

# Collection of Family Health Histories: The Link Between Genealogy and Public Health

Donald O. Case

*College of Communications and Information Studies, LCLI 341, University of Kentucky, 500 South Limestone, Lexington, KY 40506-0224. E-mail: dcase@uky.edu*

Although a number of investigations have been conducted on the information behavior of family historians, we know little about the degree to which they systematically collect information on the causes of death and major illnesses of ancestors. Such information, if reliable and accessible, could be useful to family physicians, the families themselves, and to epidemiologists. This article presents findings from a two-stage study of amateur genealogists in the USA. An initial state-wide telephone survey of 901 households was followed by in-depth interviews with a national sample of 23 family historians. Over half of the responding households in the general survey reported that someone in their family collects ancestral medical data; this practice appears to be more common among respondents who are women, older persons, and those with higher incomes. In-depth interviews revealed that this information is commonly collected by family historians, and typically comes from death certificates, secondarily from obituaries, and thirdly from word-of-mouth or family records; most of these respondents collected health information for reasons of surveillance of their own health risks. Social-networking approaches to encourage gathering of family data could aid in increased awareness and surveillance of health risks. Implications for health information seeking and applicable theories are discussed.

## Introduction

According to a poll conducted in 2000 (Maritz, 2000), 60% of the U.S. adult population is interested in genealogy. Furthermore, 25% of Americans (over 60 million people) have conducted genealogical research online (Pew Internet and American Life Project, 2007). That makes the pursuit of family history one of the most popular hobbies in North America.

A growing number of studies have appeared, focused on the information gathering of these amateur genealogists. Starting in 1983 (Sinko & Peters, 1983) and accelerating in

number with the proliferation of genealogy Web sites in the 1990s, most recent investigations have used small samples and have emphasized the use of Internet sources (Duff & Johnson, 2003; Dulong, 1986; Fulton, 2005; Humble, 1999; Kuglin, 2004; Wood, 2004; Yakel, 2004). Notable exceptions to these two generalizations were a postal survey with 1,348 responses (Lambert, 1995a, 1995b, 1996) and Web-based surveys by Drake (2001, with 4,109 respondents) and Veale (2008, with 5,724 replies); all three employed large samples, and the earliest was also conducted before the wide-spread use of the World Wide Web.

Despite a growing number of studies, there is an important element of family history that has been almost entirely ignored in investigations of either genealogists or the general public: the medical data that they collect. The earlier investigations of family historians neglected to ask an important, health-related question: What information do they typically collect regarding the diseases and causes of death of their ancestors and family members? The most explicit mention of health matters is from Lambert, who merely notes that (1995a, p. 229) "Respondents in the health professions, for example, typically expressed an interest in tracking diseases in their family trees"; he does not say how widespread the motivation was among other respondents. Correspondingly, the only relevant survey of the general public was conducted in 2004 by Parade magazine and Research!America (Charlton Research Company, 2004); it found that 96% of Americans thought family history to be important to health, but fewer than 30% said they collected this information from relatives.

## Family History and Medical History

The term *family history*, as typically used in the medical literature, refers to a patient's response to a roster of standard questions asked by a medical doctor or nurse, referring to the incidence of disease among the patient's immediate relatives. Yet other medical publications discuss *family history* in a broader sense: as reports from multiple relatives or other

---

Received March 18, 2008; revised June 19, 2008; accepted July 8, 2008

© 2008 ASIS&T • Published online 14 August 2008 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20938

sources, extending back several generations. Various articles in medical journals—including Guttmacher, Collins, and Carmona (2004); Murff, Spigel, and Syngal (2004); Eerola, Blomqvist, Pukkala, Pyrhonen, and Nevanlinna (2001); Harlow and Fernandez (2005); Hunt, Gwinn, and Adams (2003); Yoon et al. (2002); and Yoon, Scheuner, and Khoury, (2003)—have suggested that this latter type of family medical history is potentially very useful in surveillance of cancer, heart disease, and other illnesses among family members.

Yoon et al. (2002, p. 304) declare that “family history of a common, chronic disease is associated with relative risks ranging from two to five times those of the general population.” One example is Type II diabetes, in which the risk is 2.4 times greater if one’s mother suffers from diabetes, and four times greater if both maternal and paternal relatives have it. A more extreme correlation is found with prostate cancer, in which one’s risk is eleven times greater than normal if three first-degree relatives have been diagnosed with the condition.

While only about 5% of cancer cases have a strong basis in family history, “a family history of cancer often is the strongest known epidemiological factor that can be identified” (Stopfer, 2000, p. 348). Regarding coronary heart disease, the connection between families and individuals is even stronger: According to Kardia, Modell, and Peyser (2003, p. 145) “the evidence appears quite strong that simple family-history tools can be very efficient and accurate ways of assessing familial occurrence of disease.”

### Research Questions

Can the two types of “family medical histories” be usefully connected? Discussions with an officer of the Cancer Information Service (CIS) of the Centers for Disease Control and Prevention (CDC) led me to consider a two-stage investigation of this question. The main objective of the first stage of this investigation was to determine the incidence of collection of information about relatives’ diseases and causes of death. Such an objective connects to the research agenda of the Cancer Information Service, to the concerns of the U.S. Surgeon General’s Family History Initiative (U.S. Department of Health and Human Services, 2007), and to earlier work of mine regarding the uptake and utility of genetic screening tests, with special emphasis on Appalachian populations (Case, Johnson, Andrews, Allard, & Kelly, 2004; Case, Johnson, Andrews, & Allard, 2005). For the initial investigation, the research question was as follows:

- R1: What percentage of the general population collects information on the diseases and causes of death of their family members?

Given an opportunity to contribute a question to a state-wide telephone survey, I saw the chance to obtain at least a partial answer to this question. The results would not, however, provide details regarding specific practices of collecting family health histories. Therefore I later added a second stage to the study (described below), which concerns specific practices among family historians for the collection of medical

information. The following were the main research questions for this stage:

- R2: Who are family historians (in terms of age, gender, experience, and research habits)?  
 R3: How often do they collect information about death and disease among family members?  
 R4: From what sources does information about death and disease originate?  
 R5: For what purpose do they collect such health-related information?

### Stage 1 Methodology

The Stage 1 data were derived from responses to questions that were part of a state-wide telephone survey of noninstitutionalized Kentucky residents 18 years of age or older. I was allowed to ask only one, multipart, question in this survey, which was to limit what I could learn through this method. The survey was conducted from August 14 to September 6, 2006, by the University of Kentucky Survey Research Center. The survey employed the standard Waksberg random-digit dialing procedures, as well as the ACS-Query Computer-Assisted Telephone Interviewing (CATI) system. A total of 2,732 phone numbers were called. There were a total of 314 ineligible persons contacted, and 1,517 refusals or incomplete surveys, resulting in a total of 901 interviews completed. The response rate among eligible persons was 44.5% (CASRO, 2008). The margin of error for the sample was approximately  $\pm 3.3\%$  at the 95 percent confidence level.

The survey as a whole included a total of 96 questions submitted by a number of University of Kentucky researchers on a variety of topics; due to the omnibus nature of the 96 questions, there was no screening of respondents for any particular question (e.g., interest in genealogy). For this project, one 8-part question was included asking about the practices of family historians regarding the collection of health-related information about their relatives and ancestors.

### Telephone Survey Questions

The main variable of interest was measured with a single question: “Does anyone in your extended family keep a record of the following medical conditions among your relatives?” followed by a list of eight conditions: any type of cancer, heart attacks or other heart disease, stroke, diabetes, blindness, asthma, causes of death, and “other major diseases or conditions not already mentioned.”

Several potential independent variables were measured using standard demographic questions (see Table 1). These data included gender, age, ethnicity, marital status, education, employment, total household income (before taxes), and community size.

All data analyses were performed using SPSS Version 12. Both linear and binary logistic regressions were performed to determine which variables were significant predictors of collecting family health history information. A level of 0.05 was selected as the criterion for statistical significance for all analyses.

TABLE 1. Background characteristics of the sample ( $N = 901$ ).

Gender	
Female	64%
Male	36%
Age (in years)	$M = 53$ ( $SD = 16$ )
Marital status	
Married	65%
Not married	35%
Education (summarized from 11 categories)	
Less than high school graduate	13%
High school graduate or equivalency (GED)	36%
Some college	27%
Bachelor's degree	12%
Some graduate school or graduate degree	12%
Household income (summarized from 14 categories)	
\$20,000 or less	23%
\$20,000–40,000	24%
\$40,000–70,000	27%
\$70,000 or more	27%
Racial/Ethnic background	
White	93%
African American	4%
Hispanic, Asian, or Multiple Ethnicities	3%

## Stage 1 Results

### Telephone Survey Sample Characteristics

Of the 901 respondents (see Table 1), 36.4% were male and 63.6% were female, with a mean age of 53. Respondent characteristics were compared with 2005 Kentucky census data estimates (U.S. Census Bureau, 2008). The survey population appeared roughly representative of the population of Kentucky in demographic terms, yet somewhat whiter (93.4% in the sample versus 90.4% in the population), and somewhat better educated (87% high school graduates in the sample, versus 74% in the population). The sample was, however, predominantly female, as is the case with many telephone surveys; women are more likely to be home and to agree to be interviewed. By design the sample was older than the state average for the adult population, as only those 18 years or above were sampled. Given that the main question asked about the practices of the family rather than the individual, however, these demographic differences should not matter greatly.

### Responses to Questions About Family History

A majority of the respondents reported that someone in their family kept track of medical conditions within their family and of ancestors, including causes of death. The most common conditions tracked were blindness (81.3%), asthma (70%) and “any other major disease” (82.6%); most of the other conditions drew a “yes” response from between 53% and 63% of the respondents. See Table 2 for a summary of these responses.

It is intriguing that blindness, asthma, and unspecified “other major diseases” are the most commonly recorded;

TABLE 2. Percentage of sample whose family records medical conditions ( $N = 901$ ).

% Answering “Yes” to Each Medical Condition	
Any other major disease not named below	83%
Blindness	81%
Asthma	70%
Stroke	63%
Cause of death	58%
Diabetes	56%
Heart attack or cardiac	54%
Cancer	53%

however, it may be that these are more easily observed than internal conditions, such as cancer and heart disease. Likely candidates for these “other major diseases” could be arthritis, respiratory conditions, or allergies; a longer checklist would have helped identify them. Perhaps respondents were thinking of the possibilities found on the extensive checklists frequently used by physicians in taking a family medical history.

### Logistic Regressions

Several binomial logistic regressions on the different outcome variables (e.g., collecting information on relatives' blindness or asthma) were conducted to determine predictors of whether or not respondents would record data on health conditions of their relatives. In this case, logistic regression analysis creates an equation that predicts whether respondents will say “yes” or “no” to the question about whether their family records such information. The results are expressed as an *odds ratio*, a measure of effect size in which the ratio of the odds of a “yes” response occurring are compared to the odds of a “no” response, for the particular variables examined. An odds ratio of 1 indicates that either response would be equally likely. See Table 3 for the details of these results, by medical condition recorded.

For most of the analyses two of four variables—gender, and either age, household income before taxes (as a set of 14 categories), or presence of elderly in the household—proved to be predictive of responding “yes” to each of the eight questions.

In the case of collecting relatives' causes of death and instances of stroke, the most powerful predictor was female gender, followed by income. However, for both cancer and heart disease as recorded conditions, the two most predictive variables were gender and age. Regarding the remaining health conditions, the results were either other inconclusive or weak. No variables predict the recording of blind relatives, for example. For asthma, only age was strongly predictive, and female gender marginally so. For “any other major disease not mentioned” only female gender and the presence of elderly (65 years or older) in the household were predictive of data collection.

In summary, the two most predictive variables for each health condition were used to classify respondents as to their

TABLE 3. Logistic regression of yes-no responses for tracking family health conditions.

Health condition	Predictive variable	Odds ratio	95% Confidence ratio interval (lower-upper)	Significance	Model
Cause of death	Female gender	1.65	1.22–2.27	$p < 0.01$	$\chi^2(2, N = 901) = 14.28, p < 0.01$
	Total household income	1.06	1.01–1.11	$p = 0.01$	
Cancer	Female gender	1.55	1.16–2.07	$p < 0.01$	$\chi^2(2, N = 901) = 19.68, p < 0.01$
	Age (in years)	1.01	1.01–1.02	$p < 0.01$	
Heart disease	Female gender	1.41	1.06–1.87	$p = 0.02$	$\chi^2(2, N = 901) = 20.20, p < 0.01$
	Age (in years)	1.02	1.01–1.03	$p < 0.01$	
Diabetes	Female gender	1.38	1.04–1.84	$p < 0.03$	$\chi^2(2, N = 901) = 23.37, p < 0.01$
	Age (in years)	1.02	1.01–1.03	$p < 0.01$	
Stroke	Female gender	1.51	1.09–2.09	$p = 0.01$	$\chi^2(2, N = 901) = 15.07, p < 0.01$
	Total household income	1.08	1.03–1.14	$p < 0.01$	

likelihood of collecting information. The results revealed that information seeking regarding four health conditions (causes of death, cancer, heart disease, and diabetes) could be predicted by examining the two demographic variables in each model (gender, and either age, income, or presence of elderly in household). For another two conditions, stroke and “other diseases,” gender and income were somewhat predictive.

It is important to note that, despite robust levels of significance, the success at correct classification was relatively modest: from 57.3 to 69.5% correct in the successful analyses; in other words, the formulas were only able to improve on chance assignment to a “yes” or “no” answer by only a few percentage points in several of the conditions.

#### Multiple-Regression Analysis

One final statistical analysis was conducted. A multiple-regression analysis was conducted to assess how well female gender and household income predicted how many “yes” responses were given to the series of eight questions (see Table 4). An index of “yes” responses was created (mean of 3.27, median of 2), and regressed with the two most predictive variables: gender and total household income before taxes. With two predictor variables entered simultaneously, the model was significant,  $F(2, 694) = 7.436, p = .001$ . However, only a tiny percentage ( $R^2 = 0.021$ ) of the variance was accounted for by the two predictor variables. Individual coefficients assessed how well each alone predicted the criterion variable. Female gender was the strongest predictor of recording multiple health conditions ( $Beta = -0.120, p < 0.01$ ). Household income also predicted more collection of health data ( $Beta = -0.106, p = 0.01$ ). Given the frequent

number of “yes” responses there could be a ceiling effect for the index, which would render the statistics less powerful.

That variables such as gender, income, and age can predict information practices, of course, not surprising in itself; this has been a common finding of other information-seeking studies. Women are much more likely to practice genealogy than men, for example, and older and wealthier people often have more free time to spend on hobbies. Being older oneself, or having an older adult around the home would likely encourage the respondent to be more aware of medical conditions.

#### Stage 1 Limitations

Given that this investigation took place in the context of an omnibus telephone survey of 96 disparate questions, the number of questions that could be asked about information practices was limited to one. Due to this fact, there are a number of things we don’t know about the responses, particularly their accuracy. In addition to the usual problems with self-reports of behavior, the initial question (“Does anyone in your extended family keep a record . . .”) asks about the behavior of others as well as oneself. It is easy to imagine that responses would overestimate the degree to which health conditions are recorded by other family members in cases in which the respondent was not the one doing the recording. In addition the sample may not be representative of a larger population. While this was a random sample of 2,732 phone numbers (from among a population of more than 4 million), there were many refusals, leading to an effective response rate of just 44.5% of eligible respondents. We know that the result over-sampled women (63.6% of the respondents), a common problem in surveys of this type. Also, it may be that the population of Kentucky differs in systematic ways that bias the results.

More important than issues of sampling are problems with the unit of analysis. The survey asked whether someone in the family collects health-related information on the family; we do not know if the person who answered the survey was indeed the person who records such data, or if someone else in the family does. So the demographic data is of limited

TABLE 4. Summary of multiple regression analysis for variables predicting collection of facts regarding health problems of relatives ( $N = 901$ ).

Variable	B	SEB	Beta
Female gender	0.81	0.26	0.12*
Household income	0.10	0.04	0.11*

Note.  $R^2 = .02$ . \* $p < .01$

use, as gender and age apply to the individual level of analysis rather than the household; however, household income and family composition data *do* apply to the individual answering the questions.

So there are several reasons why these data are flawed and might be inconclusive; a fuller study of another large sample, with more specific questions, could shed light on the true incidence and depth of the collection of family health history. The second-stage interviews were intended to clarify how extensive and reliable are the data collected by family historians.

## Stage 2 Methodology

The telephone survey had two troubling limitations. First, I was restricted to a single question. Also, the methodology did not allow for any screening that would have allowed me to focus on families who practiced genealogy. Therefore I followed up on the telephone survey with in-depth interviews with a sample of family historians. The goal of this stage of the investigation was to learn more about the practices of amateur genealogists, what kinds of health data they collect (if any), the source and accuracy of such data, and how people view genetic influences on their health. A total of 29 questions were addressed in the 30-to-45-minute interviews.

## Stage 2 Results

### *Sample Characteristics*

In Stage 2 a nonrandom sample of 23 respondents were interviewed. Participants for the interviews were solicited through the Kentucky Genealogical Society (13 respondents) and Internet genealogical discussion groups (10 participants, originating from eight different U.S. states). Between April 2007, and January 2008, 23 interviews were conducted (20 by telephone and three in person). Thirteen respondents (57%) were female and 43% male—a gender breakdown parallel to the Stage 1 sample, and possibly representative of a hobby that is predominately female (e.g., in Drake's 2001 study of 4,109 genealogists, 72% were female, and among Veale's 2008 sample of 5,724, or 71%, were female). Seven respondents were in their 50s, seven in their 60s, and four in their 70s, with only five being younger than 50 years of age. They had been collecting information on their families for between 2 and 50 years, with an average of 19 years of experience. The amount of time spent on genealogy ranged from an estimated 1 hour per month to 160 hours, with an average of 37 hours per month (about half of the sample were retired individuals, and two were doing full-time research for other family historians, which accounts for these high numbers) and a median of 18 hours/month. Every respondent used a computer for keeping track of family history information, and all made ample use of the Internet, particularly the Ancestry.com site, which is the largest gateway to genealogy databases and discussion lists. Thirteen respondents were able to say approximately (sometimes exactly) how many unique individuals were recorded

in their personal GEDCOM database; totals ranged from 500 to 60,000 names, with an average of roughly 16,000 and a median of 4,600 names.

Regarding the key health-related questions indicated above, 43% of those interviewed said that they "always" record ancestors' cause of death, and 26% "almost always" record it, whenever such information is available; only 9% "never" record cause of death. Respondents said that the source of that information is typically a death certificate (61%), less often an obituary (39%; multiple answers were allowed), or and sometimes another source, such as a family Bible or reports of family members (i.e., "word-of-mouth"; 30%). While death certificates are the most authoritative source, problems exist with these, including multiple causes of death, wrong diagnosis (surprisingly common in autopsies, according to Ravakhah, 2006, and Kircher, Nelson, & Burdo, 1985), and antique terminology (e.g., "dropsy") that does not map directly onto current diagnoses. Obituaries may have additional problems, particularly vague terminology due to ignorance or politeness.

Regarding major illnesses, the pattern is less clear. Sixty-one percent of respondents said that they "sometimes" capture such data, and 30% "always" or "almost always" do; yet every single respondent pointed out that such information is often missing. When available, information about disease tends to come from personal documents (e.g., individual letters or narratives), which are obviously of limited reliability. The only reliable sources of disease information are death certificates, and (more rarely) special files kept by public-health agencies during epidemics of cholera, yellow fever, or flu.

Regarding the purpose of recording family health information over multiple generations, the most common response was that it could be useful in spotting "trends," i.e., in surveillance of potential risks; 86% said that they did see such health trends, the most common being some form of heart disease, identified by 53%, followed by cancer (32%), tuberculosis (21%), and stroke (13%). Seven other conditions were mentioned. A small number said that they collected health information simply because they collected every fact they could find.

### *Stage 2 Limitations*

Obviously the data gathered in Stage 2 suffers from a small, nonrandom sample; they may not be representative of family historians in North America, much less elsewhere. Also, in all cases respondent self-reports could not be verified, and may suffer from the demand characteristics of the interview situation.

## Discussion

One might reasonably ask whether there is anything to be gained by questioning amateur genealogists and other non-health professionals whether they collect health-related information, especially given problems with the reliability of information collected. There is an important reason for doing

so, despite the aforementioned limitations of the present studies: to form an understanding of this particular activity and social group, such that the considerable energy that they devote to their hobby can be harnessed in the service of promoting public health. Indeed, the motivation for the U.S. Surgeon General's Family History Initiative is to promote awareness of genetic and lifestyle risk factors, and to encourage adults to record relevant data to pass on to both their physicians and their children. Several years of publicity campaigns by health organizations have made only modest changes in the behaviors of both the public and their physicians, less than half of whom take a complete health history of two or more generations (Rich, et al., 2004). Thus, a new approach needs to be developed.

### *The Link between Genealogy and Public Health*

Because of the disease burden imposed on society by various forms of cancer, along with heart disease, diabetes, asthma, and a few other common illnesses, public-health officials in the United States and other nations have repeatedly sponsored public-information campaigns (Backer, Rogers, & Sopory, 1992; Fishbein & Cappella, 2006; Hornik, 2002; Rice & Atkin, 2001) to call more attention to such diseases and change individual behaviors that might contribute to their incidence.

In the past most public-information campaigns employed the mass media (television, radio, and print) to spread the word regarding health hazards and healthy behaviors. In recent years an increasing awareness of the importance of social networks has led to more community-based approaches to influencing public knowledge and behavior (Lomas, 1998; Renger, Steinfeldt & Lazarus, 2002; Viswanath, 2008). For example, Viswanath, Steele, and Finnegan (2006) make the case that virtual communities offer an opportunity to change health outcomes through exploitation of the social ties that bind such groups together, because their electronic interactions are "likely to contribute to interpersonal discussion and interactions that may heighten attention to messages in the environment that could lead to better learning about health" (p. 1460).

The evidence is that patients often seek information from sources outside the health care system—from the mass media, telephone helplines, friends, relatives, and, increasingly, via online sources (Carlsson, 2000; Dutta-Bergman, 2005; Lambert & Loiselle, 2007; Rice & Katz, 2001; Snipes, Ingram & Jiang, 2005). In addition, various scholars (e.g., Cassell, Jackson, & Cheuvront, 1998; Cline & Haynes, 2001; Hardey, 2001; Neuhauser & Kreps, 2003) make the important point that health-related Internet sites result in patients forming relationships with other health-care consumers, transforming them into *producers* of health knowledge (i.e., providing information about themselves interactively), rather than merely passively consuming health information posted by organizations. The high-profile, electronic health-record projects of Microsoft and Google will accelerate this trend (Lohr, 2008; Schonfeld, 2008). Both Microsoft and Google

have devised Web-based systems for patients to keep their own medical records online, based on information they exchange with their health-care provider. Both systems anticipate the eventual arrival of personal health records (PHR), probably in the form of persistent URLs.

As indicated by various earlier studies (Veale, 2008; Yakel, 2004) family historians already have extensive online communities composed of individuals who correspond regularly via message boards and through email. Veale (2004) explains that

As a hobby, genealogists rarely participate in isolation—rather genealogy requires, and creates, community . . . genealogists correspond with others to enquire on kinship connections, previous research, and sources of family history data. The relationships forged are rarely discarded; rather, initial conversations often turn into long-term relationships.

Some of these communities are formed based on where members live, ranging from geographic entities like cities, counties, states, provinces, and regions; for example, most counties in the eastern United States have a local genealogy Web site and message board. The largest international site, Ancestry.com, claims nearly 3 million members who have posted over 17 million messages on over 161,000 message boards. Other online groups are formed by people with the same or similar last names, irrespective of geography; FamilyTreeDNA, one the largest of the genetic testing labs for genealogy, sponsors nearly 5,000 family name projects, which in turn involve over 200,000 individuals. These individuals can be both motivated to compile and update comprehensive family medical histories, and to share these histories among their family members and with their own physicians—some of whom who must also be encouraged to recognize the importance of family history (Yoon, Scheuner, & Khoury, 2003).

Thus, one approach to addressing the problem of inadequate family health histories is to draw upon the enthusiasm of self-organized clusters of family historians, who already have extensive social networks. These genealogical communities, especially those sponsored by the DNA-testing labs, are ripe grounds for the promotion of what Dutta-Bergman (2005, p. 3) calls "health consciousness," i.e., "enduring involvement in the domain of health . . . a continual interest in health issues."

### *Implications for Theory*

The implications for theory are several. It has long been appreciated that issue salience is key to motivating both information seeking and subsequent action based on the information obtained (e.g., Moorman & Matulich, 1993; Dutta-Bergman, 2005). Taking the salience of cancer risk as one example: Over 61% of the adult population of the United States have "close family members" who have had cancer. In an earlier survey of Kentucky residents (Case et al., 2004, 2005; Kelly, Andrews, Case, Allred, & Johnson, 2007), 47% said that cancer "ran in their family," and over a third stated



that they worry (“often” or “sometimes”) about their familial susceptibility to cancer. Thus, many people have a family history of cancer or other inheritable risks, are concerned about those risks, and so are open to influence regarding the topic. Czaja, Manfredi, and Price (2003, p. 531) say that studies of cancer patients and their families find that a key contextual determinant of health-related information seeking is “concern about getting cancer and having a family member or personal friend treated for cancer.”

Theories about promoting healthy lifestyles via persuasive means typically make use of “antecedents” or “background influences,” like personal experience, perceived risk for disease, and individual demographics, in order to shape attitudes, beliefs, intentions, and behaviors through some kind of intervention, such as a mass-media campaign (Fishbein & Capella, 2006). As Noar (2006) points out in his review of reports since 1996, at least 14 types of theory have been applied in recent health campaigns; among the most common have been the health-belief model, the theory of reasoned action, social-learning theory, social-cognitive theory, and the stages of change model.

Noar’s (2006) review suggests that future approaches should encourage interpersonal communication, make use of communities and social networks, and might involve “interactive health communication” (p. 32). Cassell et al. (1998), along with Cline and Haynes (2001), also advocate these ideas as regards the Internet. Cassell et al. (1998, p. 77) describe the Internet as “a hybrid channel with the persuasive capabilities of interpersonal communication” with great potential for “Internet-based public-health interventions.” Cline and Haynes (p. 687) note that “framing Internet use as health communication invites social-systems and social-influence theoretical frameworks.”

In the context of the present investigation, targeting online genealogical communities and recruiting their members as change agents would be in line with “stages of change” trans-theoretical models, such as that of Prochaska, DiClemente, and Norcross (1992), a point also made by Cassell et al. (1998) in regards to Internet interventions in general. In a review of hundreds of studies of addictive behaviors (e.g., smoking, drinking, overeating) Prochaska and his colleagues concluded that individuals cycle through five stages on their way to successful behavior changes: precontemplation, contemplation, preparation, action, and maintenance. In an initial, precontemplation, stage, family historians and their families might be asked simply to collect information in such a way that it can be used by family members to better inform their health-care providers of possible risks; they are not asked to achieve the more difficult goal of ceasing risky behaviors like smoking, or implementing changes in diet or exercise (if needed). Yet, by collecting their own information about the pattern of illnesses in their own family, they will be primed for later messages about these additional steps in the contemplation and preparation stages. In the meantime, they will have provided useful information to the health-care system that may lead to more careful monitoring of health risks by their physician.

This confluence of circumstances—a renewed appreciation for the power of social networks and the emergence of the Internet for trading of health and other personal information—can be exploited in the service of public health. A concerted effort to collect information on more distant relatives could be reasonably reliable and add valuable data to both individual health records and existing registries for cancer and other preventable conditions.

## Acknowledgments

A preliminary version of this paper was presented at the Twelfth International Symposium on Health Information Management Research, held at Sheffield, UK, 18–20 July, 2007, and appears on pages 157–166 of the Proceedings of the symposium, under the title “Do genealogists collect medical histories?”

## References

- Backer, T.E., Rogers, E.M., & Sopory, P. (1992). Designing health communication campaigns: What works? Thousand Oaks, CA: Sage.
- Carlsson, M.E. (2000). Cancer patients seeking information from sources outside the health care system. *Supportive Care in Cancer*, 8, 453–457.
- Case, D.O., Johnson, J.D., Andrews, J.E., & Allard, S.L. (2005). Avoiding versus seeking: The relationship of information seeking to avoidance, blunting, coping, dissonance and related concepts. *Journal of the Medical Library Association*, 93, 353–362.
- Case, D.O., Johnson, J.D., Andrews, J.E., Allard, S.L., & Kelly, K.M. (2004). From two-step flow to the Internet: The changing array of sources for genetics information seeking. *Journal of the American Society for Information Science and Technology*, 55, 660–669.
- CASRO (2008). On the definition of response rates. Port Jefferson, NY: Council of American Survey Research Organizations. Retrieved March 1, 2008, from <http://www.casro.org/resprates.cfm>
- Cassell, M.M., Jackson, C., & Cheuvront, B. (1998). Health communication on the Internet: An effective channel for health behavior change. *Journal of Health Communication*, 3, 71–79.
- Charlton Research Company (2004). Taking our pulse: The PARADE/Research!America health poll. Washington, DC. Retrieved August 1, 2008, from [http://www.nlm.nih.gov/archive/20060215/csi/research\\_america\\_data.pdf](http://www.nlm.nih.gov/archive/20060215/csi/research_america_data.pdf)
- Cline, R.J., & Haynes, K.M. (2001). Consumer health information seeking on the Internet: The state of the art. *Health Education Research*, 16, 671–692.
- Czaja, R., Manfredi, C., & Price, J. (2003). The determinants and consequences of information seeking among cancer patients. *Journal of Health Communication*, 88, 529–562.
- Drake, P.J. (2001). Findings from the Fullerton genealogy study. Unpublished master’s thesis, California State University, Fullerton. Retrieved June 18, 2008, from <http://psych.fullerton.edu/genealogy/>
- Duff, W.M., & Johnson, C.A. (2003). Where is the list with all the names? Information-seeking behavior of genealogists. *The American Archivist*, 66, 79–95.
- Dulong, J.P. (1986). Genealogical groups in a changing organizational environment: From lineage to heritage. Unpublished doctoral dissertation, Wayne State University, Detroit.
- Dutta-Bergman, M.J. (2005). Developing a profile of consumer intention to seek out additional information beyond a doctor: The role of communicative and motivation variables. *Health Communication*, 17, 1–16.
- Eerola, H., Blomqvist, C., Pukkala, E., Pyrhonen, S., & Nevanlinna, H. (2001). Familial breast cancer in southern Finland: How prevalent are breast cancer families and can we trust the family history reported by patients? *European Journal of Cancer*, 36, 1143–1148.

- Fishbein, M., & Capella, J.N. (2006). The role of theory in developing effective health communications. *Journal of Communication*, 56, S1–S17.
- Fulton, C. (2005). Finding pleasure in information seeking: Leisure and amateur genealogists exploring their Irish ancestry. In A. Grove, (Ed.), *Proceedings of the ASIST Annual Meeting*, 42(1), 1292–1303.
- Guttmacher, A.E., Collins, F.S., & Carmona, R.H. (2004). The family history: More important than ever. *New England Journal of Medicine*, 351, 2333–2336.
- Harlow, I., & Fernandez, C. (2005). The oral tradition meets the genetic family history: Anthropologists, folklorists, genetic health specialists, and consumers unite. *The Genetic Family History in Practice*, 2(3), 4.
- Hardey, M. (2001). "E-health": The Internet and the transformation of patients into consumers and producers of health knowledge. *Information, Communication, & Society*, 4, 388–405.
- Hornik, R.C. (2002). Public health communication: Evidence for behavior change. Mahwah, NJ: Lawrence Erlbaum.
- Humble, G.B. (1999). Genealogy and the Internet. Unpublished master's thesis, Charles Sturt University, Albury-Wodonga, New South Wales, Australia.
- Hunt, S., Gwinn, M., & Adams, T. (2003). Family history assessment: Strategies for prevention of cardiovascular disease. *American Journal of Preventive Medicine*, 24, 136–142.
- Kardia, S.L.R., Modell, S.M., & Peyser, P.A. (2003). Family-centered approaches to understanding and preventing coronary heart disease. *American Journal of Preventive Medicine*, 24, 143–151.
- Kelly, K.M., Andrews, J.E., Case, D.O., Allard, S.L., & Johnson, J.D. (2007). Information seeking and intentions to have genetic testing for hereditary cancers in rural and Appalachian Kentuckians. *Journal of Rural Health*, 23(2), 166–172.
- Kircher, T., Nelson, J., & Burdo, H. (1985). The autopsy as a measure of accuracy of the death certificate. *New England Journal of Medicine*, 313, 1263–1269.
- Kuglin, H. (2004). Granny hunting: The information-seeking behavior of genealogists in open access libraries: A quantitative study. In *Proceedings of Library & Information Association of New Zealand (LIANZA)*. Retrieved August 6, 2008, from <http://opac.lianza.org.nz/cgi-bin/koha/opac-detail.pl?bib=239>
- Lambert, R.D. (1995a). Becoming a family historian. *Families*, 34, 223–32.
- Lambert, R.D. (1995b). Looking for genealogical motivation. *Families*, 34, 73–80.
- Lambert, R.D. (1996). Doing family history. *Families*, 35, 11–25.
- Lambert, S.D., & Loiselle, C.G. (2007). Health information seeking behavior. *Qualitative Health Research*, 17, 1006–1019.
- Lohr, S. (2008, February 21). Google Health begins its preseason at Cleveland Clinic. *New York Times* [Electronic edition]. Retrieved June 7, 2008, from <http://bits.blogs.nytimes.com/2008/02/21/google-health-begins-its-preseason-at-cleveland-clinic/?ref=technology>
- Lomas, J. (1998). Social capital and health: Implications for public health and epidemiology. *Social Science and Medicine*, 47, 1181–1188.
- Maritz (2000). Recent Maritz poll shows explosion in popularity of genealogy. Retrieved May 16, 2006, from <http://www.genealogy.com/genealogy/press-051600.html>
- Moorman, C., & Matulich, E. (1993). A model of consumers' preventive health behaviors: The role of health motivation and health ability. *Journal of Consumer Research*, 20, 208–229.
- Murff, H.J., Spigel, D.R., & Syngal, S. (2004). Does this patient have a family history of cancer? *Journal of the American Medical Association*, 292, 1480–1489.
- Neuhauser, L., & Kreps, G.L. (2003). Rethinking communication in the e-health era. *Journal of Health Psychology*, 8, 7–23.
- Noar, S.M. (2006). A 10-year retrospective of research in health mass-media campaigns: Where do we go from here? *Journal of Health Communication*, 11, 21–42.
- Pew Internet and American Life Project (2006). Internet activities. Retrieved March 6, 2008, from [http://www.pewinternet.org/trends/Internet\\_Activities\\_1.11.07.htm](http://www.pewinternet.org/trends/Internet_Activities_1.11.07.htm)
- Prochaska, J.O., DiClemente, C.C., & Norcross, J.C. (1992). In search of how people change: Applications to addictive behaviors. *American Psychologist*, 47, 1102–1114.
- Ravakhah, K. (2006). Death certificates are not reliable: Revivification of the autopsy. *Southern Medical Journal*, 99, 728–733.
- Renger, R., Steinfeld, V., & Lazarus, S. (2002). Assessing the effectiveness of a community-based media campaign targeting physical inactivity. *Family and Community Health*, 25(3), 18–30.
- Rice, R.E., & Atkin, C.K. (2001). Public communication campaigns. Thousand Oaks, CA: Sage.
- Rice, R.E., & Katz, J.E. (2001). The Internet and health communication: Experiences and expectations. Thousand Oaks, CA: Sage.
- Rich, E., Burke, W., Heaton, C., Haga, S., Pinsky, L., Short, M., & Acheson, L. (2004). Reconsidering the family history in primary care. *Journal of General Internal Medicine*, 19, 273–280.
- Schonfeld, E. (2008). Google Health: A quick, hands-on look. Retrieved June 17, 2008, from <http://www.techcrunch.com/2008/05/19/google-health-a-quick-peek/>
- Sinko, P.T., & Peters, S.N. (1983). A survey of genealogists at the Newberry Library. *Library Trends*, 32, 97–109.
- Snipes, R.L., Ingram, R., & Jiang, P. (2005). Information search in health care decision-making: A study of word-of-mouth and Internet information users. *Journal of Hospital Marketing & Public Relations*, 15(2), 33–53.
- Stopfer, J.E. (2000). Genetic counseling and clinical cancer genetics services. *Seminars in Surgical Oncology*, 18, 347–357.
- U.S. Census Bureau (2008). State and county quick facts: Kentucky 2007. Washington, DC: Author. Retrieved March 6, 2008, from <http://quickfacts.census.gov/qfd/states/21000.html>
- U.S. Department of Health and Human Services (2007). U.S. surgeon general's family history initiative. Washington, DC: Author. Retrieved March 1, 2008, from <http://www.hhs.gov/familyhistory>
- Veale, K.J. (2004). Discussing our family trees: A longitudinal analysis of online, community-based communication in genealogical newsgroups. *Interface: The Journal for Education, Community, and Values*. Retrieved June 17, 2008, from <http://bcis.pacificu.edu/journal/2004/04/veale.php>
- Veale, K.J. (2008). The Internet genealogical community study. Retrieved June 17, 2008, from <http://www.veale.com.au/phd/html/survey.htm>
- Viswanath, K. (2008). Social capital and health. New York: Springer.
- Viswanath, K., Steele, & Finnegan, J.R., Jr. (2006). Social capital and health: Civic engagement, community size, and recall of health messages. *American Journal of Public Health*, 96, 1456–1461.
- Wood, C.A. (2004). Toward the new genealogy: Genealogical research in archives and the Saskatchewan Genealogical Society, 1969–2004. Unpublished master's thesis, University of Manitoba, Winnipeg.
- Yakel, E. (2004). Seeking information, seeking connections, seeking meaning: Genealogists and family historians. *Information Research*, 10(1). Retrieved March 1, 2008, from [informationr.net/ir/10-1/paper205.html](http://informationr.net/ir/10-1/paper205.html)
- Yoon, P.W., Scheuner, M.T., & Khoury, M.J. (2003). Research priorities for evaluating family history in the prevention of common chronic diseases. *American Journal of Preventive Medicine*, 24, 128–135.
- Yoon, P.W., Scheuner, M.T., Peterson-Oehlke, K.L., Gwinn, M., Faucett, A., & Khoury, M.J. (2002). Can family history be used as a tool for public health and preventative medicine? *Genetic Medicine*, 4, 304–310.