CONTENTS

Using Intervention Theory to Model Factors Influencing Behavior Change:
Project RESPECT
MARTIN FISHBEIN, MICHAEL HENNESSY, MARY KAMB, GAIL A. BOLAN, TAMARA HOXWORTH, MICHAEL IATESTA, FEN RHODES, JONATHAN M. ZENILMAN, and PROJECT RESPECT STUDY GROUP 363

Location of Death in Canada:
A Comparison of 20th-Century Hospital and Nonhospital Locations of Death and Corresponding Population Trends
DONNA M. WILSON, HERBERT C. NORTHcott, CORRINE D. TRUMAN, SUSAN L. SMITH, MARJORIE C. ANDERSON, ROBIN L. FAINSINGER, and MICHAEL J. STINGL 385

The Impact of a Community-Wide Self-Care Information Project on Self-Care and Medical Care Utilization
JUDITH H. HIBBARD, MERWYN GREENLICK, HOLLY JIMISON, JEFFREY CAPIZZI, and LYNN KUNKEL 404

Assessing Attitudes Toward New Names for Chronic Fatigue Syndrome
LEONARD A. JASON, HEATHER EISELE, and RENEE R. TAYLOR 424

Self- and Peer Assessment in a First-Year Communication and Interviewing Course
DAVID W. RUDY, MICHELE C. FEIFAR, CHARLES H. GRIFFITH, III, and JOHN F. WILSON 436

Impact of an Online Analytic Skills Course
ANITA FAREL, KARI L. UMBLE, and BARBARA POLHAMUS 446

Index 460

Sage Publications Thousand Oaks • London • New Delhi
EVALUATION & THE HEALTH PROFESSIONS is designed to provide a forum for all health professionals interested or engaged in the development, implementation, and evaluation of health programs. Original manuscripts are invited that deal with results of evaluation studies, instructional innovations, progress reports, and updates. The editors welcome the submission of articles dealing with the philosophical, technical, and political aspects of evaluation, which are unique to the health professions.

MANUSCRIPTS should be submitted to the editor of EVALUATION & THE HEALTH PROFESSIONS, R. Barker Bausell, Complementary Medicine Program, University of Maryland School of Medicine, Kernan Hospital Mansion, 2200 Kernan Drive, Baltimore, MD 21207-6687. Articles should not exceed 30 typewritten double-spaced pages with notes, references, tables and figures on separate pages, and should follow the guidelines set forth in the Publication Manual of the American Psychological Association (4th edition). An abstract of 100-150 words, as well as a brief biographical paragraph describing each author's current affiliation, research interests, and recent publications, should accompany the manuscript. Since manuscripts are sent out anonymously for evaluation, the author's name and affiliation should appear only on a separate covering page.

EVALUATION & THE HEALTH PROFESSIONS (ISSN 0163-2787) is published four times annually—in March, June, September, and December—by Sage Publications, 2455 Teller Road, Thousand Oaks, CA 91320. Telephone: (800) 818-SAGE (7243) and (805) 499-9774; fax/order line: (805) 375-1700; web site: www.sagepub.com; e-mail: order@sagepub.com. Copyright © 2001 by Sage Publications. All rights reserved. No portion of the contents may be reproduced in any form without written permission of the publisher.

Subscriptions/Inquiries: Annual subscription rates for institutions and individuals are based on the current frequency. Prices quoted are in U.S. dollars and are subject to change without notice. Canadian subscribers add 7% GST (and HST as appropriate). Outside U.S. subscription rates include shipping via air-speeded delivery: Institutions: $365 (within the U.S.) / $381 (outside the U.S.) / single issue: $105 (worldwide). Individuals: $78 (within the U.S.) / $94 (outside the U.S.) / single issue: $31 (worldwide).

Address all correspondence and permissions requests to Sage Publications, 2455 Teller Road, Thousand Oaks, CA 91320. All subscription inquiries, orders, and renewals for journals published by Sage Publications with ship-to addresses in North America, South America, and Canada must be addressed to Sage Publications, 2455 Teller Road, Thousand Oaks, CA 91320, U.S.A., telephone (800) 818-SAGE (7243) and (805) 499-9774, fax (805) 375-1700. All subscription inquiries, orders, and renewals with ship-to addresses in the U.K., Europe, the Middle East, and Africa must be addressed to Sage Publications, Ltd, 6 Bonhill Street, London EC2A 4PU, England, telephone +44 (0)20 7374 0645, fax +44 (0)20 7374 8741. All subscription inquiries, orders, and renewals with ship-to addresses in India and South Asia must be addressed to Sage Publications Private Ltd, P.O. Box 4215, New Delhi 110 048, India, telephone (91-11) 641-9884, fax (91-11) 647-2426. Address all permissions requests to the Thousand Oaks office.

Periodicals postage paid at Thousand Oaks, California, and at additional mailing offices.


Back Issues: Information about availability and prices of back issues may be obtained from the publisher's order department (address below). Single-issue orders for 5 or more copies will receive a special adoption discount. Contact the order department for details. Write to the London office for sterling prices.

Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Sage Publications for libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that the base fee of 50¢ per copy, plus 10¢ per copypage, is paid directly to CCC, 21 Congress St., Salem, MA 01970. 0163-2787/2001 $5.00 + $.10.

Advertising: Current rates and specifications may be obtained by writing to the Advertising Manager at the Thousand Oaks office (address above).

Claims: Claims for undelivered copies must be made no later than six months following month of publication. The publisher will supply missing copies when losses have been sustained in transit and when the reserve stock will permit.

Change of Address: Six weeks' advance notice must be given when notifying of change of address. Please send old address label along with the new address to ensure proper identification. Please specify name of journal, POSTMASTER: Send address changes to: Evaluation & the Health Professions, c/o 2455 Teller Road, Thousand Oaks, CA 91320.
Project RESPECT was a multisite randomized trial comparing three clinic-based interventions’ ability to increase condom use and prevent infection with HIV and sexually transmitted diseases. Because Project RESPECT had guiding concepts that determined the content of the sessions, the authors investigated how the intervention operated using these theoretical variables. Growth curve analysis and structural equation modeling estimated the correlation between intentions toward condom use and self-reports of condom use and isolated the treatment effects on mediating variables—attitudes, self-efficacy, and social norms—that predict intentions. The correlations between intentions and behavior exceeded .70 for both genders, justifying the emphasis on intentions. Project RESPECT was effective through changing attitudes and self-efficacy for females in both counseling interventions. For males, only enhanced counseling had significant effects on these two mediator variables.
Project RESPECT was a multisite, randomized trial to evaluate the effectiveness of three interventions (two sessions with an informational focus, two sessions of client-centered HIV prevention counseling, and four sessions of enhanced HIV prevention counseling) to increase condom use and prevent infection with HIV and other sexually transmitted diseases (STDs) (Kamb, Dillon, Fishbein, & Willis, 1996). The outcome analysis of Project RESPECT reported that both counseling interventions significantly increased condom use and decreased incident STDs relative to the informational intervention (Kamb et al., 1998). Although the difference in STD incidence was not statistically discernable between the two counseling interventions, condom use was higher in the enhanced counseling session compared with the two other interventions, although this difference disappeared over time.

For measures of condom use (e.g. “any use” and “condoms with last sex”), enhanced counseling participants tended to report safe behaviors more often, followed by those who were assigned to the brief counseling arm and then those who were assigned to the didactic message arm. . . . At the 9 and 12 month visits, there were no significant differences between the interventions. (Kamb et al., 1998, p. 1165)

Because Project RESPECT used a set of guiding concepts to structure the content of the intervention, it is important to investigate how Project RESPECT achieved its outcome in terms of these theoretical variables (Weiss, 1997). Empirically modeling the process of intervention functioning (i.e., validating the program theory of Project RESPECT) is important for two reasons. If the program is a success,
there will be interest in replicating the program in other locations that serve other populations. If the program is a failure, it is useful to know which element of the program failed (Chen, 1990). Treating social programs as an input-output problem with no knowledge of the mechanisms within the black box of the program is a mistake (Kaneko, 1999; Lipsey, 1993; Short & Hennessy, 1994).

Specifically, Project RESPECT focused on increasing the likelihood that a male condom would always be used for vaginal sex with main and occasional sex partners. Thus, the intervention sessions were designed to increase male participants’ intentions to always use a condom during vaginal sex and to increase female participants’ intentions to get their male sex partners to always use a condom for vaginal sex. Intentions themselves were determined by the following three precursors: (a) attitudes, (b) perceived norms, and (c) self-efficacy with respect to consistent condom use (Fishbein et al., 2001). Thus, positive intentions to always use a condom for vaginal sex with the main partner is a function of feeling positively about always using a condom for vaginal sex, perceiving that important others believe that one should always use a condom, and believing that one could always use a condom (or in the case of females, always get their partner to use one) even in the face of barriers to doing so.

However, the relative importance of the three psychosocial mediators as determinants of intentions (and behavior) can vary as a function of both behavior and the respondent. For example, one behavior (e.g., always using a condom for vaginal sex with the main partner) may be under attitudinal control (e.g., a function of positively or negatively valued outcomes), whereas another (e.g., always using a condom for vaginal sex with occasional partners) may be under normative or self-efficacy control (e.g., a function of one’s perception of the ability to perform the behavior). It is also possible that although attitudes are most important in one population or culture, perceived norms or self-efficacy may be most important in others. Because the intention to use a condom is very different from the intention to get one’s partner to use one, it seems reasonable to investigate whether Project RESPECT’s interventions worked differently for male and female participants. Kamb et al. (1998) did not report on gender differences in relation to condom use, and we find here that although both counseling interventions increased condom use in the short term for women, only enhanced counseling appeared to be effective for men.
This article investigates how and why the different interventions of Project RESPECT led to differential increases in condom use. It also explores how and why the interventions had different effects between genders. First, the correlations between intentions to use condoms and self-reports of condom use are estimated. If this correlation were low, it would call into question one of the basic assumptions underlying the intervention because changing intentions to use condoms consistently was the focus of the two counseling components. Then, treatment effects on the psychosocial mediating variables due to the three counseling interventions are estimated. In the same analysis, the associations between each of the three mediators and changes in intentions to use condoms are natural by-products. All these analyses are conducted longitudinally using growth curve analysis (Duncan, Duncan, Strycker, Li, & Alpert, 1999; Raudenbush, 2001).

PROJECT RESPECT

DESCRIPTION OF INTERVENTION

Project RESPECT compared three separate HIV/STD prevention interventions as part of a randomized trial. These included an informational intervention (HIV education) and two counseling interventions (HIV brief counseling and HIV enhanced counseling). Each of the three interventions primarily focused on increasing consistent condom use during vaginal sexual intercourse with one’s main and/or other partners. Each session of the counseling interventions built on the previous session or sessions, and all three interventions were conducted within the context of an HIV pre- and posttest counseling protocol in an STD clinic. Session 1 always included a discussion of the HIV test. Test results were given from 7 to 10 days later during Session 2 for HIV education and brief counseling or in Session 3 for enhanced counseling.

HIV education. The HIV education intervention consisted of two 5-minute informational sessions about preventing STDs and HIV. The pretest message was given by the clinician (medical practitioner) who examined and when appropriate treated the study participant for STDs during the initial clinic visit. The message included information about
the HIV test and a prevention message relevant to the participant’s reported risk behaviors. The clinician explicitly did not engage the participant in an interactive counseling process. The posttest session took place from 7 to 10 days later when the participant returned for the HIV test results. During the posttest session, the clinician or counselor informed the participant about the test results and the limitations of the test, again using a didactic approach. HIV transmission risks were reiterated, and specific behaviors or circumstances that might place the participant at risk for acquiring HIV or other STDs were again identified.

**HIV brief counseling.** In response to concern that true prevention counseling was not occurring in STD clinics and with advice from experts on counseling theory and practice, the Centers for Disease Control and Prevention (CDC) developed a client-centered counseling model that has been recommended for use in all U.S. STD clinics since 1993 (CDC, 1993, 1994). The intervention consisted of two 20-minute, client-focused and interactive counseling sessions with an HIV counselor. The first (pretest) session took place at the initial clinic visit prior to the STD examination, and the second (posttest) session occurred from 7 to 10 days later when the HIV test results were available. This intervention involved assessing the participant’s risk and self-perception of risk, highlighting that consistent and correct condom use could prevent HIV and other STDs, identifying barriers to condom use, negotiating an acceptable and achievable risk-reduction plan that focused on condom use, and supporting participant-initiated behavior change. The pretest session concluded with a behavioral goal-setting exercise in which the participant arrived at a small behavioral risk-reduction step that he or she could make before the final posttest session. Project RESPECT’s brief counseling intervention differed from CDC’s model only in its suggestion that the behavior-change step focus on condom use rather than on other possible risk-reduction measures (e.g., reducing the number of sex partners). It also focused on increasing perceptions of personal HIV risk, addressed two beliefs that underlie attitudes toward condom use (i.e., that condoms prevent disease transmission and infection with HIV and other STDs), and tried to increase the participant’s self-efficacy to use condoms consistently and correctly.
**HIV enhanced counseling.** The multisession enhanced counseling intervention was based on the theory of reasoned action (Fishbein, Middlestadt, & Hitchcock, 1991), social cognitive theory, and other psychosocial concepts. This intervention consisted of four interactive counseling sessions with an HIV counselor. The first session was identical to the first session of HIV brief counseling, took place during the initial clinic visit prior to the clinical examination, and lasted 20 minutes. The remaining sessions took place over the following 2 weeks, each session lasting 60 minutes. During Session 2, the focus was on changing attitudes toward condom use. During Session 3, the focus was on changing self-efficacy for using condoms. Each of these sessions ended with the participant arriving at a strategy for taking another step toward consistent condom use before the next session. Session 4 focused on changing perceived norms and perceptions of community support concerning consistent condom use. The session ended with the participant arriving at a long-term strategy for behavior change toward the goal of consistent condom use.

**PROJECT RESPECT PARTICIPANTS**

Project RESPECT took place between July 1, 1993, and June 30, 1995, during which time participants were enrolled into one of the three experimental conditions (2,440 men and 1,860 women completed baseline interviews after randomization into one of the three counseling modes; all participants were HIV negative at the start of the study). Postintervention data were obtained through interviews conducted immediately following the final intervention session (the immediate follow-up or IFU) and also through interviews conducted at four additional visits (3, 6, 9, and 12 months) postintervention. The participating clinics (Baltimore’s Druid Clinic, the Denver Metro Health Clinic, the City of Long Beach STD Clinic, the Newark STD Clinic, and San Francisco City Clinic) varied considerably in the demographic, cultural, and behavioral characteristics of their clientele; however, all clinics served predominantly minority, low-income, and young populations.

Project RESPECT’s intervention completion rates were high. Overall, 82% of participants completed all their assigned intervention sessions; 72% completed all four sessions of the HIV enhanced counseling intervention (most of the attrition came after the first session),
86% completed both sessions of the HIV brief counseling intervention, and 84% completed both sessions of the HIV education intervention. Intervention completion overall varied slightly among men (80%) and women (83%), and the higher completion rate among women occurred regardless of intervention.

Because formative research showed different patterns of condom use for main and occasional sex partners for males and females, sampling rules were used to determine if respondents at baseline answered questions about main or occasional partners. Females with only a main partner were asked about their main partner; all other females were asked only about their other partner. Males with only other partners or only a main partner answered about these contacts. Males with both main and other partners were assigned to discuss either main or other sex partners based on an odd or even respondent identification code. However, in the follow-up interviews, respondents were queried about both types if they were reported. Because intentions to use and actual condom use with occasional partners are higher (e.g., about 20% higher overall) than intentions to use and actual use with main partners and because sample sizes are dramatically smaller for data on occasional partners (e.g., when all respondents are considered, an average of 630 males and 319 female respondents have nonmissing data on intentions re occasional partners at each wave during the follow-up periods compared with averages of 1,184 males and 1,099 females having data on intentions re main partners in each wave of the follow-up), in the analyses that follow, we consider only respondents with main partners.

PROJECT RESPECT VARIABLES

Data collected were HIV and STD risk behaviors, determinants of condom use (i.e., attitudes, norms, self-efficacy, and intentions to use condoms), and biological outcomes including incident STDs and HIV infection. In each case, data were collected following standardized procedures outlined in a common protocol (Kamb et al., 1996, 1998). The interviews took on average 30 minutes to complete. (More information on Project RESPECT, including the intervention manuals, can be found on the Web at www.cdc.gov/nchstp/hiv_aids/projects/RESPECT, and researchers interested in the Project RESPECT data should contact Dr. Kamb at mlk5@cdc.gov.)
METHOD

GROWTH CURVES

Growth curves typically have two different components. First, they require an equation for predicting an individual respondent’s longitudinal values of an outcome variable as a function of time. In this application, for example, a respondent’s self-efficacy for consistent condom use is modeled as an initial posttest value and its change over time. Thus, predicted self-efficacy during the postintervention period is a function of the following two parameters: (a) the posttest at the first data collection after the intervention (e.g., at the immediate follow-up) and (b) the slope of change over time beginning at the IFU as:

\[ \text{Self-Efficacy}_{it} = \alpha_{i0} + \beta_{i1} \times \text{Time Metric}_{it} + \text{error}_{it} \]  

where the \( i \) subscript reflects individual observations, the Time Metric is the time scale, the \( \alpha_{i0} \) the value of self-efficacy when the Time Metric is zero, the \( \beta_{i1} \) the regression coefficient relating time and Self-Efficacy for each individual, and the \( t \) subscript represents the ordering of the observations from the IFU to the final follow-up observations 1 year later. Thus, the error term defines the individual (e.g., within subject) measurement errors of self-efficacy for each observation.

Growth curve analysis is a specific sort of multilevel model (Kreft & DeLeeuw, 1998; Raykov & Marcoulides, 2000) in which the parameters of an individual’s growth model (discussed earlier) can in turn be the dependent variables of other predictors (Francis, Fletcher, Stuebing, Davidson, & Thompson, 1991; Willett & Sayer, 1996). This is important here because it is of interest to discover if the two parameters of the individual model (e.g., initial value and slope of change over time) vary as a result of the type of HIV counseling reflected in the three experimental distinctions. Thus, the second component of the growth curve analysis here estimated the treatment effects of Project RESPECT insofar as it answers questions such as “Right after the intervention, is self-efficacy higher or lower in some counseling modality than others?” or “During the follow-up period, which counseling modalities increase in self-efficacy over time and which do
not?” Here, enhanced and brief counseling are compared with HIV education (acting as the status quo treatment category), using dummy variables for the counseling interventions. Therefore, the two parameter level equations are:

\[ \alpha_{i0} = \pi_{00} + \pi_{10}\text{Enhanced Counseling}_i + \pi_{20}\text{Brief Counseling}_i + \text{error}_{i0} \quad (2) \]

\[ \beta_{i1} = \pi_{01} + \pi_{11}\text{Enhanced Counseling}_i + \pi_{21}\text{Brief Counseling}_i + \text{error}_{i1} \quad (3) \]

where the \( \pi_{00} \) and \( \pi_{01} \) are the intercept and slope for HIV education and the other \( \pi \)s are deviations from this default value due to HIV enhanced counseling and HIV brief counseling group membership (i.e., these are the treatment effect parameters on the average value and change over time). The error term in these equations represents the between-subject variability that remains unexplained.

Finally, intercept and slope parameters of one growth curve for a particular data series can be correlated with the intercept and slope parameters of another data series (Duncan et al., 1999, chapter 5). This involves nothing more than estimating the simple correlation between parameters of each of two growth curves: the correlation between the initial values (intercept) and the correlation between the individual changes over time (slope). The former correlation represents the correlation between the starting levels of the two series and the latter correlation the association between change over time in the two series. We use this method to estimate the correlation between condom use and intentions.

**THE FUNCTIONAL FORM OF TIME**

The functional form of the change over time can either be imposed or left free to be estimated (Burchinal & Applebaum, 1991). In cases where different variables will be compared, it probably is reasonable to impose a common functional form. Initial analysis showed that a nonlinear functional form (i.e., log weeks after the intervention) is superior in goodness of fit to a model of linear change in weeks. This is the time metric used for all analyses.
ESTIMATION WITH MISSING DATA

To estimate the effects of experimental status and mediating variables on the outcome, we used maximum likelihood methods in the structural equation modeling program AMOS (Hox, 1995; Kline, 1998b), which uses all available data without resorting to imputation, an approach that is generally superior both to listwise and pairwise deletion of missing data given the assumption that data are missing at random (Arbuckle, 1996; Enders, 2001; Graham, Hofer, Donaldson, MacKinnon, & Schafer, 1997; Little & Rubin, 1987, 1989-1990). However, because the follow-up data collection did not limit data collection to a particular sex partner type, variability of sex partner choice is a complicating factor here that leads to missing values when the available respondent changes sexual partner at a time subsequent to the baseline interview. In addition, missing values also occur when the respondent misses one or more of the follow-up interviews or does not have sex within the follow-up periods. In either case, psychosocial variables and/or self-reports of condom use are not available for analysis. However, an analysis of the data for predictors of missingness for the self-efficacy variables show virtually no relationship between either type of missing data, experimental status, or gender (Hennessy et al., 1999). Thus, we feel confident that the maximum likelihood approach is justified. In addition, the maximum likelihood approach to missing values during the follow-up period is more consistent with the intent-to-treat assumptions commonly used in randomized clinical trials because all postintervention follow-up data are utilized and no creaming of observations (e.g., selecting just respondents with complete longitudinal data or just respondents who finished the intervention) is done.

ASSESSING GOODNESS OF FIT

Assessment of model fit for simultaneous equation models is a complex issue. In general, $R^2$ values can be computed for each outcome, but when goodness of fit of the entire multiple measurement/structural equation system is the issue, there are many measures that vary along different dimensions (Tanaka, 1993). A $\chi^2$ test is commonly used to compare the predicted covariance matrix of the observed variables for the model with the actual covariance matrix.
Small values suggest only minor differences between the two matrices and therefore a good fit of the model to the data. However, $\chi^2$ is usually augmented by other measures that are less sensitive to sample size and that therefore represent an index of fit rather than a dichotomous decision rule (Hu & Bentler, 1995).

Therefore, two additional fit indices are used here, the root mean square error of approximation (RMSEA) and the Tucker-Lewis Index. The RMSEA is an absolute measure of fit where 0 represents perfect fit and positive departures reflect poorer fit. Values less than .05 are usually considered acceptable (Maruyama, 1998). In contrast, the Tucker-Lewis Index (also called the nonnormed fit index) is a comparison of two fit functions, the estimated covariance matrix and a matrix assuming no association between the observed variables. Positive departures from 0 reflect the relative advantage of the proposed model compared with the baseline model of no association (the Tucker-Lewis has no upper bound) (Kline, 1998a). The Tucker-Lewis Index also adjusts for parsimony and penalizes overfitting in a way that is superior to that of other null model indices (Marsh, Balla, & Hau, 1996). The Tucker-Lewis Index should be at least .90 to reflect an adequately fitting model (Kline, 1998a; Schumacker & Lomax, 1996), although there is controversy concerning the appropriateness of this arbitrary standard (Hu & Bentler, 1995).

**MEASURES**

Measures of the psychosocial mediators were as follows. Self-efficacy toward consistent condom use was measured by the question, “A lot of different things can get in the way of using a condom when people have sex. How sure are you that you can use a condom every time you have vaginal sex with your main partner?” This item was scored on a 1-to-7-point scale with 1 as *very sure I cannot* and 7 as *very sure I can*. Attitudes toward consistent condom use was measured by the average of six semantic differential items ranging from 1 to 7 with the stem, “Would you say that using a condom every time you have vaginal sex would be . . . .” The terms used were (1) *very unpleasant/very pleasant*, (2) *very unwise/very wise*, (3) *very bad/very good*, (4) *very difficult/very easy*, (5) *very unnecessary/very necessary*, and (6) *very uncomfortable/very comfortable*. The average for the respondent was used to ensure that all the psychosocial variables were in a comparable
1 to 7 metric. Subjective norms toward consistent condom use were measured by responses to the item, “Do most people who are important to you think you should or should not use a condom every time you have vaginal sex with your main partner?” This item was coded as a 1 to 7 scale with 1 as think I should not and 7 as think I should. Intentions toward consistent condom use was measured by the item, “How likely is it that from now on, for at least the next 6 months, you will use a condom every time you have vaginal sex with her/him?” This item was coded as a 1 to 7 scale with 1 as very unlikely and 7 as very likely. Condom use was computed as a percentage of condom-protected vaginal sex events (i.e., all condom-protected vaginal sex acts divided by all vaginal sex acts times 100). Anal sex was extremely rare in the participant population and is not considered here.

In contrast to the analysis of Kamb et al. (1998), which combined condom use behavior and other types of safer sex responses (e.g., abstinence) in their condom use outcome (see Figure 2 in Kamb et al., 1998), we counted persons who reported no vaginal sex during the 3 months as missing data, not as no unprotected sex. Nonetheless, the pattern of the results are similar for the two analyses—a clear pattern of condom use ranked by intervention group for the 3- and 6-month follow-ups that erodes during the last two follow-up periods. In this article, gender groups were analyzed simultaneously so that statistical tests across male and female groups could be performed.

ANALYSIS MODELS

To examine the correlation between intentions and behavior, we defined two growth curves—one for intentions and the other for self-reported behavior—and estimated the correlations between the parameters of each. That is, the intercept term of intentions-to-use-condoms variable was correlated with the intercept term of the condom-use-percentage variable, and the slope of change over time of intentions was correlated with the slope of change over time with condom use. These two correlations represent the correlation between initial values (e.g., at the immediate follow-up period) and the correlation between changes in intentions and condom use over time. The full model is shown in Figure 1. Note that the indicator variables are lagged because behavioral intentions always refer to the future. Thus, the intentions growth curve used data from the
immediate follow-up to the 9th month after the intervention, whereas the condom use curve used data from the 3rd to the 12th month after the intervention (lagging by 6 months reduces the available data too dramatically). We also controlled for correlations between the measurement errors of the different indicators at the same time (e.g., between e1 and e6).

The second use of growth curves included a structural equation model between experimental status, the mediator curve, and the intentions curve. The annotated model is shown in Figure 2. Parameters A and C are treatment effects of enhanced and brief counseling, respectively, on the average value (intercept) of the mediator variable at the immediate follow-up, whereas Parameters B and D are the corresponding treatment effects of the mediator’s slope of change over time (Figure 2 shows self-efficacy, but the mediator effects were also estimated for the two other psychosocial determinants of intentions as well).
Parameters E and F are the regression coefficients linking the average value (intercept) of the mediator to the intercept value of intentions and the change over time of the mediator to the change over time in intentions. The association between the mediator and the intentions slope and intercept (E and F) do not vary by experimental status because Project RESPECT was supposed to work through the mediating variables to change intentions. We were more concerned about whether causal mediators have low associations with intentions because attempting to alter a mediator that is poorly associated with the outcome is a waste of intervention resources and would suggest a revised intervention model. As with Figure 1, correlations between error terms (e.g., e1, e6, D1, D2) not shown for clarity.
RESULTS

STATISTICS ON CONDOM-PROTECTED VAGINAL SEX

Table 1 shows for each gender the proportions of condom-protected vaginal sex with the respondent’s main partner by experimental group. For females and their main partners, the interventions were effective in significantly increasing condom use in both enhanced and brief counseling compared with the control group for the first follow-up period and in HIV enhanced counseling for the second follow-up. For males, only HIV enhanced counseling showed significant differences at the first and second follow-up.

THE CORRELATION BETWEENIntentions AND CONDOM USE

Table 2 shows the correlations between the parameters of the intentions and condom use growth curves. The R1 and R2 correlations exceed .70 for each gender group. The gender difference between the estimates was not statistically discernable: The association between changes in intentions and changes in condom use did not vary by gender. It appears that the focus on behavioral intentions was justified.

### TABLE 1

<table>
<thead>
<tr>
<th></th>
<th>3 Months</th>
<th>6 Months</th>
<th>9 Months</th>
<th>12 Months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females (n = 1,860)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enhanced counseling</td>
<td>54%</td>
<td>53%</td>
<td>50%</td>
<td>49%</td>
</tr>
<tr>
<td>Brief counseling</td>
<td>48%</td>
<td>48%</td>
<td>49%</td>
<td>49%</td>
</tr>
<tr>
<td>HIV education</td>
<td>48%</td>
<td>46%</td>
<td>50%</td>
<td>47%</td>
</tr>
<tr>
<td>Males (n = 2,440)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enhanced counseling</td>
<td>56%</td>
<td>46%</td>
<td>48%</td>
<td>43%</td>
</tr>
<tr>
<td>Brief counseling</td>
<td>48%</td>
<td>43%</td>
<td>43%</td>
<td>41%</td>
</tr>
<tr>
<td>HIV education</td>
<td>46%</td>
<td>41%</td>
<td>45%</td>
<td>41%</td>
</tr>
</tbody>
</table>

NOTE: Condom-protected vaginal sex computed as the number of condom-protected vaginal sex acts as a proportion of all vaginal sex acts.

a. Statistically discernable difference from the HIV education control group.
b. Indicates one-tailed test.
TREATMENT EFFECTS ON MEDIATORS

Table 3 shows the treatment effects on the parameters of the mediator growth curves and the associations between the mediator curve and intentions for consistent condom use. We discuss the results for self-efficacy in detail and summarize the results for the other two mediators.

In HIV education (the control condition), the mediator self-efficacy growth curve was estimated as: males: $5.72 - .11\text{ (time)}$ and females: $5.65 - .06\text{ (time)}$. For females, both enhanced and brief counseling resulted in higher average values (intercepts) of self-efficacy at the IFU (Estimates A and C in Table 3, self-efficacy section) but had no effects on the change over time (slopes) compared with the control condition (Estimates B and D in Table 3, self-efficacy section). For males, only HIV enhanced counseling had a statistically discernable effect on the average value of self-efficacy, and the treatment effects of HIV enhanced counseling were not different between the genders (e.g., .28 for females and .22 for males).

Combining the increments to intercepts and slopes due to experimental status with the control group equations resulted in equations modeling self-efficacy over time as:

HIV enhanced counseling: $\delta 5.94 - .14\text{ (time)}$; $\varphi 5.93 - .09\text{ (time)}$

HIV brief counseling: $\delta 5.71 - .14\text{ (time)}$; $\varphi 5.84 - .08\text{ (time)}$

HIV education: $\delta 5.72 - .11\text{ (time)}$; $\varphi 5.64 - .06\text{ (time)}$

### TABLE 2
Correlations Between Average Values and Changes Over Time in Intentions and Condom-Protected Vaginal Sex—Main Partner

<table>
<thead>
<tr>
<th></th>
<th>Females</th>
<th>Males</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept-intercept correlation (correlation between values at immediate follow-up)</td>
<td>.78</td>
<td>.76</td>
</tr>
<tr>
<td>Slope-slope correlation (correlation between changes over time)</td>
<td>.72</td>
<td>.77</td>
</tr>
</tbody>
</table>

$\chi^2 = 308, df = 50, p < .05$; root mean square error approximation = .035, Tucker-Lewis Index = .989

NOTE: All correlations discernable from 0 at the .05 level or less. The difference between the corresponding correlation in each gender group is not discernable from 0. See Figure 2 for the growth curve model.
Note that HIV enhanced counseling males had higher intercepts (5.94) than other males, whereas females in enhanced and brief counseling had higher intercepts of self-efficacy (5.93 and 5.84, respectively) than females in HIV education (5.64) immediately after the
intervention is over. The slope estimates show that males declined faster from their initial values in all treatment conditions (slopes of −.14 and −.11) compared with the observed decline for females (−.09, −.08, −.06 in each experimental group), although none of these slopes’ differences were significantly different from the declines estimated in HIV education for either gender.

Finally, the results for the mediator/intentions parameters (E and F in Table 3, self-efficacy section) show that self-efficacy slopes and intercepts were significantly associated with intentions; a one-unit change in self-efficacy produced an essentially equivalent change in intentions for both males and females.

In HIV education (the control condition), the mediator attitude growth curve for each gender was estimated as: (1) males: 5.48 – .05(time) and (2) females: 5.84 – .03(time). Females in both counseling interventions had higher intercepts, and females in HIV brief counseling declined significantly faster than those in HIV education after the intervention was over. However, only HIV enhanced counseling males showed a treatment effect for attitude change: They were .39 higher on average after the intervention was over. Males in both enhanced and brief counseling declined from their initial levels faster than the males in the control condition.

In HIV education (the control condition), the mediator norm growth curve for each gender was estimated as: (1) males: 5.90 – .10(time) and (2) females: 6.16 – .03(time). Note that these initial (i.e., intercept) values were very high even for the control group (recall that the scale went from 1 to 7). This suggests the possibility of ceiling effects for the norm outcome. Only females in HIV enhanced counseling showed a discernable average difference at the end of the intervention in both intercept and change over time.

In general, Project RESPECT altered the attitude and self-efficacy mediators for females in both enhanced and brief counseling. For males, however, the intervention changed attitudes and self-efficacy only for HIV enhanced counseling. In addition, changes in attitudes, self-efficacy, and norms were associated with changes in intentions. However, Project RESPECT was generally unsuccessful in affecting subjective norms toward consistent condom use with main partners for all experimental groups and genders except for females in HIV enhanced counseling, although changes in norms were related to changes in intentions for both genders.
CONCLUSIONS

It is difficult to refine or tailor intervention programs when the theoretical variables underlying program operation are not investigated. Project RESPECT is unusual in that it combined a rigorous research design and comprehensive data collection protocol with a detailed and measurable theoretical foundation. Because of the level of measurement intensity and the explicitness of the underlying theory, it is possible to investigate both the empirical plausibility of the underlying assumptions and how the prevention counseling program operated in practice using structural equation modeling (Hennessy & Greenberg, 1999).

The focus on changing intentions to use condoms consistently was justified. The correlations between changes in condom use intentions and condom use self-reports were high and statistically equivalent for both genders, a finding that is consistent with the underlying operational hypotheses of Project RESPECT and its emphasis on changing intentions as the critical determinant of behavior.

The fact that male and female participants had equivalent correlations between intentions and behavior is somewhat surprising. Because condom use is a behavior for men but condom negotiation the behavior for women, it is often assumed that women are less able to act on their condom use intentions than men (Fishbein et al., 2001). However, von Haeften, Fishbein, Kaspryzk, and Montano (2000) also reported that men and women were equally capable of acting on their intentions to use condoms with their main partners. They suggested, therefore, that goals may be appropriate outcomes in some intervention contexts. In addition, these findings imply that the issue of inequitable power relationships between male and female sex partners might be overemphasized.

Although men and women appear equally capable of acting on their intentions to use condoms with their main partner, it does appear that the intervention as a psychosocial phenomenon operated differently for men and women. Women in both prevention counseling interventions increased in attitudes and self-efficacy compared with women in the educational intervention. For men, in contrast, only enhanced counseling demonstrated discernable changes. Thus, for increasing condom use, both counseling interventions seemed to be effective for female participants, but the more intensive counseling was more effective
for men in affecting the range of important mediating psychosocial variables.

As reported elsewhere (e.g., Fishbein & Pequegnat, 2000; Pequegnat et al., 2000), there is no simple relationship between condom use and STD (or HIV) incidence because the change in STD incidence as a function of increased condom use will depend on many factors, including the prevalence of the STD in the population being considered, the sexual mixing patterns in that population, the degree of infectivity of the donor, characteristics of the host, and the type and frequency of sexual practices. Nevertheless, there is evidence that although correct and consistent condom use has little effect on STD incidence among those at low risk for acquiring an STD, correct and consistent condom use does significantly reduce incidence among those at high risk of acquiring an STD (DeVincenzi, 1994; Fishbein & Jarvis, 2000; Saracco, Musicco, & Nicolosi, 1993).

Despite the fact that there were no differences in STD incidence between the enhanced and brief counseling interventions, the differential effects of these on the participants’ condom use and the interpretable patterns of mediated causal influence for each gender are important findings that need to be taken into account in developing recommendations for HIV counseling and testing. For people for whom condom use is an appropriate and acceptable HIV/STD risk-reduction goal, it may be important to take into account the gender, risk status, and the relative influence of the theoretical mediators (e.g., attitudes, self-efficacy, and social norms) when choosing the best intervention strategy.

REFERENCES


This report compares 20th-century Canadian hospital and nonhospital location-of-death trends and corresponding population mortality trends. One of the chief findings is a hospitalization-of-death trend, with deaths in hospital peaking in 1994 at 80.5% of all deaths. The rise in hospitalization was more pronounced in the years prior to the development of a national health care program (1966). Another key finding is a gradual reduction since 1994 in hospital deaths, with this reduction occurring across all sociodemographic variables. This suggests nonhospital care options are needed to support what may be an ongoing shift away from hospitalized death and dying.

AUTHORS’ NOTE: The Statistics Canada mortality data were purchased through a Social Sciences Research grant, with the initial analysis of this data supported by a Support for the Advancement of Scholarship grant, both of which were obtained through the University of Alberta. Excellent data analysis support was provided by Corrine Truman. The larger investigation of barriers to palliative care, including an analysis of historic influences on location and death and this analysis of location of death trends, was supported by a National Health and Research Development Program operational research grant (No. 6609-2096-96).
Providing the right care in the right place is not simple. This is particularly true when a person is terminally ill or dying (Curtin, 1996). The seriousness of this concern is illustrated on one hand by considerable advances in both curative and noncurative, or palliative, care (Canadian Palliative Care Association, 1997; Turner et al., 1996) and on the other hand by an increasingly obvious requirement to use health services more wisely (Curtin, 1996; Gray, 1998; Kaplow, Charest, & Benaroya, 1998; Lavis & Anderson, 1996; Litwin & Lightman, 1996; National Forum on Health, 1997; Wolff & Schlesinger, 1998).

Some reports indicate hospital deaths are common in Canada (Ajemian, 1992; Fainsinger, Bruera, & MacMillan, 1997; Heyland, Lavery, Tranmer, Shortt, & Taylor, 2000) and in other developed countries (Brock & Foley, 1998; Costantini et al., 1999; H. Davis, Schoendorf, Gergen, & Moore, 1997; Field & James, 1993; Grande, Addington-Hall, & Todd, 1998; B. M. Thomas, Starr, & Whalley, 1997; Tolle, Rosenfeld, Tilden, & Park, 1999). Yet, hospital deaths can be problematic for a number of reasons. First and foremost is the concern that the care of terminally ill and dying persons is treatment oriented as opposed to comfort oriented (Ajemian, 1992; Curtin, 1996; Heyland et al., 2000; Wilson, 1997, 2000). Other concerns include the high cost of hospital care (Health Canada, 1996) and the opportunity costs of using acute care hospital beds for care that can be provided elsewhere (National Forum on Health, 1997).

As the care provided to dying persons is extensively influenced by factors inherent in the place of death and no description of Canadian location-of-death trends is available for planning or policy purposes, a research investigation was conducted to identify 20th-century Canadian locations of death and corresponding population mortality trends. This report compares hospital and nonhospital locations of death and corresponding population mortality trends.

**RESEARCH METHOD**

To compare Canadian location-of-death trends and corresponding population mortality trends, the researchers sought archived population mortality data. Although vital statistics databases were not specifically designed for the purpose of evaluating health facility utilization
and for supporting subsequent health services or health system reforms, mortality statistics have long been considered an important source of health information (G. G. Davis, 1997). Through such statistics, health officials have been informed of leading primary and secondary causes of death, public awareness has been raised over conditions leading to death, and health care policy and planning have been advanced (H. Davis et al., 1997; Laraque, Greene, Triano-Davis, Altman, & Lin-Greenberg, 1996; Nordenholtz et al., 1995; Samkoff, Hockenberry, Simon, & Jones, 1995; Shkolnikov, Leon, Adamets, Andreer, & Deer, 1998). Although some ongoing concern has been expressed about the validity and reliability of mortality records, these types of records frequently present the most accessible and complete source of population health data. A growing list of publications illustrates the diversity and usefulness of these databases as well as a current trend toward creating population databases by linking smaller ones (Backlund, Sorlie, & Johnson, 1997; Boyle & Dobson, 1995; Christenson & Johnson, 1995; Coulta & Hughes, 1996; Every et al., 1997; Gideon & Mannino, 1996; Goodin & Hanzlich, 1997; Guite & Burney, 1996; Hanzlich, 1996; Kraus, Peek, Silberman, & Anderson, 1995; Mackenbach, Kunst, Lautenbach, Oei, & Bijlsma, 1997; Maudsley & Williams, 1996; Messite & Stellman, 1996; Myers & Farquhar, 1998; Nordenholtz et al., 1995; Rushton & Romaniuk, 1997; Samkoff et al., 1995; Sheddon & Williams, 1997; B. M. Thomas et al., 1997; Wolleswinkel-Van den Bosch, Van Pollel, Tabeau, & Mackenbach, 1998; Wood, Sallar, Schechter, & Hogg, 1999).

To this end, a Statistics Canada computerized mortality database containing aggregate data on all deaths in Canada was purchased following research ethics approval, administrative approval of access by Statistics Canada, and notification of research grant funding (see Authors’ Note). Statistics Canada is a national government organization that gathers, stores, and often analyzes data on a wide range of subjects important to the health and well-being of Canada and its peoples. As this database only contained 1950-1997 data, 1900-1950 provincial data were sought. A second computerized database, identifying location of death for residents of British Columbia from 1927 through 1950, was obtained without charge on request from that province’s Department of Vital Statistics. A third database, identifying location of death for residents of Alberta, was constructed by the principal investigator during a search of archived 1905-1950 annual
hospital and vital statistics records. These provincial databases are limited to the annual incidences of hospital deaths and total deaths in province.

The Statistics Canada mortality database is a compilation of computerized death certificate data obtained on an annual basis from all Canadian provinces and territories. The data variables include year of death, province/territory where the death occurred, birthplace (Canada or other country of origin), gender, age at time of death, marital status, location of death (hospital or various other sites), and primary cause of death (International Classification of Diseases, ninth version [ICD-9] code). Data were obtained on 7,851,499 deceased persons, covering all deaths from 1950 through 1997, a span of 48 years. In most instances, data were complete or nearly complete. Location of death had 6.1% missing data; only marital status (6.7%) and birthplace (26.9%) had higher rates of missing data. The missing location-of-death data were primarily concentrated in certain years and provinces. Location of death was not reported by any province or territory in 1965, British Columbia did not report location-of-death data in 1961-1962 and 1993-1996, Quebec did not report location-of-death data in 1975-1977 and 1988-1989, and Newfoundland did not begin reporting data until 1953.

Another potential problem with the Statistics Canada and archived provincial data concerns the issue of a death registered in an acute care hospital or continuing care facility involving a stay of days or weeks or only minutes, as in the case of a person who is pronounced dead on or shortly after arrival. Although this is an important issue for quantifying the extent of utilization, resources are used whenever health facilities are accessed. The intention to use health care resources through seeking admission is a related consideration. Another issue is that the degree of inconsistency in reporting practices from province to province and by province over the years is not known.

For this reason, caution was taken not to overinterpret minor fluctuations in data. Analysis was also largely confined to simple descriptive statistics, with comparisons normally limited to $t$ tests (bivariate ratio or interval data), ANOVA (multivariate ratio or interval data), and chi-square analysis (ordinal or nominal data). The .05 probability standard was used in determining if differences between or among groups were statistically significant.
TABLE 1
Deaths in Canada, 1950-1997

<table>
<thead>
<tr>
<th></th>
<th>All Deaths</th>
<th>1950 Deaths Only</th>
<th>1994 Deaths Only</th>
<th>1997 Deaths Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>All deaths (N)</td>
<td>7,851,499</td>
<td>123,590</td>
<td>201,499</td>
<td>216,061</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>65.7</td>
<td>56.1</td>
<td>71.7</td>
<td>72.8</td>
</tr>
<tr>
<td>Median</td>
<td>72</td>
<td>66</td>
<td>76</td>
<td>76</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>23.0</td>
<td>28.5</td>
<td>18.3</td>
<td>17.7</td>
</tr>
<tr>
<td>Age (years) &lt; 65</td>
<td>34.4</td>
<td>46.3</td>
<td>24.5</td>
<td>22.4</td>
</tr>
<tr>
<td>Age (years) 65+</td>
<td>65.6</td>
<td>53.7</td>
<td>75.5</td>
<td>77.6</td>
</tr>
<tr>
<td>Birthplace</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In Canada</td>
<td>72.4</td>
<td>72.1</td>
<td>76.5</td>
<td>75.7</td>
</tr>
<tr>
<td>Outside Canada</td>
<td>27.6</td>
<td>27.9</td>
<td>23.5</td>
<td>24.3</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>56.4</td>
<td>56.4</td>
<td>52.9</td>
<td>51.7</td>
</tr>
<tr>
<td>Female</td>
<td>43.6</td>
<td>43.6</td>
<td>47.1</td>
<td>48.3</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>47.2</td>
<td>44.5</td>
<td>45.2</td>
<td>44.4</td>
</tr>
<tr>
<td>Widowed</td>
<td>30.7</td>
<td>26.2</td>
<td>34.4</td>
<td>35.8</td>
</tr>
<tr>
<td>Single</td>
<td>19.2</td>
<td>29.0</td>
<td>14.6</td>
<td>13.5</td>
</tr>
<tr>
<td>Divorced</td>
<td>2.5</td>
<td>0.3</td>
<td>5.2</td>
<td>5.5</td>
</tr>
<tr>
<td>Separated</td>
<td>0.4</td>
<td>0.0</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>46.9</td>
<td>45.5</td>
<td>45.2</td>
<td>44.4</td>
</tr>
<tr>
<td>Not married</td>
<td>53.1</td>
<td>55.5</td>
<td>54.8</td>
<td>55.6</td>
</tr>
</tbody>
</table>

NOTE: Numbers given are percentages unless otherwise noted. Missing data excluded from analysis.

FINDINGS

STATISTICS CANADA DATA

Total deaths. In keeping with a growing population and despite an ongoing decline in the mortality rate throughout the 20th century (Statistics Canada, 2000), the total number of deaths each year increased steadily. In 1997, there were 216,061 deaths, a 75% increase over 1950 (see Table 1). An increasing number of deaths in every province and territory was also found, with Ontario consistently having the highest number and proportion of Canadian deaths each year (38.0% over 48 years), followed by Quebec (23.9% over 48 years). The sparsely populated Yukon Territory and Northwest Territories
(now divided into the Northwest Territories and Nunavut) had the lowest numbers and thus proportions of Canadian deaths each year and over all 48 years (0.1%, respectively).

**Location of death.** Each year since 1950, more deaths were recorded as having taken place in hospitals than in any other single place or all other places combined (Table 2). Just more than two thirds of all deaths (68.7%) took place in hospitals over the 48-year period (missing data excluded from analysis). A long-standing hospitalization-of-death trend was also found. In 1950, hospitals accounted for 50.9% of death places, followed by an ongoing increase in incidence until 1994, when hospital deaths peaked at 80.5% (see Figure 1). A decline in hospital deaths followed 1994, with deaths in hospitals reaching 75.3% in 1997 from 78.0% in 1995 and 76.3% in 1996. Regardless of this decline, in 1997, three quarters of all deaths were recorded as having occurred in acute care hospitals (compared to 2.9% in continuing care facilities, 20.7% in all other known locations combined, and 4.4% in places either unknown or unrecorded).

Regardless of the recent decline in hospital deaths, there was a significant increase over the 48 years in the proportion of deaths in hospitals compared to deaths in all other locations combined (missing data excluded, \( p < .001 \)). Furthermore, linear regression analysis indicated that from 1950 to 1964, the slope of the line fitted to the percentage of deaths occurring in hospitals reflected an average yearly increase of 1.20% (95% confidence interval equals 1.05%, 1.35%). From 1966 to 1994, the slope of the line reflected an average yearly increase of only 0.45% (95% confidence interval equals 0.37%, 0.49%). These slope coefficients were substantially and significantly different from each other (\( p < .000 \)), indicating that the hospitalization-of-death trend was more pronounced from 1950 to 1964 than it was from 1966 to 1994 (see Figure 1).

Although an overall decline in hospital deaths followed 1994, provincial and territorial data show it began much earlier in some regions. Saskatchewan was the first province to record an ongoing decline in hospital deaths after hospital deaths peaked in 1981 at 70.5%. A decline in hospital deaths was subsequently observed in all other provinces and territories.

Considerable cross-Canada variation in the use of hospitals for death and dying purposes was also found. The Northwest Territories
**TABLE 2**
Location of Deaths in Canada, 1950-1997

<table>
<thead>
<tr>
<th></th>
<th>All Deaths (N = 7,363,454)</th>
<th>1950 Deaths Only (n = 108,060)</th>
<th>1994 Deaths Only (n = 199,255)</th>
<th>1997 Deaths Only (n = 206,620)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hospital</td>
<td>Nonhospital</td>
<td>Hospital</td>
<td>Nonhospital</td>
</tr>
<tr>
<td>All deaths</td>
<td>68.7</td>
<td>31.3</td>
<td>50.9</td>
<td>49.1</td>
</tr>
<tr>
<td>n</td>
<td>5,060,692</td>
<td>2,302,762</td>
<td>54,993</td>
<td>53,067</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>66.1</td>
<td>65.1</td>
<td>52.1</td>
<td>61.0</td>
</tr>
<tr>
<td>Median</td>
<td>72</td>
<td>71</td>
<td>63</td>
<td>70</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>22.6</td>
<td>23.7</td>
<td>29.3</td>
<td>26.9</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 65</td>
<td>66.3</td>
<td>33.7</td>
<td>58.8</td>
<td>41.5</td>
</tr>
<tr>
<td>65+</td>
<td>70.0</td>
<td>30.0</td>
<td>44.1</td>
<td>55.9</td>
</tr>
<tr>
<td>Birthplace</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In Canada</td>
<td>69.7</td>
<td>30.3</td>
<td>47.8</td>
<td>52.2</td>
</tr>
<tr>
<td>Outside Canada</td>
<td>72.3</td>
<td>27.7</td>
<td>58.8</td>
<td>27.9</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>68.4</td>
<td>31.6</td>
<td>52.8</td>
<td>47.2</td>
</tr>
<tr>
<td>Female</td>
<td>69.2</td>
<td>30.8</td>
<td>48.4</td>
<td>51.6</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Marital status</th>
<th>All Deaths (N = 7,363,454)</th>
<th>1950 Deaths Only (n = 108,060)</th>
<th>1994 Deaths Only (n = 199,255)</th>
<th>1997 Deaths Only (n = 206,620)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hospital</td>
<td>Nonhospital</td>
<td>Hospital</td>
<td>Nonhospital</td>
</tr>
<tr>
<td>Married</td>
<td>54.0</td>
<td>46.0</td>
<td>48.4</td>
<td>51.6</td>
</tr>
<tr>
<td>Widowed</td>
<td>49.9</td>
<td>50.1</td>
<td>44.3</td>
<td>55.7</td>
</tr>
<tr>
<td>Single</td>
<td>65.9</td>
<td>34.1</td>
<td>60.2</td>
<td>39.8</td>
</tr>
<tr>
<td>Divorced</td>
<td>66.6</td>
<td>33.4</td>
<td>68.2</td>
<td>31.8</td>
</tr>
<tr>
<td>Separated(^a)</td>
<td>78.8</td>
<td>21.2</td>
<td>78.8</td>
<td>21.2</td>
</tr>
</tbody>
</table>

NOTE: Numbers given are percentages unless otherwise noted. Missing data excluded from analysis.
\(^a\) Data from 1951-1953 are missing.
and Yukon Territory had the lowest incidences of hospital deaths over 48 years (34.7% and 53.2%, respectively), whereas British Columbia and New Brunswick had the highest (76.3% and 71.6%, respectively). In 1997, the Northwest Territories again had the lowest annual rate (39.0%), followed by Alberta (55.7%), whereas Quebec and British Columbia had the highest (87.4% and 77.7%, respectively).

STATISTICS CANADA SOCIODEMOGRAPHIC POPULATION DATA

As indicated previously, Statistics Canada variables are limited to year of death, province/territory where the death occurred, birthplace (Canada or other), gender, age, marital status, location of death, and primary cause of death (ICD-9 code). As presented next, each of these variables was compared on the basis of hospital or nonhospital location of death, with both annual and combined 48-year comparisons made. Corresponding population data analysis findings follow.

Age. The average age of all persons who died in hospitals or other locations over the 48 years varied significantly ($p < .001$). Yet, the people who died in hospitals were only 1 year older on average (see Table
2). Some age-based variance from year to year was noted, however. From 1950 to 1963, persons younger than age 65 had a higher rate of hospital deaths each year than persons aged 65 and older (seniors). After 1963, seniors had a higher annual rate of hospital death than younger persons did. Regardless, hospital deaths peaked in 1994 for both younger persons and seniors (see Table 2), with the incidence of hospital deaths then declining for both groups. The reduction from 1994 through 1997 was greater for seniors than it was for younger persons (5.3% and 4.7%, respectively).

Corresponding population mortality data analysis was also revealing. Over 48 years, the average age of all deceased persons increased significantly ($p < .001$). As indicated in Table 1, there was a 30% increase from 1950 to 1997 in the annual average age at time of death. The average age over all 48 years combined was only 65.7, yet 65.6% of all deaths were seniors (see Table 1). Deaths of seniors increased steadily from 53.7% of all deaths in 1950 to 77.6% in 1997 (see Table 1). Population aging was also illustrated by a mode age increasing from infancy (birth to age 1) for the years 1950 through 1976 to 74 years of age in 1977 and then gradually upward to 82 years of age in 1997.

**Birthplace.** Birthplace was also related to location of death. Over all 48 years combined, persons born outside of Canada had a significantly higher rate of death in hospitals than persons who were born in Canada ($p < .001$). Yet, hospital deaths were common to both groups; as illustrated in Table 2, a difference of only 2.6% separated them. The incidence of hospital deaths increased for both native-born and foreign-born persons until 1994, when 80.5% of native-born Canadian deaths were taking place in hospitals compared to 79.4% of foreign-born Canadian deaths (see Table 2). The incidence of hospital deaths declined for both groups after 1994 (see Table 2), with foreign-born persons having a greater decline in hospital deaths (6.9% compared to 5.0% for native-born Canadians). It is also relevant to note that most (72.4%) deaths over the 48 years involved persons who were born in Canada (see Table 1), and that the proportion of deaths of native-born persons increased significantly over the years ($p < .001$).

**Gender.** Gender was also related to location of death. Over all 48 years combined, females died significantly more often in hospitals
then men did \( (p < .001) \). This gender difference was less than 1%, however (see Table 2). Furthermore, females did not have the highest rate of hospital deaths every year. Males had higher rates of hospitalized death in the years 1950-1955, 1987-1989, and 1991-1997. Hospital deaths peaked in 1994 for both males and females (see Table 2). Following 1994, hospital deaths declined more rapidly for females (5.7%) than for males (4.5%).

This information should also be contrasted with the finding that there are more deaths of males than females (see Table 1), a gender difference that was statistically significant each year \( (p < .001) \) and over all 48 years combined \( (p < .001) \). Yet, the gap in incidence of deaths by gender narrowed over time, most particularly during the 1990s. More specifically, a gender gap of 12.8% existed in 1950 (the greatest gap); by 1997, this gap had reduced to 3.8% (see Table 1).

**Marital status.** Marital status was also related to place of death. As shown in Table 2, persons who were separated at the time of death had the highest rate of hospitalized death over all 48 years combined (78.8%), followed by divorced and single persons. Widowed persons had the lowest rate of hospital deaths, although approximately half of these deaths still took place in hospitals. Differences in hospital versus nonhospital location of death rates by marital status were statistically significant each year \( (p < .001) \) and over all 48 years combined \( (p < .001) \).

When the categories of single, divorced, and widowed were combined and married and unmarried persons were then compared in regard to their respective hospital utilization over 48 years (see Table 2), married persons had a higher rate of hospital deaths than unmarried persons did (a difference of only 4.8%, however). This difference was significant each year \( (p < .001) \) and over all 48 years combined \( (p < .001) \). The hospital death rate peaked in 1994 for both married and unmarried persons (see Table 2), with a higher rate of hospitalization again among married persons. Following 1994, there was a slightly faster rate of decline in hospital deaths among unmarried persons (5.3%) than among married persons (5%). As of 1997, married persons continued to have a higher rate of hospital deaths than unmarried persons did.

Furthermore, more deaths over the 48 years involved married persons than widowed, single, divorced, or separated persons (see Table
1). However, when the categories of single, divorced, and widowed were combined for analysis, significantly more ($p < .001$) deaths over the 48 years involved unmarried persons (52.8%) than married persons (47.2%). This population pattern was also evident each year ($p < .001$).

*Cause of death.* The location of death also varied by cause of death. There were 999 distinct causes of death recorded over the 48 years (each a separate ICD-9 code). These ICD-9 codes were grouped into 10 commonly reported categories for analysis. The highest rate of hospitalized death over all 48 years combined was among persons who died of congenital (birth defects or congenital malformations present at birth) disorders (87.0%), followed by chronic liver diseases (84.8%) and cancer (81.1%). The lowest rate of hospitalized death and the only cause of death for which fewer than half of all deaths over the 48 years occurred in hospitals was for persons who died of injuries or suicide (40.6%). Each year, with the exception of 1950 through 1955 (when cardiovascular and respiratory disease rates were lower), deaths by injury or suicide were the least likely to occur in hospitals. Similarly, congenital disorders had the highest incidence of hospital deaths each year.

Regardless, the rate of hospitalized death increased gradually each year until the 1990s for all 10 cause-of-death categories. Hospitalized infectious and parasitic disease deaths peaked first, in 1991, at 83.2% of the total, followed by hospitalized cancer deaths, which peaked in 1993 at 85.2%. In 1994, the hospitalized death rate peaked for the 8 other cause-of-death categories (ranging from 61.0% for injuries and suicides to 95.0% for congenital disorders). An ongoing decline in hospital death rates was subsequently noted after the peak for all 10 cause-of-death categories. The decline in hospital death rates from 1994 though 1997 was greatest for the *all other* category (from 80.2% to 71.8%, or 8.4%), followed closely by the nervous system/sense organ category (from 73.9% to 66.2%, or 7.7%). The smallest decline was for infectious and parasitic diseases (from 80.3% to 79.6%, or 0.7%), followed by congenital disorders (95.0% to 92.6%, or 2.4%) and respiratory disorders (from 81.5% to 78.7%, or 2.8%). In 1997, the lowest hospital death rate remained among those persons who died of injury or suicide (56.5%), whereas the highest hospital death rate...
remained among those persons, mainly infants and children, who died of congenital disorders (92.6%).

It is also notable that nearly half of all deaths over the 48 years were considered to be due to cardiovascular diseases (45.2%). Although the incidence of death by cardiovascular diseases declined considerably over time, from 45.5% of all deaths in 1950 to 36.8% in 1997, after peaking in 1966 and 1967 at 50.2%, it remained the most common cause of death in 1997 (36.8%). The incidences of death from infectious and parasitic diseases (from 3.1% of deaths in 1950 to 1.2% of deaths in 1997) and all other disorders (from 12.1% in 1950 to 10.9% in 1997) also declined. The greatest decline, however, was among congenital disorders (from 8.7% of deaths in 1950 to 0.9% in 1997). Over the same period of time, the incidences of other disorders increased: cancer (15.1% to 27.3%), chronic liver (0.5% to 1.9%), diabetes (1.3% to 2.7%), nervous system and sense organ disorders (1.8% to 3.3%), respiratory disorders (6.6% to 9.4%), and injury or suicide (5.3% to 5.7%).

ALBERTA AND BRITISH COLUMBIA ARCHIVED DATA

Archived pre-1950 British Columbia and Alberta provincial data similarly illustrate a hospitalization-of-death trend. In 1927, the first year computerized location-of-death data were available for British Columbia, 43% of all deaths were recorded as having occurred in hospitals. A steady increase in hospital deaths followed; by 1950, 61% of all deaths in British Columbia were recorded as having occurred in hospitals. British Columbia hospitals in those years were defined as institutions providing inpatient medical and/or surgical care as well as care for persons suffering from tuberculosis and mental diseases. Institutions that only provided custodial or domiciliary care were not considered hospitals.

Hospitals were similarly defined in early 20th-century Alberta government documents. Total hospital deaths and total deaths in Alberta did not begin to be recorded until 1930. In 1930, 39% of all deaths in Alberta took place in hospitals. Following 1930, a steadily increasing number and percentage of deaths took place in hospitals. By 1950, 51% of all Alberta deaths were recorded as having occurred in hospitals.
DISCUSSION AND IMPLICATIONS

Despite some database and data limitations, the 1950-1997 Statistics Canada mortality data and pre-1950 archived provincial data provide useful information. One of the most important findings was a pronounced and sustained hospitalization-of-death trend across all variables. By the middle of the 20th century, hospitals were the most common site of death in Canada, with the hospital rate increasing until 1994. Although statistically significant population differences were found when hospital and nonhospital death rates were compared, these do not clearly distinguish location of death. The descriptive comparisons of location of death for married and unmarried persons, for young and old persons, and for all other tested variables revealed more similarities than dissimilarities in the use of hospitals as death places.

Interestingly, the rise in hospital death rates was more rapid prior to the 1966 passage of the federal Medicare Act that created universal access to hospital, diagnostic, and medical care than it was after 1966 when free hospital care was assured to all Canadians. This finding suggests that rising hospital utilization contributed to the development of Canada’s Medicare system.

In short, and with the possible exception of deaths due to injury or suicide, hospital deaths have been common across Canada during much of the 20th century. All other sites, most particularly continuing care facilities, are clearly secondary in incidence and thus significance. As much of the care immediately preceding death also takes place in hospitals, with this care of short or long duration, it is apparent that hospitals assumed crucial social and health care roles in regard to death, dying, and end-of-life care.

Among other considerations, it is important to determine if hospitals are appropriate places for end-of-life care. In Canada, there has been a declining number of acute care beds despite population increases since the 1980s (Tully & Saint-Pierre, 1997). During the mid-to late 1990s, 25% of acute care beds across Canada were closed to reduce government expenditure (Wilson, 1996). Using scarce beds for care that can be provided elsewhere is a concern. Another concern is that a literature review found that the availability of hospital beds influences their use for end-of-life care purposes (Grande et al., 1998). Tolle et al. (1999) directly reported that the “use and availability of
beds in acute care hospitals has been confirmed to be the principal determining factor in location of death” (p. 681).

In Canada, hospitals are not often or openly considered by the public or by the nurses, physicians, or administrators who work there as places of death and dying (Roy, 1999). Only approximately 1,000 of the 135,000 current acute care hospital beds across Canada are dedicated to palliative care (Canadian Healthcare Association, 1998; Health Canada Working Group, 1997). Although palliative care can be provided in any hospital bed, the majority of hospitals in Canada today do not have palliative care teams or palliative care programs to plan care for dying persons (Health Canada Working Group, 1997; Roy, 1999). This, along with concerns that end-of-life care in hospitals is often treatment oriented (Ajemian, 1992; Curtin, 1996; Heyland et al., 2000; Wilson, 1997, 2000), raises the need for investigations of the efficacy of hospital-based end-of-life care.

It is also relevant to note that some deaths in hospitals are unexpected. Sudden heart attacks, strokes, and other severe disorders, which are currently responsible for an estimated 20% to 25% of deaths in Canada (Heart and Stroke Foundation, 1998; S. P. Thomas & Hrudey, 1997), may unexpectedly cause death in hospitals. However, deaths in Canada and other developed nations are increasingly less premature and unexpected (Kane, 1996). Since the mid-1950s, an increasing proportion of deaths in Canada has resulted from progressive noncurable or chronic conditions and aging as opposed to acute curable illnesses (Statistics Canada, 2000). Given the high rate of death in Canadian hospitals, many hospital deaths are likely to be anticipated and unpreventable. In these cases, end-of-life care could take place in another location. Alternatively, dying needs to be openly acknowledged and better addressed in hospitals.

The relatively recent but notable reduction in hospital deaths across Canada is another finding of interest. Although death still takes place more often in hospitals than in all other places combined, a reduction in hospital deaths is important to monitor and understand, if not foster. From 1994 to 1997, this reduction involved all persons who died, regardless of province of residence, age, gender, or any other available sociodemographic characteristic. Yet at the same time, it was more pronounced among females, seniors, unmarried persons, persons who were not born in Canada, and persons suffering from nervous system/sense organ (including multiple sclerosis and amyotrophic lateral
sclerosis) or other disorders (including failure to thrive or undiagnosed conditions). These people are normally considered more vulnerable to hospitalization. In contrast, Grande et al.’s (1998) literature review found persons who die at home and access palliative home care are typically younger, male, socioeconomically advantaged, and possessing informal caregiver networks. Other studies suggested similar patient-related factors as influences on location of death (Brock & Foley, 1998; Costantini et al., 1999; H. Davis et al., 1997; Field & James, 1993; Grande et al., 1998; B. M. Thomas et al., 1997). A personal desire to avoid hospital care, as established by Tolle et al. (1999) through interviewing Oregonians, could also influence location of death. Sadly, Grande et al.’s (1998) literature review found few people who want to die at home are able to do so.

Regardless, the reduction since 1994 in hospital deaths raises the issue of nonhospital support for dying persons. Concern over an unfunded or unsupported shift of care to the home is already apparent in Canada and other developed countries (Robinson, 1997). An increase in the number and scope of community-based palliative care programs is occurring in Canada, although these programs are already said to be inadequate for meeting current needs (Health Canada Working Group, 1997). Similarly, home care has been deemed inadequate for meeting the current personal care needs of community-dwelling Canadians (Wilkins & Park, 1998), let alone future needs of an aging population (Statistics Canada, 2000).

In conclusion, three national and provincial mortality databases provided a useful understanding of historic and current mortality trends. The findings of this investigation may be useful for forecasting or perhaps mitigating the future. For this reason, the recent decline in hospital deaths across Canada, following what can only be described as a pronounced and long-standing hospitalization-of-death trend, is of particular interest. Whether or not this reduction signals the beginning of a continuing shift away from hospitalized death should become a major focus of attention. It is possible that a shift away from the hospital is occurring with little or no direct intervention. It is more likely, however, given the long-standing practice of hospitalized death and limited support for community-based home palliative care in Canada (Chochinov & Kristjanson, 1998; McWhinney, Bass, & Orr, 1995) that a concerted effort to support nonhospital dying and death will be needed in the challenging years ahead.
REFERENCES


This study assesses the effects of the Healthwise Communities Project (HCP) on use of self-care resources and health care utilization. The intervention included the distribution of the Healthwise Handbook, the provision of a telephone advice line, and a Web site. All of these products use a symptom-based approach and are aimed at a general population. A quasi-experimental design was used with two comparison communities. Measurements over time assessed the effects of the HCP while controlling for secular trends. Survey and utilization data are used to assess the effect of the intervention. Findings indicate that the community intervention increased the use of self-care resources. Users believe that these products help them make better decisions regarding when to seek care and how to self-treat problems. Most believe that using the self-care resources saved them from seeking unnecessary care. The findings from the utilization data provide some evidence to support this conclusion.

THE IMPACT OF A COMMUNITY-WIDE SELF-CARE INFORMATION PROJECT ON SELF-CARE AND MEDICAL CARE UTILIZATION

JUDITH H. HIBBARD
University of Oregon
MERWYN GREENLICK
Oregon Health Sciences University
HOLLY JIMISON
Center for Health Research, Portland, Oregon
JEFFREY CAPIZZI
Multnomah County Health Department, Portland, Oregon
LYNN KUNKEL
Oregon Health Sciences University

AUTHORS' NOTE: Funding for this project was provided by the Robert Wood Johnson Foundation.
Self-care interventions involve the dissemination of health information to empower consumers to make more appropriate health care decisions. Evaluations of self-care interventions measure the effect of this information on consumers and their health care utilization. Whereas most self-care interventions have focused on employee or health plan populations, the Healthwise Communities Project (HCP) is unique in that it is a community-wide intervention. As there are more opportunities for reinforcing messages via the formal and informal channels in the community, a community-wide intervention has the potential for a greater effect than one limited to a single site or population group. The theory is that a community-based intervention provides a more sustained effect on a large segment of the population (Thompson, Wallack, Lichtenstein, & Pechacek, 1991). This study examines the effect of the community-wide intervention on consumer use of self-care information, self-care behavior, and utilization of medical care.

GOALS OF THE INTERVENTION

The goal of the evaluation was to assess effects of the HCP on use of self-care resources, self-care behavior, health care utilization, health care costs, and satisfaction with care in the target population. The intervention was carried out by Healthwise Inc., a nonprofit organization that produces information and decision-support products to aid consumers in managing their own health. Their *Healthwise Handbook* is a widely used self-care book. The intervention was funded by a grant from the Robert Wood Johnson Foundation. The self-care resources, including the *Healthwise Handbook*, a telephone advice line, and a Web site, all use a symptom-based approach and are aimed at a general population. These self-care products provide decision support about how to handle a symptom or problem (e.g., earaches in children), delineate when it is appropriate to seek care or engage in self-care, and if self-care is called for, give instructions on the self-care measures to take. Preventive information is also included along with more general consumer information, such as how to work in partnership with the physician.

The intervention was intended to arm consumers with the necessary skills and resources to be more active participants and managers of their own health and health care. The assumption was that increasing
access to self-care resources would activate consumers to make more informed choices that would in turn decrease unnecessary utilization of health care and reduce costs. The intervention was also based on the assumption that increasing patients’ role in medical decisions would increase patient satisfaction.

Major employers in the community were partners in the intervention, as were insurers and health care providers in the community. Community partners were involved in the planning and provided resources to support the intervention. The project was aimed at 250,000 residents of four southwest Idaho counties and ran from May 1996 through December 1998.

BACKGROUND

To date, almost all self-care programs have focused on interventions aimed at a worksite population or a population enrolled in a health care delivery system. A recent assessment of the efficacy of self-care manuals, funded by the Robert Wood Johnson Foundation, summarizes the evidence (Carney et al., 2000). After searching the literature, 14 randomized trials were found to be relevant to effectiveness and were evaluated and summarized (Anderson, Morrell, Avery, & Watkins, 1980; Bertakis, 1986; B. L. Hansen, 1995; B. W. Hansen, 1990; Kemper, 1982; Moore, LoGerfo, & Inui, 1980; Morrell, Avery, & Watkins, 1980; Rasmussen, 1989; Smith, Pina, Carnagey, & Toothaker, 1997; Terry & Pheley, 1993; Usherwood, 1991; Vickery et al., 1983; Vickery, Golazewski, Wright, & Kalmer, 1988, 1989). Outcomes in these trials were health care utilization and/or cost (data derived from administrative or medical records). In some studies, there was evidence that the distribution of self-care manuals reduced physician utilization, and in other studies, the distribution had no apparent effect. Out of 14 randomized trials examining the effect of the dissemination of self-care manuals on costs and/or health care utilization, about half showed reductions in either cost or utilization or both. The remaining showed no effect or an effect opposite to that hypothesized. Thus, there is evidence that these self-care interventions have the intended effects at least some of the time. However, the factors associated with achieving the desired outcomes are less clear.

In this study, we assessed the degree to which a community intervention was successful in increasing the use of self-care resources, the
factors associated with use, and the effects on different types of utilization. The study questions were:

- Is there greater access to and use of self-care products in the intervention community?
- Who is the most likely to use self-care products?
- How are the self-care products used?
- What are the reported effects of using self-care resources?
- Are these reported effects greater in the intervention community?
- Is there any evidence that use of the self-care products reduces overall utilization?

METHOD

DESCRIPTION OF THE INTERVENTION

The Healthwise Communities Project included several key elements, including the following:

- A Healthwise Handbook was mailed to every household in a four-county area surrounding Boise (112,000 households).
- A Healthwise Web site (an expanded version of the manual) included more information on different treatment options and what is known about their efficacy and outcomes.
- A telephone nurse advice line was also set up. The nurses used the information from the Web site and the manual to advise callers.
- Resource centers were set up to provide access to the Web site, manuals, and other health education materials. Many of these centers were at work sites, health care delivery sites, and libraries.
- Workshops for health care providers were conducted on how to integrate the self-care resources into clinical practice.
- Workshops for consumers, often at their place of work, were provided on how to use the self-care resources.
- A media campaign ran throughout the intervention period and included billboards, radio, newspaper ads, magnets mailed to households, and weekly newspaper articles.

EVALUATION DESIGN

A quasi-experimental design was used, with one intervention community and two comparison communities and nonrandom assignment
to conditions. Multiple measurements over time assessed the effects of the HCP in the intervention community while controlling for the secular trends observed in the comparison communities. The design included process, effect, and outcome measurements. The two comparison communities were chosen because they bracketed the intervention community demographically in terms of population density and in particular, in the degree of managed care penetration. The intervention community consisted of the four counties surrounding Boise, Idaho. The two comparison communities were Billings, Montana, and Eugene/Springfield, Oregon. The evaluation components included three waves of consumer surveys in each of the three communities, utilization data collected from insurers and hospitals in the three communities, and a process evaluation, including focus group and the use of archival data.

Process evaluation. The implementation of the intervention was monitored throughout the project. Also monitored were the environmental trends related to self-care in the intervention and two comparison communities. The goal of the process evaluation was to develop an understanding of the context of the intervention in Boise and environmental influences on self-care in all three communities. The results of the process evaluation served as a background for interpreting the quantitative results from the utilization and survey data.

Consumer surveys. To measure self-care resource use and behaviors, consumer surveys were conducted in both the intervention and comparison communities at baseline, 24, and 36 months postintervention. The design followed a panel throughout the observation period and augmented the panel with a sample of new community residents at 24 and 36 months. This created samples representative of the community at large. The 24-month survey was used as a midpoint assessment of the intervention components, and, finally, the 36-month survey allowed an assessment of postintervention effects. Response rates for each of the surveys are shown in Table 1. The analysis shown here focuses on the baseline data and the 36-month data. Most of the survey data are analyzed using bivariate analyses (e.g., chi-square or t tests).
Utilization data. The utilization analysis was undertaken to test the hypothesis that the HCP reduced the overall health care utilization and costs in the Boise, Idaho community. The two comparison communities, Billings, Montana, and Eugene/Springfield, Oregon, were selected to provide information on secular trends in utilization in the Northwest to help interpret the Idaho community data.

Utilization was assessed in all three communities from the beginning of 1994 through the end of 1998. This provided 2 full years of comparison data prior to the beginning of 1996 and 2 full years after the end of 1996. The HCP intervention was initiated in April 1996. Data were collected from the two major hospitals and from two major insurance plans in each of the intervention and comparison communities. These data were supplemented in the intervention community by data from a major corporation’s self-insured plan. For the final analysis, the data within each community were combined, with weighting by the number of members within the plans in order to develop a single estimate of the utilization of services per person per year (estimated monthly) in each community.

A set of measures was estimated for each community, including total charges, ambulatory visit charges, ambulatory visit rates, doctor office visit rates, emergency room visit rates, and a variable called TLAS (time-limited acute symptoms) visit rates. TLAS was defined by a set of International Classification of Diseases (ninth revision) diagnostic codes and included such conditions as minor burns, backaches, headaches, asthma, diarrhea, and such. TLAS conditions were hypothesized to be the most sensitive to the Healthwise intervention. During the baseline period, these visits made up 20% to 25% of the total ambulatory visits of each community.

<table>
<thead>
<tr>
<th></th>
<th>Baseline Completed</th>
<th>Baseline % Response Rate</th>
<th>24 Months Completed</th>
<th>24 Months % Response Rate</th>
<th>36 Months Completed</th>
<th>36 Months % Response Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>959</td>
<td>43</td>
<td>807</td>
<td>46</td>
<td>718</td>
<td>30</td>
</tr>
<tr>
<td>Comparison</td>
<td>1,954</td>
<td>67</td>
<td>1,801</td>
<td>69</td>
<td>1,617</td>
<td>69</td>
</tr>
</tbody>
</table>

TABLE 1
Response Rates for Consumer Surveys
A time series intervention approach was used to assess departures from ongoing utilization trends in the three communities. This technique, described by Box and Tiao (1975), employs dummy variables to measure the effect of an intervention based on an autoregressive integrated moving average (ARIMA) model of the preintervention pattern of utilization. Each observation in the analysis was a monthly mean rate, and inspection of each series of these observations suggested specific traits that defined the degree of serial correlations and short- and long-term cyclical patterns. ARIMA models of each preintervention series were based on autocorrelation function and partial autocorrelation function plots that defined the autoregressive and/or moving average processes best characterizing each series. A series that displayed a monotone increase or decrease was differenced as part of the ARIMA model (i.e., a monthly rate was subtracted from rate of the following month) so that the results would indicate if there were any disturbances apart from the ongoing trend. Models were assessed for fit by looking at the autocorrelations of residuals. The ARIMA model was then applied to the whole study period with a dummy variable associated with each month of the intervention. ARIMA analysis calculated beta coefficients for each of the dummy variables. A covariance matrix was then used to calculate a sum of betas and standard error, and a Wald test and associated $p$ value indicated whether any statistically significant postintervention disturbances occurred in the trend.

We observed intervention effects on utilization for two separate time periods. Period 1 is the first year following the intervention. Period 2 is the second year following the intervention. Other interventions have noted a 1-year time lag in observing effects (Carney et al., 2000). Observation Period 2 is designed to detect any possible lagged effects (see Table 2).

**RESULTS**

The overarching question addressed in the analysis was, What effects were observed that could be attributed to the HCP? We examine this larger question within several smaller questions. The first several questions are explored using the survey data, whereas the ques-
Utilization and survey data were analyzed both across the intervention and comparison communities and over the 3 years of the evaluation. Table 2 presents the annualized mean rates of utilization per 1,000 enrolled.

### TABLE 2

Annualized Mean Rates of Utilization Per 1,000 Enrolled

<table>
<thead>
<tr>
<th>Period</th>
<th>Idaho</th>
<th>Oregon</th>
<th>Montana</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total charges/1,000 enrolled year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>1,233,404</td>
<td>1,185,399</td>
<td>1,221,498</td>
</tr>
<tr>
<td>Period 1</td>
<td>1,491,590</td>
<td>1,168,900</td>
<td>1,054,881</td>
</tr>
<tr>
<td>Period 2</td>
<td>1,696,560</td>
<td>1,085,478</td>
<td>937,909</td>
</tr>
<tr>
<td>Total ambulatory charges/1,000 enrolled year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>619,986</td>
<td>584,169</td>
<td>597,553</td>
</tr>
<tr>
<td>Period 1</td>
<td>775,741</td>
<td>601,000</td>
<td>534,606</td>
</tr>
<tr>
<td>Period 2</td>
<td>863,761</td>
<td>566,542</td>
<td>506,034</td>
</tr>
<tr>
<td>Total ambulatory visits/1,000 enrolled year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>4,115</td>
<td>3,215</td>
<td>4,908</td>
</tr>
<tr>
<td>Period 1</td>
<td>4,789</td>
<td>3,028</td>
<td>4,098</td>
</tr>
<tr>
<td>Period 2</td>
<td>4,853</td>
<td>2,555</td>
<td>3,715</td>
</tr>
<tr>
<td>Total doctor office visits/1,000 enrolled year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>2,823</td>
<td>2,889</td>
<td>4,353</td>
</tr>
<tr>
<td>Period 1</td>
<td>3,248</td>
<td>2,700</td>
<td>3,619</td>
</tr>
<tr>
<td>Period 2</td>
<td>3,840</td>
<td>2,241</td>
<td>3,313</td>
</tr>
<tr>
<td>Total time-limited acute symptoms (TLAS) visits/1,000 enrolled year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>909</td>
<td>768</td>
<td>656</td>
</tr>
<tr>
<td>Period 1</td>
<td>1,020</td>
<td>705</td>
<td>571</td>
</tr>
<tr>
<td>Period 2</td>
<td>951</td>
<td>541</td>
<td>561</td>
</tr>
<tr>
<td>Total emergency room visits/1,000 Enrolled year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>135</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>Period 1</td>
<td>135</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td>Period 2</td>
<td>124</td>
<td>131</td>
<td></td>
</tr>
</tbody>
</table>


The intervention focusing on utilization effects relies more on the utilization data. Utilization and survey data were analyzed both across the intervention and comparison communities and over the 3 years of the evaluation.

**IS THERE GREATER ACCESS TO AND USE OF SELF-CARE PRODUCTS IN THE INTERVENTION COMMUNITY?**

The intervention community started with a higher base rate, and there was a significantly greater increase in the percentage of households reporting the possession of a self-care manual over the study period (see Table 3). Thus, there is evidence that the intervention
significantly increased the percentage of households in the intervention community with a self-care manual.1

Use of the self-care manuals also increased over the study period in the intervention community. The use of self-care manuals increased from 49% to 66% over the 3-year period of observation in the intervention community (see Table 4). Although there was also an increase in use observed in one comparison community (Oregon), the intervention community increase was significantly greater. There was almost no increase in the other comparison community. The increase in use of the manual in the intervention community was observed at both 24 months and at 36 months after the initiation of the intervention. It is noteworthy that this increase in usage was sustained for the full 3 years, particularly in light of the fact that the intervention period ended at 30 months.

Respondents in the intervention community reported significantly more frequent use of self-care manuals than those in the comparison communities. Among those using self-care manuals, the intervention respondents reported they had used it an average of 3.5 times in the last 6 months. Respondents in the comparison communities reported using a manual an average of 2.6 (Oregon) and 2.7 (Montana) times. Use of the telephone advice nurse stayed flat in all three communities over the observation period. Use of a computer or online program increased significantly for all three communities (see Table 4). Thus, the most apparent effect of the intervention was the increased use of self-care manuals.

<table>
<thead>
<tr>
<th>TABLE 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of a Self-Care Manual in Home (by Community)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percentage that has a health or medical self-care manual at home</th>
<th>Intervention Baseline</th>
<th>Intervention 36 Months</th>
<th>Intervention Difference</th>
<th>Comparison Baseline</th>
<th>Comparison 36 Months</th>
<th>Comparison Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td>90</td>
<td>20***</td>
<td></td>
<td>66</td>
<td>71</td>
<td>5</td>
</tr>
</tbody>
</table>

***p < .001 (paired t test).
WHO IS MOST LIKELY TO USE SELF-CARE PRODUCTS?

Using bivariate analysis (e.g., chi-square), we examined the socio-demographic characteristics of those using and not using self-care products. Self-care manual users were more often women, those married, and those with children in the household. Those with more than a high school education were more likely to use the manuals than those with high school or less. The advice line was more often used by those with poorer health status, those with children at home, and those who are married. Computer use for self-care information was unrelated to age or education. Those with private health insurance were more likely to use a computer for self-care information than those with no insurance.

HOW ARE THE SELF-CARE PRODUCTS USED?

About 90% of those who use self-care products reported that they were able to find the information they were looking for (see Table 5). About 42% of the manual users reported that the last time they used the manual, they did so to decide if they needed to see a doctor, whereas 28% ended up going to the doctor. That is, about two thirds of those using the manual to decide to go to the doctor ended up going.

### Table 4

Comparison of Self-Care Resource Usage Within Communities

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>36 Months</th>
<th>Difference</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health or medical self-care manual</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention</td>
<td>49</td>
<td>66</td>
<td>+17</td>
<td>**</td>
</tr>
<tr>
<td>Comparison</td>
<td>41</td>
<td>46</td>
<td>+1</td>
<td></td>
</tr>
<tr>
<td>Telephone advice nurse</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention</td>
<td>25</td>
<td>26</td>
<td>+1</td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>31</td>
<td>28</td>
<td>–3</td>
<td></td>
</tr>
<tr>
<td>Computer for health or medical information</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention</td>
<td>8</td>
<td>22</td>
<td>+14</td>
<td>**</td>
</tr>
<tr>
<td>Comparison</td>
<td>8</td>
<td>21</td>
<td>+13</td>
<td>**</td>
</tr>
</tbody>
</table>

**p < .01 (chi-square).
More respondents in the intervention community reported using a manual to decide whether to see a doctor than in the comparison communities (see Table 6). The advice nurse is more often used to decide whether or not to see a doctor than the computer or a self-care manual. About two thirds who do use the advice nurse for deciding actually seek care. Far fewer respondents indicated they used a computer program or online program to decide whether to seek care (20%).

**WHAT ARE THE REPORTED EFFECTS OF USING SELF-CARE RESOURCES?**

Only respondents who reported using a resource in the last 6 months responded to these items. It appears that the most widely reported effect of using a self-care resource was to reduce worry about a symptom or a health problem (80% to 81% for manual or advice-nurse users). Reducing worry was also the most important effect for
TABLE 7
Effect of Using a Health or Medical Reference Manual in the Last 6 Months (36-Month Data)

<table>
<thead>
<tr>
<th></th>
<th>Intervention</th>
<th>Comparison</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helped you to reduce your worry about a symptom or health problem?</td>
<td>81</td>
<td>73</td>
<td>ns</td>
</tr>
<tr>
<td>Helped you to self-treat a symptom or health problem?</td>
<td>73</td>
<td>58</td>
<td>*</td>
</tr>
<tr>
<td>Helped you to decide to visit a doctor?</td>
<td>45</td>
<td>46</td>
<td>ns</td>
</tr>
<tr>
<td>Saved a visit to the doctor?</td>
<td>51</td>
<td>39</td>
<td>**</td>
</tr>
<tr>
<td>Helped you to decide to visit an emergency room?</td>
<td>8</td>
<td>6</td>
<td>ns</td>
</tr>
<tr>
<td>Saved you a visit to the emergency room?</td>
<td>22</td>
<td>17</td>
<td>ns</td>
</tr>
</tbody>
</table>

*p < .05; **p < .01 (chi-square).

those using a computer or online program; however, it resulted in fewer reports of reduced worry than the other two self-care resources. The second most widely reported effect of using one of the self-care resources was that it helped the user to self-treat a symptom. Again, this was more often reported with regard to use of the manual (68%) and the nurse advice line (70%). Only 43% of computer users report that the program helped them to self-treat a symptom. The nurse advice line is the resource most likely to both help respondents decide to visit the emergency room and to save them a visit to the emergency room. Manual users were three times more likely to report that the manual helped save them a visit to the emergency room than it helped them to decide to visit the emergency room.

ARE THESE REPORTED EFFECTS GREATER IN THE INTERVENTION COMMUNITY?

The effects observed in Table 7 for manual use appeared to be greater in the intervention community than in the comparison communities. The intervention respondents were more likely to indicate that using a self-care manual helped them self-treat a symptom and saved them a visit to the doctor. The magnified effect observed in the intervention community for manual use does not occur for users of a nurse advice line or computer program. Thus, it appears that the intervention increased the use of manuals as well as the effects of using a manual.
There is evidence to suggest that the media campaign and the reinforcements associated with the HCP encouraged consumers to use the self-care manual. Using bivariate analysis (e.g., chi-square), we examined the effect of the media campaigns and reinforcements on use of the self-care products. Respondents in the intervention community were significantly more likely to report that a doctor or health professional had encouraged them to use a self-care manual than respondents in the comparison communities ($p < .001$). Those who had been encouraged by a health care professional were significantly more likely to have used that self-care resource ($p < .001$). Furthermore, those in the intervention community who had heard of the HCP were significantly more likely to use the computer ($p < .01$) or manual ($p < .001$) than intervention respondents who had not heard of the project.

The higher usage and the greater effects could also be related to the type of manual in use. Respondents in the comparison communities reported using a variety of self-care or medical reference books. In the intervention community, 83% of respondents reported using the *Healthwise Handbook*, whereas only 8% and 11% of the comparison community respondents indicate they used a *Healthwise Handbook*.

In addition to those discussed earlier, we also examined other hypothesized effects. For example, we assessed differences between the intervention and comparison communities in terms of confidence in self-care, preferences for shared decision making, and satisfaction with medical care. The intervention was designed to increase satisfaction, desire for shared decision making with providers, and satisfaction with care. There were no significant differences among the communities on any of these measures.

Furthermore, respondents in the intervention community were no more likely to report self-treatment for common symptoms than those in comparison communities. Five common symptoms for adults and five symptoms for children were evaluated in the study. These symptoms generate high rates of utilization in formal care. Although the frequency for experiencing one of these symptoms was similar across the communities, there were no differences in the frequency with which respondents sought formal medical care or self-treated these common symptoms.
IS THERE ANY EVIDENCE THAT USE OF THE SELF-CARE PRODUCTS REDUCES OVERALL UTILIZATION?

Total charges per person per year and ambulatory charges per person per year were quite similar in the three communities during the baseline period, approximately $1,200 and $600, respectively. Visits rates were a bit different, with Eugene generally being lower and Billings being higher, as might be expected given the managed care market situations. In ambulatory visits, Eugene was about 20% lower than Idaho, and Billings was about 20% higher at baseline (see Figure 1).

The differences between the intervention and comparison communities were in the nature of the trend in utilization over time. In the comparison communities, there was a straight, monotonic decline in total charges, ambulatory charges, ambulatory visits, and doctor office visits across the three time periods. This represents strong evidence of a regional secular trend of reducing medical care costs and visits. In the intervention community, there was a straight, monotonic increase in costs and visits during the same time period, with a net increase of 20% to 35% across the variables (see Figures 1 and 2).
There was a somewhat different picture for TLAS and emergency room visits. TLAS visits also showed a steady decline in the comparison communities. In the intervention communities, however, TLAS visits increased during the first follow-up year but decreased during the second period. They remained higher during that period than during the baseline and were higher in the intervention communities than in the other two comparison communities by a factor of nearly two. But they were lower in the second follow-up period than in the first. Emergency rooms visits definitely began to decline in the intervention group following the intervention, against an increase in Eugene. It was not possible to estimate the emergency room rates in Billings because of coding problems in the plans’ data set.

A time series intervention approach was used to assess the trends in the three communities, and this analysis indicated that there were changes in the utilization patterns in the intervention group component that could be consistent with an intervention effect. This methodology tested each indicator as a series of monthly observations for discontinuities from the baseline trend during the follow-up period. The data were tested against the hypotheses of an intervention effect beginning April 1996 and an intervention effect beginning April 1997.
The time series analysis indicated no discontinuity in the comparison communities’ rates. They both had a decreasing pattern that began during the baseline period and continued throughout the follow-up. This can be seen in the graphs for total charges, ambulatory visits, emergency room visits, and TLAS visits shown on the figure. In the intervention group, the rates for those three variables increased until mid- to late 1997. After that time, the rates began to decrease in the intervention group: TLAS visits \( p = .03 \), emergency room visits \( p = .00 \).

There were no discontinuities in the analysis of total charges or ambulatory charges. Hospital utilization data are not presented here. The data for hospital utilization indicated that the intervention community utilization did not change during the study period and had a pattern quite similar to the two comparison communities.

There are two possible interpretations for the findings of this section. The first interpretation is that there have been strong secular forces for reducing utilization in the Northwest, and that these forces did not affect the intervention community as soon as the comparison communities. These forces eventually came into play in the intervention community, and then the utilization patterns began to resemble those in other Northwest communities. This interpretation suggests there is no intervention effect.

An equally reasonable interpretation is that there was a trend in the Northwest toward lower utilization, and the implementation of the HCP helped to usher in that change in the intervention community. The implementation of the HCP tipped the balance. Therefore, the community, as a result of the intervention (and perhaps other factors), has begun to move toward lower utilization. This interpretation would be consistent with the view that the intervention had the effect of reducing utilization.

The analysis from the utilization data provides some support for the conclusion that the intervention reduced utilization rates. The areas where there was reduced utilization were the areas most likely to be affected by the intervention. The time series analysis showed reductions in utilization for TLAS, doctors’ office visits, and emergency room visits. However, because of the observed secular trends in the comparison communities, it is not certain that the observed effect can be attributed to the intervention. But taken together, the findings from
the survey and the utilization data could be interpreted as a modest utilization effect, particularly under the hypothesis of a 1-year lag effect.

CONCLUSIONS

In summary, the citizens of the intervention community clearly used self-care resources more frequently as a result of the community-wide distribution of the Healthwise Handbook and the community intervention activities. And, they reported, they found these resources very valuable. The findings indicate that consumers are more likely to use self-care resources when they are made more accessible and when they are encouraged to do so by health care providers and the media. Consumers report they find these resources useful and indicate a high degree of satisfaction when using them. Self-care information users believe that these products help them make better decisions regarding when to seek care and how to self-treat problems. Most believe that using the self-care resources saved them from seeking unnecessary care. And although there are alternative interpretations, the findings from the utilization data provide some evidence to support this conclusion.

These observed positive effects of the intervention should be interpreted within the context of the overall design. There are at least two factors inherent to the design of the intervention and the evaluation that increased the probability of an overall false negative finding or at least made it difficult to interpret apparently mixed findings. First, the methodologies for the evaluation of community intervention studies are still evolving.

It is recognized that intervening at the community level is potentially more powerful or at least more cost-effective than interventions at the individual level. But the approaches for evaluating the effects of community interventions are generally underpowered. A key limiting factor for community intervention studies has been underestimating the secular change occurring in comparison communities (and the larger society), leaving the trials with insufficient power to detect changes beyond secular trends (Murray, 1995).

Second, it is inherently difficult to measure differences in variables characterized by nonnormal distributions or large variances. Because medical care utilization distributions are the classic example of this
problem, it is extremely difficult to clearly identify the relatively small magnitude in changes that could be expected as a result of interventions such as the one under study in this report. Utilization data are not normally distributed and are characterized by large variances. Consequently, tests of significance that rely on the assumption of normality are inherently unstable in measuring these data. In addition, people do not use physician services evenly across time. The circumstance to use a self-care manual may arise only occasionally, maybe not at all within a period of a year. Therefore, the intervention could be producing intended outcomes but not be at a threshold level to detect within the evaluation time frame. Thus, the general concern with community intervention evaluations and with the assessment of utilization data is the risk of a false negative finding, that is, a finding of no significant difference in the intervention community when a difference actually exists.

However, even with these potential inherent biases toward a negative finding, the results showed that the intervention had a significant effect in the intended direction on many of the outcomes. A significant increase in the use of self-care manuals was observed, and most of the reported effects of using the manual were greater in the intervention community (e.g., helped to self-treat and saved a doctor visit). The observed secular trends in the comparison communities make it less certain that the observed reduction in utilization can be attributed to the intervention. However, taken together, the findings from the survey and the utilization data could reasonably be interpreted as a modest utilization effect, particularly under the hypothesis of a 1-year lag effect.

Based on observing the three communities over this 3-year period, it is clear that self-care products are widely used and their use is growing. As Web-based information sources become more plentiful and consumers more accustomed to using them, self-care information usage is apt to grow. However, as these information sources grow, consumers need some assurance about the legitimacy of the sponsor and the accuracy and the completeness of the information. Endorsement of particular information sources from medical providers and reinforcing messages from the media and medical personnel will likely enhance usage.

Providing consumers with access to the same type of information used by clinicians is changing the roles of both patients and providers.
As models for delivering care to a defined population continue to evolve, the importance of the provision of information to support both clinician and consumer decisions will also expand. Supporting consumer decisions with timely and accurate information has the potential to enhance the quality of care, the appropriateness of care, and health outcomes. Unfortunately, these outcomes are typically not examined in most self-care information interventions, including this evaluation. Nevertheless, they should be the focus of future work. The findings do, however, suggest that using a planned community-level intervention to provide increased access and reinforcements for the use of self-care information does appear to be a viable approach.

NOTE

1. The question on the survey asked if their household had any self-care books.

REFERENCES


A questionnaire was distributed at the American Association of Chronic Fatigue Syndrome’s biannual convention in Washington in January 2001 as well as through various Internet Web sites and listservs during early February and March of 2001. The sample consisted of 432 respondents. Most respondents (86%) indicated they wanted a name change, although more patients than scientists were in favor of this change. It was also apparent that the patients and physicians were clearly split between adopting a name such as myalgic encephalopathy versus one such as neuro-endocrine immune disorder. Also, among those respondents who selected either of these two choices for a new name, less than 30% of them supported the other name. Although the majority of respondents feel the name should be changed at this time, this survey suggests there are different stakeholders involved in the name-change process, each with strong and sometimes disparate feelings about changing the name.

AUTHORS’ NOTE: Financial support for this study was provided by National Institute of Allergy and Infectious Diseases Grant No. AI36295. We appreciate the work contributed by the members of Chronic Fatigue Syndrome Coordinating Committee Name Change Workgroup (i.e., Carol Lavrich, K. Kimberly Kenney, Charles Lapp, John Herd, Daniel Kahn, Susan Levine, Nancy G. Klimas, and Leonard A. Jason), who created the survey used in this study. Requests for reprints should be sent to Leonard A. Jason, Ph.D., DePaul University, Center for Community Research, 990 W. Fullerton Ave., Chicago, IL 60614.
Many patients with chronic fatigue syndrome (CFS) have felt stigmatized or misunderstood by medical professionals. For example, Anderson and Ferrans (1997) found that 77% of individuals with CFS reported past negative experiences with health care providers. Another survey found that 57% of respondents were treated badly or very badly by their doctors (David, Wessely, & Pelosi, 1991). Green, Romei, and Natelson (1999) also found that 95% of individuals seeking medical treatment for CFS reported feelings of estrangement, and 70% believed that others attributed their CFS symptoms to psychological causes.

It is possible that negative attitudes toward people with CFS might help explain the consistent finding that patients with CFS have mixed experiences with the health care system. Shlaes, Jason, and Ferrari (1999) developed a CFS Attitudes Test and found that if someone believes that people with CFS are responsible for their illness, it is likely that they will also believe that people with CFS have negative personality characteristics, such as being compulsive or overly driven. It is possible that negative attitudes might be a function of past negative portrayals of CFS as either nonexistent or as a function of neurotic, overworked, stressed lifestyles (Jason et al., 1997).

The patient community has felt that the term chronic fatigue syndrome trivializes the seriousness of this illness, as the illness is typified by many severe symptoms in addition to fatigue, and fatigue is generally regarded as a common symptom experienced by many otherwise healthy individuals in the general population (Taylor, Friedberg, & Jason, 2001). In addition, CFS is frequently confused with chronic fatigue, which is a symptom of many illnesses, including some psychiatric disorders. The negative stigma that is associated with CFS could also be due to the trivializing name that has been given to this disorder. The name selected to characterize an illness, such as chronic fatigue syndrome, can influence how patients are perceived and ultimately treated by medical personnel, family members, and work associates. Although it was expected that the CFS name would be eventually replaced as more information became available, this name has remained and has become the most commonly used label within the United States (Friedberg & Jason, 1998).

Many health care professionals and research scientists feel that if the name were to be changed, it would be best to have a scientific basis for the change. Unfortunately, few data have been collected to help
guide the process of revising the name. Two recent studies reviewed whether different names for CFS indeed prompt different attributions regarding its cause, nature, severity, contagion, and prognosis among samples of medical trainees and university undergraduates (Jason, Taylor, Plioplys, Stepanek, & Shlaes, in press; Jason, Taylor, Stepanek, & Plioplys, 2001). In these studies, three different diagnostic labels (chronic fatigue syndrome, Florence Nightingale disease, and myalgic encephalopathy) were tested to determine their effects on the attributions of medical trainees (Jason et al., in press) and college undergraduates (Jason et al., 2001) regarding this syndrome. Participants were randomly assigned to one of three groups, with the only difference between groups being in the type of diagnostic label given as the diagnosis for the same case study of a patient with prototypic symptoms of chronic fatigue syndrome.

Results of these studies suggested that participants’ attributions about CFS change based on the different diagnostic labels used to characterize it. The myalgic encephalopathy label was associated with the poorest prognosis, and this term was more likely to attribute a physiological cause to the illness and less likely to consider the patient in the case study as a potential candidate for organ donation.

Results of a third investigation (Taylor, Jason, Kennedy, & Friedman, 2001) indicated that physician recommendation for treatment of CFS can influence subsequent attributions about the illness among mental health practitioners. When a medically based treatment (ampligen) was recommended, the practitioners were significantly more likely to believe that the patient was significantly more disabled than those prompted with recommendation for a psychologically based treatment (cognitive-behavior therapy with graded activity). These findings highlight distinctions in attributions that can result from recommending a medically based orientation toward CFS treatment versus one based heavily on minimizing medical factors, emphasizing psychiatric factors.

The studies reviewed earlier highlight the effect of diagnostic labeling and the salience of similar types of brief communications used to describe illnesses such as CFS. They suggest that the name given to a syndrome can influence public attributions about those with the syndrome. During the summer of 1997, the Chronic Fatigue and Immune Dysfunction Syndrome (CFIDS) Association of America conducted a survey of 120 members to determine their opinions about changing
the name. Eighty-five percent of respondents indicated they wanted
the name changed (Name-Change Survey Results, 1997). Another sur-
vey of 182 respondents by the editor of the Chronic Fatigue Syndrome
Electronic Newsletter indicated that 92% wanted the CFS name
changed (Burns, 1998).

The Department of Health and Human Services Chronic Fatigue
Syndrome Coordinating Committee appointed a Name Change
Workgroup during the summer of 2000 to consider changing the name
of chronic fatigue syndrome to a name that more accurately reflects
both the severity of the disease and the organ systems affected by the
disease. The workgroup began to have regular meetings primarily by
telephone conference calls. The workgroup agreed that there was a
need to develop a new name and that the name should not be limited to
a single symptom, as had been done with the name chronic fatigue
syndrome.

At the State of the Science CFS meeting in Washington during
October of 2000, the workgroup had a chance to meet in person to dis-
cuss potential names that might be adopted: The name NEID (neuro-
endocrine immune disorder) was developed as one possibility as there
was evidence of dysfunction in the neurological, neuro-endocrine,
and immunological functioning in patients with this syndrome. In
addition, there was also interest in considering the name myalgic
encephalopathy (ME) or a version of this name. The term ME had
been used prior to the use of the term CFS. In an effort to collect infor-
mation from a wide variety of individuals on their existing feelings
regarding the need for a name change and these name change candi-
dates, a name change survey questionnaire was developed in January
of 2001 to be broadly disseminated to the various stakeholder groups.
This study presents the results of this survey.

METHOD

A questionnaire was distributed at the American Association of
Chronic Fatigue Syndrome’s (AACFS) biannual convention in Wash-
ington in January 2001 (370 individuals registered for this confer-
ence) as well as through various Internet Web sites and listservs during
early February. The Name Change Workgroup wanted to assess the
term neuro-immune endocrine disorder because there is scientific
evidence that patients with CFS have abnormalities in their neurological, immunological, and endocrine systems (Friedberg & Jason, 2001). This term does convey a more serious and possibly debilitating condition than CFS. The term myalgic encephalopathy was also evaluated as it had been used by patients before a group of Centers for Disease Control scientists developed the CFS name in 1988. Many patients both within the United States and in other countries continue to use this term or a version of it (i.e., myalgic encephalomyelitis). Two other names were also considered, polyalgic asthenia, a name that attempts to describe key symptoms, and Ramsay’s disorder, an eponym for a famous physician who had studied this syndrome.

At the AACFS conference, 108 questionnaires were completed by conference participants, and an additional 324 questionnaires were filled out after the conference and mailed in either by Internet e-mail or hard copy. The two samples were then combined, and the final sample consisted of 432 respondents.1

Respondents were classified into one of the following four groups: patients, family members, clinical care professionals, or research professionals. If an individual checked more than one category and one of the categories checked was research professional (e.g., research professional and clinical care professional), the person was classified as a research professional. If the person was not a research professional but one of the categories checked was clinical care professional (e.g., clinical care professional and family member), the person was classified as a clinical care professional.

RESULTS

Of those completing questionnaires, 69% were patients, 10% were family members of patients, 9% were professionals involved in the clinical care of patients, and 12% were professionals actively conducting research related to this disease.

Table 1 indicates that most respondents (86%) indicated that they wanted a name change at this time. Using a chi-square, there was a significant difference between the four groups of respondents, \( \chi^2(3, n = 414) = 54.19, p < .01 \). When each of the groups was contrasted with each other group using chi-square analyses, the research professionals responded affirmatively significantly less than patients,
χ²(1, n = 335) = 53.61, p < .01, and family members, χ²(1, n = 94) = 11.49, p < .01. In addition, patients responded affirmatively significantly more often than clinical care professionals, χ²(1, n = 320) = 11.54, p < .01.

Table 2 indicates the support for different names that have been proposed. There does not appear to be a consensus supporting one specific name, although two names were at the forefront, NEID and ME. When respondents were asked whether they would support the use of a name such as ME or NEID, patients were the most positive in supporting these names (ME = 57.7%, NEID = 61.5%), whereas family members’ (50%, 44%) and clinical care professionals’ ratings (46.7%, 59.3%) were a little lower, and the lowest ratings were from professionals in CFS research (34.1%, 36.1%). Table 2 also indicates that support for other names, such as polyalgic asthenia and Ramsay’s disorder, is considerably less than names such as ME and NEID.

Table 3 reports on the top choice for a name for each of the respondent groups. It is apparent that the patients and physicians are clearly split between adopting a name such as ME and NEID. When asked if they had to choose one name, 38.1% of patients selected ME and 45.9% selected NEID; 36.6% of family members selected ME and 48.8% selected NEID; 44.1% of clinical care professionals selected ME and 41.2% selected NEID; and 47.6% of research professionals selected ME and 31.0% selected NEID. For these figures, the ME term referred to myalgic encephalopathy or a name like it; the survey also contained a category to list other names, and if respondent used the term myalgic encephalomyelitis or a term like this in the other category, it was included in the count for ME.
### TABLE 2
Preferences for Different Names for Chronic Fatigue Syndrome (CFS) (Would You Support the Following Names?)

<table>
<thead>
<tr>
<th>Names</th>
<th>Yes n</th>
<th>Yes %</th>
<th>No n</th>
<th>No %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myalgic encephalopathy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professionals in CFS research⁹</td>
<td>48</td>
<td>17</td>
<td>31</td>
<td>64.6</td>
</tr>
<tr>
<td>Professionals in CFS clinical care⁹</td>
<td>32</td>
<td>11</td>
<td>21</td>
<td>65.6</td>
</tr>
<tr>
<td>Patients who have CFS⁹</td>
<td>252</td>
<td>137</td>
<td>116</td>
<td>45.6</td>
</tr>
<tr>
<td>Family members of PWCs and others⁹</td>
<td>32</td>
<td>8</td>
<td>24</td>
<td>75.0</td>
</tr>
<tr>
<td>Name like myalgic encephalopathy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professionals in CFS research⁹</td>
<td>44</td>
<td>15</td>
<td>29</td>
<td>65.9</td>
</tr>
<tr>
<td>Professionals in CFS clinical care⁹</td>
<td>30</td>
<td>14</td>
<td>16</td>
<td>53.3</td>
</tr>
<tr>
<td>Patients who have CFS⁹</td>
<td>208</td>
<td>120</td>
<td>88</td>
<td>42.3</td>
</tr>
<tr>
<td>Family members of PWCs and others⁹</td>
<td>26</td>
<td>13</td>
<td>13</td>
<td>50.0</td>
</tr>
<tr>
<td>Neuro-endocrine immune disorder</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professionals in CFS research</td>
<td>50</td>
<td>18</td>
<td>32</td>
<td>64.0</td>
</tr>
<tr>
<td>Professionals in CFS clinical care</td>
<td>31</td>
<td>14</td>
<td>17</td>
<td>54.8</td>
</tr>
<tr>
<td>Patients who have CFS</td>
<td>264</td>
<td>148</td>
<td>116</td>
<td>43.9</td>
</tr>
<tr>
<td>Family members of PWCs and others</td>
<td>37</td>
<td>18</td>
<td>19</td>
<td>51.4</td>
</tr>
<tr>
<td>Name like neuro-endocrine immune disorder</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professionals in CFS research⁹</td>
<td>36</td>
<td>13</td>
<td>23</td>
<td>63.9</td>
</tr>
<tr>
<td>Professionals in CFS clinical care⁹</td>
<td>27</td>
<td>16</td>
<td>11</td>
<td>40.7</td>
</tr>
<tr>
<td>Patients who have CFS⁹</td>
<td>221</td>
<td>136</td>
<td>85</td>
<td>38.5</td>
</tr>
<tr>
<td>Family members of PWCs and others⁹</td>
<td>25</td>
<td>11</td>
<td>14</td>
<td>56.0</td>
</tr>
<tr>
<td>Polyalgic asthenia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professionals in CFS research</td>
<td>46</td>
<td>3</td>
<td>43</td>
<td>93.5</td>
</tr>
<tr>
<td>Professionals in CFS clinical care</td>
<td>28</td>
<td>3</td>
<td>25</td>
<td>89.3</td>
</tr>
<tr>
<td>Patients who have CFS</td>
<td>253</td>
<td>35</td>
<td>218</td>
<td>86.2</td>
</tr>
<tr>
<td>Family members of PWCs and others</td>
<td>31</td>
<td>4</td>
<td>27</td>
<td>87.1</td>
</tr>
<tr>
<td>Name like polyalgic asthenia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professionals in CFS research</td>
<td>37</td>
<td>4</td>
<td>33</td>
<td>89.2</td>
</tr>
<tr>
<td>Professionals in CFS clinical care</td>
<td>26</td>
<td>3</td>
<td>23</td>
<td>88.5</td>
</tr>
<tr>
<td>Patients who have CFS</td>
<td>208</td>
<td>31</td>
<td>177</td>
<td>85.1</td>
</tr>
<tr>
<td>Family members of PWCs and others</td>
<td>21</td>
<td>3</td>
<td>18</td>
<td>85.7</td>
</tr>
<tr>
<td>Ramsay’s disorder</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professionals in CFS research</td>
<td>47</td>
<td>6</td>
<td>41</td>
<td>87.2</td>
</tr>
<tr>
<td>Professionals in CFS clinical care⁹</td>
<td>28</td>
<td>1</td>
<td>27</td>
<td>96.4</td>
</tr>
<tr>
<td>Patients who have CFS⁹</td>
<td>248</td>
<td>50</td>
<td>198</td>
<td>79.8</td>
</tr>
<tr>
<td>Family members of PWCs and others⁹</td>
<td>34</td>
<td>2</td>
<td>32</td>
<td>94.1</td>
</tr>
<tr>
<td>Name like Ramsay’s disorder</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professionals in CFS research</td>
<td>13</td>
<td>3</td>
<td>10</td>
<td>76.9</td>
</tr>
<tr>
<td>Professionals in CFS clinical care</td>
<td>10</td>
<td>1</td>
<td>9</td>
<td>90.0</td>
</tr>
<tr>
<td>Patients who have CFS</td>
<td>177</td>
<td>39</td>
<td>138</td>
<td>78.0</td>
</tr>
<tr>
<td>Family members of PWCs and others</td>
<td>17</td>
<td>2</td>
<td>15</td>
<td>88.2</td>
</tr>
</tbody>
</table>

NOTE: Similar superscript letters by two respondent groups indicate that the groups are significantly different. PWCs = persons with CFS.
To derive a measure of how strongly respondents felt about their selections for ME or NEID, the questionnaire data were tabulated to discover whether respondents who selected as their first choice either ME or NEID would be willing to support the other term. For those respondents who selected NEID as their first choice, only 28.3% of those respondents would support ME as a name to replace CFS. For those respondents who selected ME as their first choice, only 22.1% of those respondents would support NEID as a name to replace CFS.

**DISCUSSION**

Findings from the survey suggest that there are different stakeholders with strong feelings about changing the name. Most respondents that seem to like NEID do not support ME, and vice versa. In addition, data from the survey suggest that about 50% of the researchers feel that the term CFS should not be changed. It is likely that many of these scientists feel that over the past decade, the term CFS has gained acceptance throughout the world and that changing the term
now would lead to confusion. Many of these scientists feel that making a change in the term right now would be a mistake unless there is a scientific breakthrough that would warrant a new name.

The findings from this study suggest that different stakeholders are committed to different names. If one name is proposed, it is likely that one or more of the other groups whose names are not selected will feel alienated and might not support the proposed name. For example, those who support ME feel that it was a mistake to use the term CFS, and they now want the original name reinstated. However, others feel that neuro-immune endocrine disorder is an even better name, and they feel that to start fresh with a new term would generate a better chance of gaining the support that is needed to build consensus for a new name. It is possible that as long as the focus is on selecting one name, it might be difficult to gain the support that is needed to make any changes.

When scientists use the term CFS, they are referring to the Fukuda et al. (1994) research case definition, and this case definition is used to designate a very specialized group of patients for research studies. Although the term CFS and its criteria were developed to be a research case definition, it has been used, for all practical purposes, to define all individuals with this condition (Fukuda et al., 1994). Many patients with CFS feel that not only is the term CFS offensive but that the criteria for this case definition are not appropriate (there are many exclusionary criteria with this research case definition, and many patients who feel that they have CFS do not meet the research case definition due to those rules) (Friedberg & Jason, 1998). One possible compromise solution would be to endorse the current name of CFS as a research case definition and recommend that the term CFS be used for this purpose. Scientists would then continue to use the term CFS and its criteria to identify a homogeneous group of participants to make comparisons across different settings. Such a strategy could appeal to the scientific audience.

If the scientists were to accept this recommendation, it would also be important to find a way to satisfy the other two constituent groups that feel strongly about the terms ME and NEID. A term such as NEID could be used to refer to a hypothetical construct that would be broader than the research case definition. Just as with the terms cancer or heart disease, a term such as NEID could refer to a larger category
comprising individuals with this syndrome, and it would still be possible for there to be particular subtypes of this syndrome, such as the research case definition, which is called CFS.

A consensus panel in Canada has recently proposed a clinical case definition that is called ME/CFS and specified different criteria than the research case definition of CFS (N. Klimas, personal communication, April 20, 2001). The current Fukuda et al. (1994) standard refers to a research case definition for CFS, and this was developed to identify a homogeneous group of patients for research purposes. The Canadians have used the term ME/CFS to refer to a clinical case definition (which will be broader than the research case definition) but have indicated that they are waiting for the U.S. Name Change Workgroup to make a decision about the actual name. If ME referred to the clinical case definition, CFS to the research case definition, and a term such as NEID to an overall construct within which both ME and CFS could be contained, each of the constituent groups would be able to have their term applied to this syndrome.

There are several limitations in this study, including the fact that the sample was collected at a research conference and through the Internet, and thus, this sample might not be representative of the patient community or medical personnel who treat this illness. In addition, it is possible that the sample was also biased in that people might have been more likely to respond to the questionnaire if they favored a name change.

Many diseases and syndromes have more than one name, and it is possible that by having several names, each with slightly different criteria, the different stakeholders might be willing to accept such a compromise. The current survey does suggest that there are at least three different constituent groups, those scientists who want to preserve the CFS label and those respondents who want either ME or a term such as NEID. Clearly, given how strongly each group feels about the name, finding a way to reach a compromise might be the only way to gain the approval for a name change. This effort is a bit unwieldy, and attempting to please multiple constituencies is a difficult task. However, regardless of the outcome, the findings from this survey have played a useful role in providing the members of the Name Change Workgroup as well as the broader CFS community data on perceptions of stakeholders, and these data are being used to guide the workgroup in its
recommendations. We are not aware of other examples of changing the name of a disease in the manner that is being attempted with CFS.

NOTE

1. There were differences between the two samples. The percentages of the types of respondents differed greatly between them. The respondents at the conference were much more evenly distributed (professionals in research 32%, professionals in clinical care 20%, patients 32%, and family members 16%), whereas the results of the type of postconference respondents were much more skewed (6%, 5%, 81%, and 9%, respectively). However, in comparing the results of the two samples, they were very similar. In addition, as both samples were using the exact same questionnaire and were pulling from similar groups of people (people interested in and at least somewhat knowledgeable of chronic fatigue syndrome), we felt it was appropriate to combine them.

REFERENCES


Peer and self-evaluation are crucial in the professional development of physicians. However, these skills must be learned, and there are barriers to their acceptance and successful utilization. To overcome these obstacles, it has been suggested that these concepts should be addressed longitudinally throughout medical education. Therefore, first-year medical students were introduced to peer and self-assessment as part of a videotape review during an interviewing course by having students complete written peer and self-assessments of the interviews. Students’ self-assessments were compared with the assessments of peers and faculty. Written evaluations showed peers were more lenient than faculty and students were most critical of their own performances. Students could provide balanced assessments of their peers but were predominately negative regarding their own performances. It appears first-year students are capable of evaluating their peers but have difficulty accurately assessing their own performance. Further interventions are needed to foster self-assessment skills in first-year students.
An important goal of medical education is to provide the physician with the knowledge and skills required for life-long professional development. A critical skill in this regard is the ability to monitor and assess one’s personal performance (Gordon, 1997; Kennell, Tempio, & Marcia, 1973). After all, in our current medical education process, once formal training is completed, further professional development is predicated on the clinician’s ability to recognize his or her areas of weakness and take appropriate measures to remedy deficiencies. Thus, the ability to accurately self-assess one’s strengths and weaknesses may be the essential skill to ensure that life-long learning actually occurs (Wooliscroft, Tenhacken, Smith, & Calhoun, 1993). Nevertheless, little attention has been given to self-assessment in the medical curriculum (Gordon, 1991).

In addition, peer review may be an important process to ensure continued professional development by providing essential feedback on areas of strength and weakness not always gleaned from self-assessment or even assessment by instructors. For example, studies of peer review in residency programs have noted that residents are more discriminating than faculty in judgments of their peers’ team relationships, physician-patient relationships, teaching ability, and physical examination, suggesting that peers may have unique insight into these areas that are not always consistently directly observed by faculty (Thomas, Gebo, & Hellmann, 1999; Van Rosendaal & Jennett, 1994). Nevertheless, although research has shown that medical students can make reliable and valid assessment of their peers (Arnold, Willoughby, Calkins, Gammon, & Eberhart, 1981), peer assessment is not widely used in medical student evaluation (Association of American Medical Colleges, 1999), and there are few descriptions of such programs in the literature (Antonelli, 1997; Gruppen et al., 1997; Stuart, Goldstein, & Snope, 1980).

Few studies have compared self-assessment and peer assessment to external measures of evaluation, such as faculty evaluation. Morton and Macbeth (1977), in a report from a fourth-year surgical clerkship in New Zealand, noted that faculty and peer assessments were highly correlated but that 4% of students failed themselves on self-assessment although passed by faculty and peers. These students were subsequently found to have “serious psychological problems.” In a study of second-year medical students at the University of Michigan, students’ assessment of their physical examination abilities (as viewed on a
videotape) were similar to their peers’ assessment but were not as highly correlated with faculty assessment (Calhoun, Wooliscroft, Ten Haken, Wolf, & Davis, 1988). In a study of interns at University of California San Francisco, intern performance was viewed somewhat differently by supervisors, intern peers, and interns themselves, with faculty ratings best predicted by earlier medical school faculty ratings, self-ratings predicted best by a combination of personality inventory measures and medical school grades, and peer ratings best predicted by a unique combination of personality inventory factors such as tolerance (Kegel-Flom, 1975). Thus, it appears in these few studies that faculty, peer, and self-evaluations correlate to some degree but that each method provides some unique insights lacking in the other methods.

We believed peer and self-assessment would be especially useful in a first-year interviewing course because interpersonal skills seems to be an area where peers might provide unique feedback (Van Rosendaal & Jennett, 1994), because early indoctrination in the method of peer and self-assessment would facilitate early acceptance of these modes of evaluation and would promote early on the habit of critical self-inquiry and evaluation necessary for continued life-long learning, and because early formative feedback on this critical skill would allow students more time to modify their behavior (Helfer, 1972). Nevertheless, no study has reported comparisons of peer, self-, and faculty evaluation of interviewing skills for medical students. The purpose of this study was to compare faculty, peer, and self-assessment of interviewing skills during a first-year communication and interviewing course. We hypothesized that although evaluations derived from each method would highly correlate, each method would provide unique insights into the student’s interviewing ability.

METHOD

SUBJECTS

Ninety-seven students in the first-year medical class during the 1998-1999 academic year participated in an interviewing course in which self- and peer assessment was an evaluation component.
PROCEDURE

In the class orientation, students are advised on how to give and receive feedback and practiced giving feedback after viewing a videotaped doctor/patient interview. The course is conducted in small groups of approximately eight students who meet weekly with a faculty preceptor. Group discussion focuses on the structure and content of the medical interview and the development of specific communication skills. A major component of the course is individual practice with standardized patients. These brief interviews are conducted in the presence of the faculty preceptor and the other students in the small group. Students are encouraged to assess their own performance and to ask the group for advice. Each interview is followed by oral feedback from the preceptor, peers, and the standardized patient. Evaluation of the course consists of a summative evaluation by the preceptor, an evaluation of a videotaped interview, and a practical examination with a standardized patient who grades the student using an item-specific checklist.

The focus of this article is on the faculty, peer, and self-evaluation of the videotaped standardized patient interview. This interview occurred midway through the course. The small groups with their faculty preceptor viewed the videos. Following each video, students were asked to complete an evaluation form containing questions about interviewing performance. The videotaped interview was chosen for this study because all evaluators, including the student, could evaluate the same encounter at the same time. The video as the focus of the interview may help reduce the tendency for students to make a global assessment of their peers as well as themselves (Arnold et al., 1981).

MEASURES

Interviewing performance was assessed using three 15-point Likert-type scale questions assessing interviewing style (empathy, self-presentation, interest, respect, and rapport), interview structure (ability to follow guidelines, information flow, and completion of topics), and interviewing techniques (open/closed questions, summarization, legitimization, transition statements, and appropriate opening and closing). The 15-point numbered scales were labeled from 1 to 15 with unacceptable as 1, below expectations as 4, meets expectations as 8,
above expectations as 12, and excellent as 15. After viewing each videotape, students and the faculty preceptor rated interviewing performance on the aforementioned questions and were asked to write comments regarding the strengths of the interview and to write comments about areas where improvement was needed.

The following three evaluation variables were derived from the interviewing assessment form: interviewing performance composite, number of positive comments, and number of negative comments. The interviewing performance composite was calculated by forming the simple mean of the three 15-point Likert-type ratings of interviewing style, structure, and techniques. Coefficient alpha for this three-item linear composite was .85. Each student’s self-rating score was based on their rating of their own performance. Each student’s peer-evaluation score was the mean of the ratings of their performance by the other students in their group (coefficient alpha = .88). The faculty assessment was the mean of the three items rated by each student’s preceptor (coefficient alpha = .84).

Written comments to the open-ended question about strengths and areas in need of improvements were first enumerated (broken into conceptually different segments) and then coded as positive or negative by two graduate student coders. Percentage of agreement between coders was 93%. Differences about segmenting responses into categories or in coding into positive or negative domains were successfully reconciled by a consensus discussion with a third coder. Self- and faculty scores on numbers of positive and negative comments were generated for each student, and a peer score was calculated as the mean number of positive and negative comments of students from their small group.

ANALYSIS

Differences between self-, peer, and faculty evaluations were examined using multiple regression approaches derived from the general linear model (Cohen & Cohen, 1983) in a repeated measures framework. Dependent variables were the interviewing performance composite, number of positive comments, and number of negative comments. A statistically significant overall $F$ ratio was followed by single degree of freedom planned contrasts between self and peer, self and faculty, and faculty and peer ratings. Because of missing values,
the effective sample size was 82 for analyses with the interviewing composite score and 91 for the positive and negative comments variables.

RESULTS

Of the 97 students who participated in the course, missing data on self-evaluation ratings and/or comments reduced the effective sample size for self composite comparisons to 82 and to 91 for positive and negative comment comparisons. Peer and faculty evaluations of students with missing data on self-evaluations did not differ from students for whom self-ratings were present (\(p > .05\)).

Correlations among peer, self, and faculty composite score ratings were moderate. Pearson correlations between self and peer ratings (\(r = .29, df = 80, p = .008\)) and between faculty and peer ratings (\(r = .50, df = 86, p = .0001\)) were statistically significant. The correlation between self and faculty composite scores only showed marginal statistical significance (\(r = .19, df = 80, p = .08\)).

For the variable indexing the number of positive comments, only the correlation between peer and self scores was statistically significant (\(r = .30, df = 89, p = .003\)). The correlation between self positive and faculty positive was \(r = -.12, df = 89, p = .24\) and between faculty and peer positive was \(r = .05, df = 95, p = .59\). For the variable indexing the number of negative comments, all three correlations were moderately positive: peer-self \(r = .25, df = 89, p = .02\), self-faculty \(r = .22, df = 89, p = .04\), and faculty-peer \(r = .31, df = 95, p = .002\).

Means and standard deviations for the composite rating, positive comments, and negative comments scores for peer, self, and faculty raters can be found in Table 1. Peer ratings were highest, followed by faculty and then self ratings, overall \(F(2, 162) = 15.86, p < .0001\); single degree of freedom contrasts: self-faculty \(F(1, 81) = 6.13, p = .015\); self-peer \(F(1, 81) = 27.66, p < .0001\); faculty-peer \(F(1, 81) = 15.92, p < .0001\).

Faculty supplied the largest number of positive comments, followed by peers and then comments by self: overall \(F(2, 180) = 104.95, p < .0001\); single \(df\)/contrasts: self-faculty \(F(1, 90) = 141.85, p < .0001\); self-peer \(F(1, 90) = 342.25, p < .0001\); faculty-peer \(F(1, 90) = 48.30, p < .0001\). Individuals described themselves with the largest number
of negative comments, followed by faculty and then by peers: overall $F(2, 180) = 26.12, p < .0001$; single df contrasts: self-faculty $F(1, 90) = 19.71, p < .0001$; self-peer $F(1, 90) = 50.19, p < .0001$; faculty-peer $F(1, 90) = 4.63, p = .034$. Also displayed in Table 1 is a population ratio of positive to negative comments, indicating the degree to which self-assessments have a negative bias in comparison to peer and faculty comments. An example of written responses for one student is displayed in Table 2 to give the reader an example of a typical student’s protocol.

**DISCUSSION**

It is valuable for physicians to have an understanding and proficiency in self- and peer-assessment skills in developing and maintaining clinical competence. Self- and peer assessment should be addressed throughout the medical curriculum. However, there are potential barriers to peer assessment among medical students and residents such as distrust of faculty intentions, fear of harming each other’s grades, or disrupting collegiality (Gordon, 1992; Van Rosendaal & Jennett, 1992). Our students did not appear reluctant to evaluate their peers in that 95% of the students were willing to provide written peer assessments. The instruction in giving and receiving feedback, the supportive environment of the small group sessions, the anonymous nature of written responses, and use of peer assessment for feedback as opposed to evaluative purposes may have contributed to the high response rate (Farh, Cannella, & Bedeian, 1991).

Students were willing to provide positive as well as corrective written feedback regarding their peers. Peer balance between positive and
negative feedback occurred at a ratio of four to one. Such a ratio of positive to negative comments may be useful in promoting acceptance of corrective feedback by the receiver.

Peer-assessment skills did not appear to transfer to self-assessment skills. In spite of students’ willingness to evaluate their peers in an appropriately balanced fashion, their self-assessments were overwhelmingly critical. Other investigations in students and residents have shown similar results (Arnold et al., 1981; Linn, Arostegui, & Zeppa, 1975; Morton & Macbeth, 1977; Stuart et al., 1980). There are several explanations for these findings. Socially, students may be uncomfortable saying anything positive about their performance; rather, they submit to humility (Stackhouse & Furnham, 1983). Due to lack of experience, students may have unrealistic goals and expectations for their level of training (Stackhouse & Furnham, 1983). There may be difficulty overcoming the traditional culture of evaluation in medical education that has been judgmental and punitive (Kurtz, Silverman, & Draper, 1998). Students may judge peers by their performances but judge themselves by their perceived capacities in a

---

**TABLE 2**

Written Comments for One Student (these comments are representative of the relative negative self-evaluation comments when compared to that of peers and faculty)

<table>
<thead>
<tr>
<th>Peer positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Very professional! Have you done this before?”</td>
</tr>
<tr>
<td>“Very organized, attentive. Established comfort.”</td>
</tr>
<tr>
<td>“Good avoiding of jargon.”</td>
</tr>
<tr>
<td>“Great nonverbal communication.”</td>
</tr>
<tr>
<td>“Flow of questions seemed natural.”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Peer negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Avoid leading questions (e.g., ‘So you have a pretty strong relationship with your husband?’)”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Self-evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Don’t say OK so much.”</td>
</tr>
<tr>
<td>“Don’t interrupt.”</td>
</tr>
<tr>
<td>“Be more confident.”</td>
</tr>
<tr>
<td>“Say things once.”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Faculty comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Good summarizing.”</td>
</tr>
<tr>
<td>“Asked patient attribution.”</td>
</tr>
<tr>
<td>“Asked good open-ended question to get into social history.”</td>
</tr>
<tr>
<td>“Very good job.”</td>
</tr>
</tbody>
</table>
global fashion (Gordon, 1991). This global self-assessment may be especially problematic to communication skills that are closely tied to self-concept (Kurtz et al., 1998). The use of videotape reviews may have also contributed to a negative self-assessment by causing anxiety and heightening self-consciousness (Hargie & Morrrow, 1986). Further interventions to foster student self-assessment skills may be indicated.

There are several limitations to our investigation. It was conducted on a single class at a single institution. The evaluation form used for the numeric scores was developed at our university. Perhaps the use of a validated communication and interviewing checklist of specific behaviors such as the SEGUE Framework for Teaching and Assessing Communication Skills (Makoul, 1998) or the Calgary-Cambridge Observation Guide (Kurtz et al., 1998) would allow for more generalizability.

We conclude that first-year students are willing to provide peers with specific written feedback with an appropriate ratio of positive to negative comments. This skill did not transfer to self-evaluation, which remained predominately negative. We plan to continue the exercise using open-ended comments only and to survey students asking their perceptions of the utility, accuracy, and comfort level with self- and peer assessment of interviewing skills.

Questions for future investigation may include the following: How well does peer evaluation of communication and interviewing predict student performance in interpersonal skills during their clinical rotations? How does early exposure to self- and peer assessment influence students’ attitudes to these forms of feedback?

REFERENCES


This article describes the effect of an online analytic skills training course on professional development and practice and discusses recommendations for using this training modality in the public health workforce. The Enhancing Data Utilization Skills Through Information Technology initiative trained professionals in maternal and child health from 13 Southern-tier state and local health departments to collect, analyze, and interpret data via a year-long Web-based course. The evaluation of this initiative was based on a model of change for health professionals that holds that training influences behavior by increasing knowledge, influencing beliefs related to the behavior, enhancing self-efficacy, and improving skills. Participants' knowledge, beliefs, and self-efficacy all increased significantly during the course. Participants' self-assessed skill levels increased significantly for each of 12 selected skills and overall for all skills combined. Distance learning is potentially an effective means for professionals to advance their skills while continuing to fulfill their work-related responsibilities.
In a climate of increasing complexity and accelerating performance and accountability standards (Task Force on Public Health Workforce Development, 1999), public health professionals need training to help them deploy their resources effectively and efficiently. In particular, the public health workforce needs considerable expertise in using technology to analyze and evaluate data (Hirano & Dillenberg, 1998). The technical and data skills training needs of public health staff members, however, are not being met (Lasker, 1998). Distance learning is potentially an effective means for professionals to advance their skills while continuing to fulfill their work-related responsibilities. We describe the effect of an online analytic skills training course on professional development and practice and discuss recommendations for using this training modality in the public health workforce.

THE ENHANCING DATA UTILIZATION SKILLS THROUGH INFORMATION TECHNOLOGY COURSE

The Enhancing Data Utilization Skills Through Information Technology (EDUSIT) initiative trained professionals in maternal and child health from 13 Southern-tier (Alabama, Arizona, Arkansas, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, New Mexico, South Carolina, North Carolina, and Virginia) state and local health departments to collect, analyze, and interpret data via a yearlong Web-based course. A regional approach was taken because of strong existing relationships between the University of North Carolina School of Public Health and state health departments in this geographic area of the United States. The course was offered through the Department of Maternal and Child Health (MCH) in the School of Public Health at the University of North Carolina at Chapel Hill in collaboration with the school’s Center for Distance Learning. The overall project goal was to strengthen the analytic skills of the MCH workforce. The course was composed of six major topic areas, each of which was taught through a separate module lasting 1 month. The module topics, selected as a result of our experience conducting
analytic skills training and Title V directors’ assessments of their needs, were the following:

- basic epidemiological and statistical concepts
- qualitative data collection and analysis
- measurement of social inequalities in health
- development of data collection instruments and planning data analysis
- economic analysis
- use of Geographic Information Systems software.

Faculty members with expertise in the specific content area developed the modules with the assistance of an instructional designer/Web developer. Each module included specific content, references, examples, and opportunities to practice knowledge and skills and to receive feedback from faculty. In addition, participants worked in state-specific teams to select and study a health status problem and apply the skills learned in the course. The final report of the team was in the form of a Web page that was linked to the state health page. Six hours of graduate credit were offered. For a more detailed description of the course, see Polhamus, Farel, and Stephens (2000).

The emphasis on teamwork was stimulated by evidence of the role of social networks in reinforcing use of technology (Fulk, Schmitz, & Steinfield, 1990). To ensure a focus on the analytic skills being taught in the course, all participants were required to pass a short, online, precourse module that covered the essential technical skills as a prerequisite for enrollment. For more information about the precourse technology skills module, see Polhamus, Farel, and Trester (2000).

This article reports results from the evaluation of the course, emphasizing the course’s impact on the participants and their practice of public health.

**METHOD**

**SAMPLE**

Conference calls were conducted with each state to discuss the project and answer questions. State MCH directors disseminated
information about the conference call. Forty state and local health department staff members enrolled in the course, which included individuals selected by the state MCH director and others who volunteered. Participants were 40 years of age or older (82%), well educated (67% had master’s degrees or higher), and their last degree was awarded more than 10 years prior to the project. Most participants (90%) worked at the state rather than the local level. As a cohort, this group had positive attitudes about using data and technology, although confidence in their abilities to use data and demonstrate technology skills was low.

Twelve participants dropped out of the project because they left the agency, moved to a different position within the health department and the course did not fit into their new job, or had responsibilities added to their position that did not leave time to complete assignments.

THEORETICAL FRAMEWORK

The evaluation was based on a model of change for health professionals enrolled in training programs derived from Umble, Cervero, Yang, and Atkinson (2000) (see Figure 1). The model holds that training influences behavior by increasing knowledge, influencing beliefs related to the behavior (i.e., learner opinions of the value of particular skills on the job), increasing confidence in one’s ability to perform specific behaviors (self-efficacy), and improving skills. Application of skills is shaped further by work-setting factors such as peer and supervisor support, time, and the appropriate supplies and equipment. A wide range of research in professional continuing education and training supports the model (Green, Eriksen, & Schor, 1988; Kraiger, Ford, & Salas, 1993; Ottoson, 2000; Umble & Cervero, 1996).

The project sought to influence participants’ knowledge, beliefs, self-efficacy, and skills through training in addition to organizational support for learning and applying the skills. Supervisors were kept apprised of the project through meetings and periodic reports from project staff members.

The evaluation measured the course’s effects on participants’ knowledge, beliefs, self-efficacy, skill level, and practice of the skills taught.
DESIGN AND MEASUREMENTS

All data were collected online. A one-group pretest-posttest design (McMillan & Schumacher, 1989) examined changes in knowledge, beliefs, and self-efficacy. The pretest and posttest were conducted online immediately before and after completion of each module.

Knowledge. A participant’s knowledge (Kraiger et al., 1993) of the content in each module was measured by totaling the correct responses to a set of five to eight multiple-choice questions. For example, “Which of the following is a method of analyzing qualitative data? (a) t tests, (b) code and retrieve, (c) normal distribution, (d) focus groups” was one of the questions for the qualitative methods module.

Beliefs. Changes in beliefs (Fishbein & Ajzen, 1975) were measured by asking participants to rate the extent to which they agreed with a set of two to four statements about the value of a specific skill using a 4-point Likert-type scale. For example, “Field observations are a useful tool for monitoring the implementation of a program” was an item for the qualitative methods module.

Self-efficacy. Self-efficacy refers to project participants’ beliefs in their own competence to perform a particular behavior (Maibach & Murphy, 1995). Self-efficacy has been shown to influence goal setting and goal commitment, coping and persistence in learning difficult
tasks, and interpretation of feedback (Applebaum & Hare, 1996; Gist & Mitchell, 1992). Self-efficacy has also been shown to predict the behavior of health professionals following training (Umble et al., 2000). Changes in self-efficacy were measured with three to five questions for each module. For each item, respondents used a 4-point scale ranging from not at all confident to completely confident. For example, one item for the qualitative methods module asked,

Your state team has recently conducted 20 qualitative interviews with women and their children regarding a new service in your state. How confident are you in your ability to use the code and retrieve method of analysis to analyze these interviews if you had to do that right now?

Skill. A retrospective pretest-posttest design (Howard, 1980) used a 6-month follow-up survey to measure changes in self-assessed skill level due to the program. This design is similar to the traditional pretest-posttest design except that the pretest is done retrospectively at the end of the program rather than at the beginning. This procedure has been shown to eliminate the common response shift bias that occurs when participants have different internal scales by which they rate themselves before and after an intervention. This procedure violates the fundamental requirement of the pre-post design that the same measurement scale be used before and after the intervention. For example, a participant in a training program on interviewing skills may be fairly confident and rate himself or herself as being at Level 4 out of 5 on a traditional pretest. After the program, however, he or she may be more realistic about the difficulty of conducting interviews and have a different internal rating scale. Thus, although the participant’s skill has increased, he or she may still rate his or her skill at Level 4 or even reduce it to Level 3. In a retrospective pretest, the same trainee may rate his or her pretraining skill level at 2 compared to his or her skill now. When this response shift bias occurs due to a training program, the trainee’s pre-post measurements are using a different scale, and the treatment effect will be underestimated. Numerous studies have documented that this measurement problem can lead to negative, reduced, or nonsignificant treatment effects in comparison with the retrospective pretest-posttest design (Bray, Maxwell, & Howard, 1984; Howard, 1980; Howard & Dailey, 1979; Pratt, McGuigan, & Katzev, 2000; Skeff, Stratos, & Bergen, 1992).
Participants rated their level of proficiency (Kraiger et al., 1993) on 12 skills emphasized in the course on a 7-point scale (1 = low, 7 = high) after the course (the traditional posttest) but also reflected back and rated their skills on the same items before they started the course (the retrospective pretest). The skills were as follows:

1. selecting an appropriate secondary source
2. conducting a Web search
3. collecting qualitative data through interviews
4. analyzing and interpreting qualitative data
5. using the guidance of Office of Management and Budget Directive 15 (standards for providing consistent data on race and ethnicity)
6. creating a map in ArcExplorer
7. considering implications of data that are reported by race and ethnicity
8. writing questions for a questionnaire
9. formatting a questionnaire
10. determining the cost per participant for a health program
11. using cost data to make program decisions
12. writing a summary of findings from a data analysis.

*Frequency of practice.* Participants were asked to indicate the number of times they used each of 12 skills listed earlier in the 6 months before the course started (0 times, 1 time, 2 or more times), the number of times they had used the skill on the job since the course ended (6 months), and whether or not they expected to use the skill in the future. Due to a change in the evaluation design during the intervention, this measurement also used the retrospective pretest-posttest design. However, it would have been preferable to use the traditional pretest-posttest design to measure changes in self-reported frequency of practice.

The 6-month follow-up survey also asked whether the participants’ interest in using distance learning for continuing education had increased during the project. Other questions asked if they had shared knowledge and material from the course with coworkers individually or through presentations. Qualitative data were collected through a question asking participants to describe the “ways the EDUSIT course has been useful to you on your job.” Qualitative data have been shown to be an important supplement to quantitative professional training evaluations because changes due to training are shaped by the work.
context and often subtle (Wergin, Mazmanian, Miller, Papp, & Williams, 1988).

**DATA ANALYSIS**

Initial tests examined whether the modules differed significantly from each other in their effects on knowledge, beliefs, and self-efficacy. The modules were evaluated separately for effects on each outcome with a generalized repeated measures analysis of variance using SAS PROC GENMOD involving methods for generalized estimating equations (McCullagh & Nelder, 1989). Differences between the pretest and posttest scores constituted the outcome, and the module was treated as a class variable. The pretest score was included as a covariable, and the correlation structure was exchangeable. No statistically significant differences were found among the effects of the modules on knowledge, beliefs, or self-efficacy.

Because there were no significant differences in the effects of the individual modules, the effect of the intervention on each outcome (knowledge, beliefs, and self-efficacy) was assessed across all modules combined. Specifically, the effect of the intervention on each outcome was assessed by means of paired t tests on the average difference score for each individual across all modules combined for each outcome.

To combine difference scores across modules, each module’s pre- and posttest knowledge and belief scores had to be rescaled such that all had the same range. For instance, in the knowledge outcome, Modules 1, 2, and 5 had scores with a range of 0 to 5. Modules 3, 4, and 6 had ranges of 0 to 7, 0 to 8, and 0 to 6, respectively. Therefore, to combine across all modules for a single knowledge analysis, Modules 3, 4, and 6 had to be rescaled such that their scores also had a range of 0 to 5. Similarly, belief pre- and posttest scores were rescaled such that they ranged from 0 to 3 in all modules. The confidence outcome did not need rescaling as all modules were scored 0 to 3 originally.

Skill level changes for the 12 selected skills were analyzed using the paired samples t test. Changes in frequency of practice of those skills were analyzed using the Wilcoxon signed ranks test (Pett, 1997).
Qualitative data from the evaluation instruments were coded and sorted into themes using the content analysis approach described by Patton (1990). Data were coded and analyzed for major themes.

RESULTS

The 28 project participants completed an average of three of the six modules; only 5 participants completed all six modules.

Participants’ knowledge \((p < .0001)\), beliefs \((p < .01)\), and self-efficacy \((p < .0001)\) all increased significantly during the course (see Table 1).

Twenty-three of the 28 participants (82%) returned the 6-month follow-up survey. Participants’ self-assessed skill levels increased significantly \((p < .001)\) for each of the 12 selected skills and overall for all skills combined (see Table 2). Across all items, the mean skill level increased by 1.75 points on the 7-point scale \((1 = \text{low skill}, 7 = \text{high skill})\), indicating a substantial rise in self-assessed skill levels.

Participants’ self-reported frequency of practice increased significantly \((p < .05)\) for 6 of the 12 skills, and a majority of participants expected to use each of the skills in the future (see Table 3). Among items for which participants did not report a change in practice frequency, most participants expected to use the skills in the future.

Eighteen out of the 23 participants (78%) who responded to the 6-month follow-up survey strongly agreed or agreed that “my interest in using distance learning for continuing education has increased during the past two years.” Seventy-four percent responded that they had “shared knowledge/material from the EDUSIT course with coworkers

<table>
<thead>
<tr>
<th>Construct</th>
<th>N</th>
<th>Pretest Score</th>
<th>Posttest Score</th>
<th>Difference Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
<td>28</td>
<td>3.39 0.76</td>
<td>4.14 0.75</td>
<td>0.75 0.90 4.43</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>28</td>
<td>1.13 0.44</td>
<td>1.94 0.46</td>
<td>0.81 0.47 9.03</td>
</tr>
<tr>
<td>Belief</td>
<td>28</td>
<td>2.40 0.39</td>
<td>2.52 0.35</td>
<td>0.12 0.39 1.63</td>
</tr>
</tbody>
</table>

a. Scale = knowledge (0 to 5), self-efficacy (0 to 3), belief (0 to 3). Individual pre- and posttest scores are averaged across modules for each respondent.
b. The \(t\) and \(p\) values are for the paired samples \(t\) test.
through one-on-one sharing,” whereas 22% had “shared knowledge/material from the EDUSIT course with coworkers by giving a presentation to coworkers on the project or on specific skills learned in the project.”

Qualitative data showed that participants gained from both informational and hands-on instructional methods, including the team project. Participants described gains in knowledge of facts, concepts, and procedures. They noted that they learned skills that were completely new to them and that skills they already had were reinforced. Whether the skills were new or enhanced, participants described many specific ways in which they found the skills useful on their jobs. Learners noted that they were now able to perform the skills taught with greater “confidence,” “comfort,” “efficiency,” speed, creativity, effectiveness, and frequency. As one learner put it,

The course has been very useful. I have used many of the techniques that I used in working with qualitative and quantitative data and mapping in conducting our five-year MCH Block Grant needs assessment.

---

**TABLE 2**

Changes in Skill Levels for Selected Skills Taught in Course (n = 23)

<table>
<thead>
<tr>
<th>Skill</th>
<th>Skill Level Before</th>
<th>Skill Level Now</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selecting an appropriate secondary data source to</td>
<td>3.61 ± 1.37</td>
<td>5.78 ± 0.90*</td>
</tr>
<tr>
<td>investigate a public health problem</td>
<td>4.22 ± 2.09</td>
<td>6.35 ± 0.71*</td>
</tr>
<tr>
<td>Conducting a Web search</td>
<td>3.48 ± 1.53</td>
<td>5.61 ± 0.89*</td>
</tr>
<tr>
<td>Collecting qualitative data through an interview</td>
<td>3.43 ± 1.80</td>
<td>5.09 ± 1.24*</td>
</tr>
<tr>
<td>Analyzing and interpreting qualitative data</td>
<td>3.26 ± 1.54</td>
<td>5.35 ± 1.07*</td>
</tr>
<tr>
<td>Considering implications of data that are reported</td>
<td>2.35 ± 1.27</td>
<td>4.39 ± 1.92*</td>
</tr>
<tr>
<td>by race and ethnicity</td>
<td>4.00 ± 0.90</td>
<td>5.70 ± 0.76*</td>
</tr>
<tr>
<td>Using the guidance of Office of Management and</td>
<td>3.96 ± 1.36</td>
<td>5.57 ± 0.90*</td>
</tr>
<tr>
<td>Budget Directive 15</td>
<td>2.70 ± 1.29</td>
<td>4.30 ± 1.40*</td>
</tr>
<tr>
<td>Writing questions for a questionnaire</td>
<td>2.96 ± 1.58</td>
<td>4.52 ± 1.47*</td>
</tr>
<tr>
<td>Formatting a questionnaire</td>
<td>1.61 ± 1.20</td>
<td>2.91 ± 2.11*</td>
</tr>
<tr>
<td>Determining the cost per participant for a health program</td>
<td>4.00 ± 1.65</td>
<td>5.09 ± 1.56*</td>
</tr>
<tr>
<td>Using cost data to make program decisions</td>
<td>3.30 ± 1.64</td>
<td>5.05 ± 1.56*</td>
</tr>
<tr>
<td>Given a data set and using ArcExplorer, creating a map</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Given results from a data analysis, writing a summary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>of the findings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined (all skills)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Based on a 7-point scale with 1 = low skill, 7 = high skill.  
* p < .05.
DISCUSSION

This study found that an online skills training course can lead to changes in knowledge, beliefs, self-efficacy, perceived skill levels, and practice among health professionals. Hands-on practice with feedback from faculty and staff working in a team and working on a state-level overall project provided helpful support and contexts for learning. Thus, Web-based analytic and technical training initiatives offer a promising means for reaching public health professionals and provide a timely alternative to more resource-intensive off-site workshops.

It has helped me look more clearly at the data generated by our program for program planning purposes.

TABLE 3
Changes in Frequency of Practice for Selected Skills Taught in Course and Expectations of Future Use (n = 23)

<table>
<thead>
<tr>
<th>Skill</th>
<th>Mean Change in Practice Frequency</th>
<th>Expect to Use the Skill in Future (% Yes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selecting an appropriate secondary data source to investigate a public health problem</td>
<td>0.61 ± 0.84**</td>
<td>96</td>
</tr>
<tr>
<td>Conducting a Web search</td>
<td>0.48 ± 0.85*</td>
<td>100</td>
</tr>
<tr>
<td>Collecting qualitative data through an interview</td>
<td>0.61 ± 1.03*</td>
<td>87</td>
</tr>
<tr>
<td>Analyzing and interpreting qualitative data</td>
<td>0.43 ± 0.79*</td>
<td>96</td>
</tr>
<tr>
<td>Considering implications of data that are reported by race and ethnicity</td>
<td>0.39 ± 0.94</td>
<td>96</td>
</tr>
<tr>
<td>Using the guidance of Office and Management Budget Directive 15</td>
<td>0.65 ± 0.71***</td>
<td>74</td>
</tr>
<tr>
<td>Writing questions for a questionnaire</td>
<td>0.17 ± 1.07</td>
<td>91</td>
</tr>
<tr>
<td>Formatting a questionnaire</td>
<td>0.09 ± 1.00</td>
<td>96</td>
</tr>
<tr>
<td>Determining the cost per participant for a health program</td>
<td>0.04 ± 0.77</td>
<td>83</td>
</tr>
<tr>
<td>Using cost data to make program decisions</td>
<td>0.09 ± 0.85</td>
<td>91</td>
</tr>
<tr>
<td>Given a data set and using ArcExplorer, creating a map</td>
<td>0.35 ± 0.65*</td>
<td>78</td>
</tr>
<tr>
<td>Given results from a data analysis, writing a summary of the findings</td>
<td>–0.09 ± 0.79</td>
<td>96</td>
</tr>
<tr>
<td>Combined (all skills)</td>
<td>0.32 ± 0.53*</td>
<td>—</td>
</tr>
</tbody>
</table>

a. Scale: 0 = 0 times used; 1 = 1 time used; 2 = 2 or more times used in the 6 months before the course and in the 6 months since the course had ended.

*p < .05; **p < .01; ***p < .001 for the Wilcoxon signed ranks test.
The primary lesson learned from this project is that on-demand, short, online training opportunities would be more useful for working professionals than the present course structure. Modules that require 15 hours to complete are not compatible with normal professional rhythms. Providing a module of that length without natural stopping points undermines the overall effectiveness of this mode of instruction. Breaking the modules into smaller segments would make it possible for a learner to select those skills needed at that moment (Farel & Polhamus, 2001).

Limitations of our findings include the small number of professionals who participated and the limited geographical region from which participants were drawn. The one-group pretest-posttest design also presents challenges to internal validity. However, participants’ statements provide strong evidence that the gains in performance were indeed attributable to the course.

This methodology may be especially appropriate for an emerging area such as the one described in this article in which it is difficult to identify precisely what parts of the intervention the participants received. However, because there is no guarantee that the use of a true pretest would have produced comparable results, it would be important in future studies to include both a pretest assessed prior to a course such as this and a retrospective pretest in order to compare the results. Threats to internal validity such as history and maturation were not likely to have been operational in this relatively short training course. Participants did not report any simultaneous training interventions that might have influenced the results (Eckert, 2000).

The convenience of using information technology for training public health professionals cannot be disputed. However, the energy and enthusiasm derived from personal interaction must be replicated in the workplace to sustain a commitment to strengthen skills. Public health professionals not only need timely access to Web-based resources but also reinforcement on the job. To provide a climate that rewards learning, organizations should include learning goals in participants’ work plans, provide opportunities for staff members to pursue similar training at the same time for mutual reinforcement, and make continuous on-the-job learning and knowledge sharing normative aspects of their culture (Broad, 1997; Watkins & Marsick, 1993).
REFERENCES


INDEX

TO

EVALUATION & THE HEALTH PROFESSIONS

Vol. 24

No. 1, March 2001 (pp. 1-104)
No. 2, June 2001 (pp. 105-232)
No. 3, September 2001 (pp. 233-360)
No. 4, December 2001 (pp. 361-464)

Authors

ABRAMS, KEITH R., see Song, F.
ANDERSON, MARJORIE C., see Wilson, D. M.
BARON, GENEVIÈVE, PHILIPPE DE WALS, and FRANÇOIS MILORD, “Cost-Effectiveness of a Lottery for Increasing Physicians’ Responses to a Mail Survey,” 47.
BAUSELL, R. BARKER, “Conclusion to the Special Issue on Meta-Analysis,” 235.
BAUSELL, R. BARKER, see Lee, W.-L.
BERMAN, BRIAN M., see Lee, W.-L.
BOLAN, GAIL A., see Fishbein, M.
BOROWSKY, STEVEN J., see Smith, M. A.
BROWN, MARGARET E., see Smith, M. A.
CAPIZZI, JEFFREY, see Hibbard, J. H.
CAREY, MICHAEL P., see Marsh, K. L.
CENTER, BRUCE, see Smith, M. A.
DAVIS, MARGARET K., see Smith, M. A.
DE WALS, PHILIPPE, see Baron, G.
DUBOIS, ANNE, and BRIGITTE SANTOS-EGGMANN, “Evaluation of Patients’ Satisfaction With Hospital-at-Home Care,” 84.
EBRAHIM, SHAH, “The Use of Numbers Needed to Treat Derived From Systematic Reviews and Meta-Analysis: Caveats and Pitfalls,” 152.
EISELE, HEATHER, see Jason, L.
FAINSINGER, ROBIN L., see Wilson, D. M.
FAREL, ANITA, KARL UMBLE, and BARBARA POLHAMUS, “Impact of an Online Analytic Skills Course,” 446.
FEJFAR, MICHELE C., see Rudy, D. W.
FERGUSON, KRISTI J., see Kreiter, C. D.

GREENLICK, MERWYN, see Hibbard, J. H.

GRIFFITH, CHARLES H., III, see Rudy, D. W.


GRUM, CYRIL M., see Williams, B. C.

HENNESSY, MICHAEL, see Fishbein, M.

HIBBARD, JUDITH H., MERWYN GREENLICK, HOLLY JIMISON, JEFFREY CAPIZZI, and LYNN KUNKEL, “The Impact of a Community-Wide Self-Care Information Project on Self-Care and Medical Care Utilization,” 404.

HOXWORTH, TAMARA, see Fishbein, M.

IATESTA, MICHAEL, see Fishbein, M.


JIMISON, HOLLY, see Hibbard, J. H.

JOHNSON, BLAIR T., see Marsh, K. L.

JONES, DAVID R., see Song, F.

KAMB, MARY, see Fishbein, M.

KASPRZYK, DANUTA, DANIEL E. MONTAÑO, JANET S. Sr. LAWRENCE, and WILLIAM R. PHILLIPS, “The Effects of Variations in Mode of Delivery and Monetary Incentive on Physicians’ Responses to a Mailed Survey Assessing STD Practice Patterns,” 3.

KLASSEN, TERRY P., see Pham, B.


KUNKEL, LYNN, see Hibbard, J. H.


LURIE, NICOLE, see Smith, M. A.


MCAULEY, LAURA, see Pham, B.

MILORD, FRANÇOIS, see Baron, G.

MOHER, DAVID, see Pham, B.

MONTAÑO, DANIEL E., see Kasprzyk, D.

NORTHCOTT, HERBERT C., see Wilson, D. M.

OLKIN, INGRAM, see Bravata, D. M.


PHILLIPS, WILLIAM R., see Kasprzyk, D.


PILLSBURY, MATTHEW S., see Williams, B. C.

PLATT, ROBERT, see Pham, B.

POLHAMUS, BARBARA, see Farel, A.

PROJECT RESPECT STUDY GROUP, see Fishbein, M.

RHODES, FEN, see Fishbein, M.


ST. LAWRENCE, JANET S., see Kasprzyk, D.

SANTOS-EGGIMANN, BRIGITTE, see Dubois, A.

SCHMID, CHRISTOPHER H., “Using Bayesian Inference to Perform Meta-Analysis,” 165.


SHELDON, TREVOR A., see Song, F.


SMITH, SUSAN L., see Wilson, D. M.


STERN, DAVID T., see Williams, B. C.

STINGL, MICHAEL J., see Wilson, D. M.

SUTTON, ALEX J., see Song, F.

TAYLOR, RENEE, see Jason, L.


TRUMAN, CORRINE D., see Wilson, D. M.

UMBLE, KARL, see Farel, A.

WILLIAMS, BRENT C., MATTHEW S. PILLSBURY, DAVID T. STERN, and CYRIL M. GRUM, “Comparison of Resident and Medical Student Evaluation of Faculty Teaching,” 53.

WILSON, DAVID B., see Lipsey, M. W.


WILSON, JOHN F., see Griffith, C. H., III.

WILSON, JOHN F., see Rudy, D. W.

ZENILMAN, JONATHAN M., see Fishbein, M.

Articles

“Assessing Attitudes Toward New Names for Chronic Fatigue Syndrome,” Jason et al., 424.

“Comparison of Resident and Medical Student Evaluation of Faculty Teaching,” Williams et al., 53.

“Conclusion to the Special Issue on Meta-Analysis,” Bausell, 235.


“Cost-Effectiveness of a Lottery for Increasing Physicians’ Responses to a Mail Survey,” Baron et al., 47.

“The Effects of Variations in Mode of Delivery and Monetary Incentive on Physicians’ Responses to a Mailed Survey Assessing STD Practice Patterns,” Kasprzyk et al., 3.
“Evaluation of Patients’ Satisfaction With Hospital-at-Home Care,” Dubois and Santos-Eggimann, 84.
“Examining the Generalizability of Ratings Across Clerkships Using a Clinical Evaluation Form,” Kreiter and Ferguson, 36.
“The Growth of Health-Related Meta-Analyses Published from 1980 to 2000,” Lee et al., 327.
“The Impact of a Community-Wide Self-Care Information Project on Self-Care and Medical Care Utilization,” Hibbard et al., 404.
“Impact of an Online Analytic Skills Course,” Farel et al., 446.
“Introduction to the Special Issue,” Bausell, 107.
“Measuring the Physician Perspective on Quality of Care in Health Plans,” Smith et al., 18.
“Methods for Exploring Heterogeneity in Meta-Analysis,” Song et al., 126.
“Missing Predictors in Models of Effect Size,” Pigott, 277.
“Self- and Peer Assessment in a First-Year Communication and Interviewing Course,” Rudy et al., 436.
“Simple Pooling Versus Combining in Meta-Analysis,” Bravata and Olkin, 218.
“The Use of Numbers Needed to Treat Derived From Systematic Reviews and Meta-Analysis: Caveats and Pitfalls,” Ebrahim, 152.
“Using Bayesian Inference to Perform Meta-Analysis,” Schmid, 165.