial results into practice: The effect of an AIDS clinical trial on prescribed antiretroviral therapy for HIV-infected pregnant women. *Annals of Internal Medicine, 130*(12), 979-986.

U.S. Department of Health and Human Services (USDHHS) (1998). *The CDER Handbook.* Wash-ington, DC: Center for Drug Evaluation and Research CDER). Available at: www.fda.gov/cder/handbook/ handbook.pdf .

U.S. Food and Drug Administration (2007). *A Brief History of the Center for Drug Evaluation and Research.* Rockville, MD: Center for Drug Evaluation and Research (CDER). Available at: http://www.fda.gov/cder/about/history/page32.htm.

U.S. Prenventive Services Task Force (1996). *Guide to Clinical Preventive Services, 2nd Edition.* Washington, DC: U.S. Department of Health and Human Services.

Wikipedia, the Free Encyclopedia (2007). *Translational medicine.* St. Petersburg, FL: The Wikipedia Foundation. Available at: http://en.wikipedia.org/wiki/Translational_medicine.

Zaza, S., Briss, P.A., & Harris, K.W. (2005). *The Guide to Community Preventive Services: What Works to Promote Health?* New York, NY: Oxford University Press.

[ii] CORRESPONDING AUTHOR: Jiann-Ping Hsu College of Public Health, Georgia Southern University, PO Box 8148, Statesboro, GA 30460-8148. VOX: 912-478-5057; FAX: 912-478-0171; e-Mail: aparrill@georgiasouthern.edu