June 2000

Automated Forms Processing and Automated Data Capture in Public Health

Jiliana R. Fenster

Follow this and additional works at: https://opencommons.uconn.edu/uchgs_masters

Recommended Citation
https://opencommons.uconn.edu/uchgs_masters/31
AUTOMATED FORMS PROCESSING AND
AUTOMATED DATA CAPTURE IN PUBLIC HEALTH

Juliane R. Fenster

B.A., University of Michigan, 1985
M.S., University of Massachusetts, 1988

A Thesis
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Public Health
at the
University of Connecticut
2000
Master of Public Health Thesis

Automated Forms Processing and Automated Data Capture in Public Health

Presented by

Juliane R. Fenster, M.S.

University of Connecticut Health Center

2000
ACKNOWLEDGMENTS

Although I have spent a great deal of time completing this project, there are a many people who have contributed over the years and made this experience possible and even slightly enjoyable.

First, I would like to thank my advisor, Dr. Stephen Walsh, for his patience, support, and guidance. I would also like to extend a warm thank you to Dr. Joseph Sheehan for “pinch hitting” in my time of need. Thank you to Dr. Jonathan Clive and Debbie Paturzo for hanging in there and reviewing this lengthy paper.

I would like to strongly acknowledge the support of the faculty and staff at the Claude Pepper Older Americans Independence Center. This project would not have been completed without their approval and support.

A warm thank you to Nita Thompson for her advice, problem solving, and friendship. To Annmarie, for her support, dog sitting, and big heart. Special thanks to Sandy for “getting me through.” Thank you to Debbie Garet, Joan Donnelly, Michelle Ahern, Alison Kleppinger, Cagney, and HP. And finally, I would like to thank my mom and dad for their support and love. I would also like to dedicate this project to the memory of my wonderful grandfather, Jack Rothman.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>1. DEVELOPMENT OF THE PROBLEM</td>
<td>1</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>5</td>
</tr>
<tr>
<td>2. REVIEW OF LITERATURE</td>
<td>6</td>
</tr>
<tr>
<td>Data Management and Research in Public Health</td>
<td>6</td>
</tr>
<tr>
<td>Assimilation and Accessibility of Public Health Data</td>
<td>6</td>
</tr>
<tr>
<td>Assessing Needs for Health and Health Services</td>
<td>9</td>
</tr>
<tr>
<td>The Collection of Adequate and Appropriate Data</td>
<td>13</td>
</tr>
<tr>
<td>Data Standards</td>
<td>17</td>
</tr>
<tr>
<td>Methodologies for Data Collection</td>
<td>20</td>
</tr>
<tr>
<td>Data Transcription and Acquisition</td>
<td>20</td>
</tr>
<tr>
<td>Computer Assisted Data Collection</td>
<td>22</td>
</tr>
<tr>
<td>Double Data Entry</td>
<td>24</td>
</tr>
<tr>
<td>Automated Data Capture</td>
<td>28</td>
</tr>
<tr>
<td>Bar Coding</td>
<td>29</td>
</tr>
<tr>
<td>Automated Forms Processing and Scanning Technology</td>
<td>34</td>
</tr>
<tr>
<td>Summary and Conclusions</td>
<td>47</td>
</tr>
<tr>
<td>3. METHODOLOGY</td>
<td>50</td>
</tr>
<tr>
<td>Volunteers</td>
<td>50</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Summary of Duplicate Data Entry Methods, Ordered by Year</td>
<td>25</td>
</tr>
<tr>
<td>2.</td>
<td>Summary of Accuracy Rates by Data Entry Methods</td>
<td>49</td>
</tr>
<tr>
<td>3.</td>
<td>Volunteer Demographics by Type of Research Experience</td>
<td>56</td>
</tr>
<tr>
<td>4.</td>
<td>Form Demographics</td>
<td>57</td>
</tr>
<tr>
<td>5.</td>
<td>Mean Number of Constrained Print Fields to Verify</td>
<td>58</td>
</tr>
<tr>
<td>6.</td>
<td>Mean Number of Fields to Correct During Verification</td>
<td>60</td>
</tr>
<tr>
<td>7.</td>
<td>Number of Character Field Substitution Errors After Verification</td>
<td>61</td>
</tr>
<tr>
<td>8.</td>
<td>Percent of Character Field Substitutions Errors in Data Set</td>
<td>63</td>
</tr>
<tr>
<td>9.</td>
<td>Number of Choice Fields to Verify</td>
<td>64</td>
</tr>
<tr>
<td>10.</td>
<td>Number of Choice Fields to Correct During Verification</td>
<td>65</td>
</tr>
<tr>
<td>11.</td>
<td>Number of Choice Field Errors in Data Set</td>
<td>66</td>
</tr>
<tr>
<td>12.</td>
<td>Percent of Choice Field Errors in Data Set</td>
<td>68</td>
</tr>
<tr>
<td>13.</td>
<td>Comparison of Those with and Without Teleform Experience</td>
<td>69</td>
</tr>
<tr>
<td>14.</td>
<td>Comparison of Constrained Print Fields with Choice Fields</td>
<td>69</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Constrained Print Fields Include Both Text and Numeric Fields</td>
<td>54</td>
</tr>
<tr>
<td>2</td>
<td>Example of Choice Fields</td>
<td>54</td>
</tr>
<tr>
<td>3</td>
<td>Mean Number of Character Fields to Verify</td>
<td>59</td>
</tr>
<tr>
<td>4</td>
<td>Mean Number of Character Fields to Correct During Verification</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>Mean Number of Character Field Substitution Errors</td>
<td>62</td>
</tr>
<tr>
<td>6</td>
<td>Mean Number of Choice Fields to Verify</td>
<td>64</td>
</tr>
<tr>
<td>7</td>
<td>Mean Number of Choice Fields Errors</td>
<td>67</td>
</tr>
<tr>
<td>8</td>
<td>Percent of Character Fields and Choice Fields to Verify</td>
<td>72</td>
</tr>
<tr>
<td>9</td>
<td>Percent of Character Fields and Choice Fields to Correct</td>
<td>72</td>
</tr>
<tr>
<td>10</td>
<td>Percent of Character Field and Choice Field Errors</td>
<td>73</td>
</tr>
</tbody>
</table>
Chapter 1

DEVELOPMENT OF THE PROBLEM

Introduction

Knowledge of human health and disease may be obtained from a variety of disciplines and resources. These resources may be divided into three different groups, the basic sciences (e.g. biochemistry, biology), clinical sciences (e.g. geriatrics, obstetrics and gynecology), and population medicine or community medicine (public health). While clinical medicine focuses on the health of individual patients, population medicine focuses on the health of communities. Like clinical medicine, population medicine requires specific information, special techniques, and skills in order to study the distribution and determinants of health and disease in the population (Mausner and Kramer, 1985). An epidemiologist is to the population what a physician is to the patient. Epidemiologists study and report on the health of communities, populations, and nations. The physician can ask the patient specific questions about his or her symptoms and may record this and any other pertinent information into a medical record. The epidemiologist, on the other hand, depends on the retrieval of information from numerous sources including medical records, vital statistics, death certificates, disease registries, health surveys, disease surveillance, hospital discharge information and health care utilization (Last, 1987). The accuracy of the physician’s diagnosis is dependent upon
the information provided from the patient about his or her condition. On the other hand, the ability to diagnose the state of the public’s health is dependent upon the quality and accuracy of the data that the epidemiologist receives from a multitude of resources. Complete and accurate information, which is made known to public health authorities, can lead to containment of disease outbreaks; show changes in patterns of disease; provide outcomes data for health care cost and utilization analysis; can lead to changes in public policy; and lead to significant changes in disease mortality or morbidity, (Haley, 1985; Mausner and Kramer, 1985; Last, 1987; Emmerson, 1995; Musser, 1996).

As with any scientific research, epidemiologic studies are subject to error. Both random and systematic error may lead an investigator to false conclusions. Instrumental error, individual variation, observer bias, selection bias, response bias, and confounding are all potential problems of measurement and classification in epidemiology (Last, 1987; Mausner and Kramer, 1985; Dawson-Saunders and Trapp, 1994). One specific area where researchers have a great potential to control for error is during the data collection or data processing step. Although there are no standardized techniques for this process, large quantities of data are usually collected on survey forms, coded, entered into a computer, and then potentially checked for error. The chance for error during this process may occur in many places (Last, 1987). What may seem as a simple and small coding error may, in fact, have serious ramifications. In one instance, researchers at Princeton University discovered startling figures related to the marital status of teenagers in the 1950 U.S. Census report. Surprisingly, the investigators discovered a significant number of widowed and divorced boys and girls under the age of fourteen years. The
researchers determined that some of the punch cards used for data entry were punched one column to the right, thus, greatly exaggerating numbers in certain rare categories (Coale and Stephan, 1962). The ramifications of such mistakes may be unknown, yet the procedures for error detection and correction are complicated and timely. Once an error has been detected several months may have elapsed, making it difficult or impossible to retrieve the original information (Kronmal et al., 1978). The Census Bureau estimated that the cost of entering an incorrect item is 200 times that of the entry of a correct item and was estimated to be about 30 cents in 1969 (O'Reagan, 1969).

Consequently, many clinical researchers employ methods such as duplicate data entry to check and minimize data entry error (Prud'homme, 1989; Reynolds-Haertle, 1992). This process, which involves duplicate keying of data, is very timely and expensive, especially when working with large quantities of data. In many public health situations, the costs associated with duplicate data entry are not feasible. For example, in New England alone, 69% of local health departments experienced budget declines prior to 1994 (Gerzoff et al., 1996). Yet, data analysis, interpretation of results, and generalization to the public are largely dependent on the quality and timeliness of the information obtained. Consequently, data quality and error prevention should be paramount concerns (Emmerson, 1995). In the last few years, however, both the private and public sectors have begun to utilize new technologies and resources for data entry.

Automated data capture (ADC) is a technological advance which provides direct entry of information into a computer database utilizing technology such as scanners, digital notepads, bar-coding, electronic patient diaries, and voice recognition. With
ADC, a single entry of information results in the capture of a data stream or image; the scanning of one bar code or one page of information can be instantly read and automatically saved to a computerized data file. ADC is appealing to researchers and clinicians for several reasons: ADC allows immediate access to data, replaces data entry operators, reduces labor costs, and implies high quality of data (Bish, 1996; Jilovec, 1996; Kasten, 1992). However, maximum accuracy and efficiency of ADC have yet to be determined. Over the last five years the computer trade literature and the media have had very enthusiastic reports about the use of such systems, i.e. voice technology. A television commercial states that one just needs to “talk and it (the computer) types”. Unfortunately, none of these reports provides definitive data concerning the accuracy, performance, and cost of the systems. Additionally, it becomes very difficult to generalize the commercial experience to that of scientific research (Kronmal et al, 1978). The present study was designed to determine the accuracy, optimal settings for enhancing accuracy, and the feasibility of utilizing an ADC software program (Teleform, Cardiff Software, INC.; San Marcos, CA) in the public health field.
Statement of the Problem

The primary purpose of this study was to determine the accuracy, the optimal conditions and software settings for the computer program named Teleform. A secondary purpose was to determine the feasibility of utilizing this technology in public health research. The specific hypotheses to be tested included:

1. Teleform software provides a very accurate methodology for data entry and thus data checking is not necessary.
2. There are no differences in accuracy due to the kind of writing instrument (pencil versus pen) used to complete forms.
3. There are no differences in accuracy due to the type of person filling in forms, trained (researcher) versus non-trained (non-researcher).
4. Teleform software is a viable alternative to data entry in the public health field.
Chapter 2

REVIEW OF LITERATURE

Data management and Research in Public Health

To date, there has been little research related to data collection methods used in public health. Research studies which test data management techniques have mainly been completed within the clinical trial setting (Kronmall R.A. et al., 1978; Prod'homme G.J. et al., 1989; Reynolds-Haertle R.A. et al., 1992; McFadden E.T., 1995). Yet, government agencies within the U.S. have been collecting vital statistics data in a structured way for more than 70 years (Feinleib, 1993a). The current data-related focus within the field of public health seems to be associated with one of several factors: 1) the assimilation and accessibility of existing public health data, 2) mechanisms to assess and document needs for health and health services, 3) collection of adequate and appropriate data for evaluating the nation’s health, and 4) development of data standards (Susser, 1993 and Feinleib, 1993a).

Assimilation and Accessibility of Public Health Data

Within the U.S. there is a multitude of health related data in a variety of different places and in a variety of different formats. According to vice president, Al Gore,

“We have generated more data, statistics, words, formulas, images, documents and declarations than we can possibly absorb. And rather
than create new ways to understand and assimilate the information we already have, we simply create more, and at an increasingly rapid pace...

Perhaps this sort of data should be called “exformation” instead of information, since it exists completely outside the brain... Indeed, by generating raw data in much larger quantities than ever before, we have begun to interfere with the process by which information eventually becomes knowledge (Gore, 1992).”

Yet, many researchers and health officials feel that much more data are needed to understand and measure health issues. Feinleib (1993a) states that, “before we sign a blank check for data collection activities, we should consider what is needed and desired from such information to ensure that it does lead to knowledge and eventually, to effective programs and better health (Feinleib, 1993a).” In addition to utilizing existing health data, there is a strong need to make existing data easy to access, utilize, and understand. In a report entitled, The Future of Public Health, the authors state that public health agencies need to “make available information on the health of the community, including statistics on health status, community health needs, and epidemiologic and other studies of health problems (Institute of Medicine, 1988).”

To date, several advances have occurred with the assimilation and accessibility of public health data. The Centers for Disease Control and Prevention (CDC) designed and implemented the program entitled CDC WONDER (Wide-ranging Online Data for Epidemiologic Research). CDC WONDER became available for use outside of the CDC in August of 1991. This online information system (http://wonder.cdc.gov) was
developed “to make it fast and easy for public health professionals to access information from a wide variety of sources.” These sources include surveys, surveillance systems, specialized studies, Morbidity and Mortality Weekly Report (MMWR), descriptions of state and local health department activities, and numerous public health databases (Friede et al, 1993). In addition to WONDER, the Epidemiology Program Office at the CDC has developed a data collection and analysis program call EpiInfo. The program provides a tool for epidemiologists to efficiently collect data pertaining to disease surveillance (Feinleib, 1993a). The use of a standard program, like EpiInfo, may also provide the means for aggregating local surveillance data at the national level.

Another such initiative has been undertaken by the National Center for Health Statistics (NCHS). The mission of NCHS is to provide statistical information that will guide actions and policies to improve the health of the American people and to “lead the way with accurate, relevant, and timely data.” NCHS has designed and implemented a number of surveys which assess many aspects of the health status of the U.S. population. All of the data are available for public use and have been used as a means to monitor the progress of achieving national health objectives (Feinleib, 1993b). An online data warehouse provides statistical tables, charts, graphs, reports, and downloadable data sets (http://www.cdc.gov/nchswww/). Additionally, NCHS has developed a system called SETS whereby public health researchers can obtain large data files that can be used on local personal computers with CD-ROM drives (Feinleib, 1993b).

To date, there are many data archives which include data on morbidity, mortality, health care, and personal health behaviors. Yet, it becomes difficult to know
exactly what data are available, where the data are located, and how to access the data. For example, one less publicized archive is the National Archive of Computerized Data on Aging (NACDA). NACDA, funded by the National Institute on Aging, is basically a large data warehouse. Data sets specifically related to gerontological research can be acquired and preserved. The organization can also be accessed on the Internet. NACDA has produced a reference book that lists all data available for scientific research (http:\www.icpsr.umich.edu\NACDA).

Assessing Needs for Health and Health Services

Perhaps one of the key areas where there are ever increasing needs for access and assimilation of existing data is within the areas of health and health services research. Policymakers and administrators want access to information on a variety of health issues including: the availability and use of health services, the costs and quality of services, patient outcomes, the health status of populations and subpopulations, levels of health in different regions, and access to care for high risk populations (Feinleib, 1993a; Roos et al, 1996; Musser, 1996). Although many data sets exist, there is no standardized way to collect or access such data. Individual HMO's and hospitals may collect acute information but this data has yet to be shared or made common. One such way to describe our health care data efforts might be to use the terms “disparate data” (Coffey et al, 1997). Brackett (1994) used this terminology to describe the state of the data coordination efforts by corporations, businesses, and other organizations. He states that, “disparate data are data that are fragmented across a variety of files,
redundant and inconsistent, poorly named and defined, poorly structured, and not well documented and understood. Their meaning, content, and format are highly variable, and they have low integrity and unknown accuracy. They are in different locations, on different databases, with different structures, and can be even stored redundantly (Brackett, 1994).” Within the medical arena, it is most likely that enormous amounts of data have been collected and that individuals may not even know what data exist, where it exists, or even what the data mean. Yet, more and more data are captured and stored. Within the U.S. some initiatives have begun to utilize existing administrative data and to standardize the collection of health outcomes data.

In 1989, six medical group practices formed an alliance to study the feasibility of collecting standard information on patients’ health and well-being, to use the information to improve quality of health care, to produce research on effectiveness, and to determine clinical policy. The plan was originally proposed by Dr. Paul Ellwood, as a methodology for “outcomes management” research. His proposal was to create a national database from which patients could be tracked and compared. To date, the American Group Practice Association Outcomes Measurement Consortia (APGA OMC) has grown to 55 practices and includes more than 400 physicians. According to Kania and colleagues (1996), there is a high level of participation by physicians and researchers due to the fact that these members are all involved in the program design, selection of data collection instruments, protocol development, data collection, and analysis. Access to the information is provided by a confidential data release agreement. Participating researchers can obtain data on thousands of patients in the U.S. In
addition to the data collection efforts, several studies have been completed. Studies have successfully utilized the data to study patient outcomes for total hip replacement (THR), cataract surgery, asthma, and diabetes. For example, data on patients with THR determined they did not recover as quickly as anticipated. This finding led to educational and procedural changes at several sites (Ellwood, 1988 and Kania, et al. 1996).

Similarly, the Physician Payment Review Commission (United States, 1994) detailed a national data strategy to Congress. The primary components of the plan included: monitoring utilization, costs, and quality of care; establishing accountability for quality and access; support of outcomes research; profiling and measuring risk. Although private initiatives by physicians may be viewed as highly commendable, Roos and colleagues (1996) state that physician plans are not population based and that key data elements were neglected. The missing elements were seen as necessary to help refocus health policy and included socioeconomic status, health status, and health care use. “Such a national data strategy neither leads toward a focus on the health of populations nor facilitates a consideration of the link between use, expenditures, and health (Roos et al., 1996).”

Population-based data summaries have been published by the National Center for Health Statistics (NCHS, 1993). Reports were provided by state on health status, health care use, and race. Yet, there was no attempt to identify any correlation across these elements (Roos et al, 1996). The most recent report by the NCHS did examine socioeconomic differences for the three largest race and ethnic categories: non-Hispanic white persons, non-Hispanic black persons, and persons of Hispanic or Mexican origin.
Interestingly, data for this report were obtained from a tremendous list of governmental and non-governmental agencies and organizations; more than thirty pages were necessary to document where the data originated. Documentation from the NCHS states, “the data in this report vary considerably with respect to source, method of collection, definitions, and reference period (NCHS, 1998)”. To date, there is no national or government supported intervention for collecting population-based health and health services data.

Musser (1996) addressed the issue of whether the collection of health care data should involve government intervention or private cooperation. She describes the division between those who opt for a public or governmental approach and those who vie for a competitive market approach. Despite a decision against national health care reform, the Federal government will most likely continue to have its hand in the health care industry, mainly to promote and foster competitive market environments. Musser states that the government will still need information to promote efficiency, regulate competition, and make purchase decisions. Individuals and corporations need data to make prudent decisions about how to ascertain the highest quality of care. “Much of the health care data collected now focuses on single, episodic contacts with the health care system. For data to be effective we must move from tracking individual providers to tracking systems of care.” Musser makes a case in her paper for adopting a system similar to the Canadian population health information system (POPULIS) (Musser, 1996).

POPULIS was developed and implemented in the province of Manitoba, Canada. The system was designed to help the Canadian public understand that more health care is
not necessarily better and to help policy makers have the means to combine population health concerns with cost containment. The system allows for comparisons of the health characteristics of regional populations and how these populations use the health care system. POPULIS builds on administrative data generated while payment is made to hospitals, nursing homes, and physicians. POPULIS describes supply, access to care, intensity of use, differential use across areas while associating indicators of socioeconomic risk and health status. POPULIS also allows for detailed information on type of services (hospital, nursing home, home care, and physician), location (i.e. small rural hospitals versus larger health centers), and costs. The authors provide several examples as to how the system has been successfully utilized. For example, four years of data were analyzed to assess whether the closure of hospital beds adversely affected access to services, the quality of care delivered, or the health of the population. The researchers found no decrease in the number of patients treated, no increase in adverse events due to early discharge, and no adverse impact on the health of the population (Roos et al., 1996).

The Collection of Adequate and Appropriate Data

Although the U.S. government does not collect information as efficiently and adequately as the POPULIS system there is a strong need to standardize what data are collected. Perhaps a mandate that required HMO's to collect data on socioeconomic status might provide a first step in integrating existing data sources. Susser (1993) attempts to address the idea of “health as a human right” from an epidemiologist’s perspective. Within his paper, Susser lists four constituents of what he deems an
“equitably distributed health right”. In addition to items such as equal access and equity for all social groups, Susser lists “evaluative mechanisms” as the third component of an equitable health right. He states that “evaluative mechanisms are necessary to monitor the distribution of both states of health and specific needs for health, including services across society.” Without evaluative methods or high-quality data, important public health decisions will be made without adequate information. For instance, data from the British National Health Service (NHS) was able to determine whether equal access to health care had been attained, this being one of the major goals of NHS (Susser, 1993). Unfortunately, there is no mechanism to address this issue within the U.S.

As stated earlier, before putting new data collection strategies into place there is a need to determine exactly what data are needed and how this data can guide effective programs and better health (Feinlieb, 1993a). Feinlieb (1993a) states that there are four functions that need to be addressed when collecting health information: assessment, explanation, prediction, and evaluation. First, data are needed to assess or determine the population’s health, the distribution of diseases, the availability and use of services, and other health characteristics. This is completed via the practice of biostatistics, descriptive epidemiology, and disease surveillance. Second, data are needed to explain or aid in the understanding of the causes of disease, determinants of health, and longevity. This function includes the development of preventive measures and therapies. The fields of analytic epidemiology, clinical trials, and biomedical research address this function. Third, data are needed for prediction: to use current information to predict trends, estimate costs, and potential outcomes of proposed programs. Prediction is necessary for
program planning, policy formulation, and priority setting. And finally, information is
needed for evaluation. This significant activity includes monitoring the performance and
outcomes of programs which have been implemented (Feinleib, 1993a). In addition to
the functions of data collection, there is an additional need to collect high quality and
useful data.

Feinleib (1993a) states that although the data may come from a variety of sources,
the data all must have certain characteristics to be valuable. These characteristics include:
relevancy, coverage (including subgroups of the population), quality, acceptability,
timeliness, accessibility, and usability. The three operational aspects that are of primary
importance for this project include quality, acceptability, and timeliness. Quality
touches on the issue of, “How good must the data be to be useful?” It seems obvious that
high quality data are important, but when compromises must be made, i.e. lack of
funding, to what extent can quality be sacrificed? Feinleib states, “if they could be, we
could routinely produce high-quality information from poor quality data.” Although
this becomes a difficult question to address, individuals who collect data must look for a
sufficient compromise for quality, cost, and timeliness. Acceptability is related to issues
of data collection methods. Some of the key issues related to acceptability include: the
acceptability of the cost and design for data collection methods, the adequacy of the
respondents to provide the requested information, the acceptability of respondent
confidentiality and whether the data being collected are valid and credible. Finally,
timeliness is related to several key questions. How recent do the data have to be and
how quickly can the data be ascertained (Feinleib, 1993a)?
One critical area where the collection of adequate and appropriate data is essential is related to the measurement of the occurrence and distribution of specific diseases, i.e. the field of epidemiology. In most instances, the diseases of concern are those with the greatest actual harm or potential to cause harm (Potter and Tauxe, 1997). Public health surveillance is the organized collection of specific disease information from those who are diagnosed with the disease. Surveillance depends on a functioning medical care system that can identify specific conditions, the willingness of clinicians and laboratory scientists to report the diagnosed cases, and the resources needed to gather, verify and summarize the information (Potter and Tauxe, 1997). Some health events under surveillance include disease incidence, morbidity, and mortality, birth defects, environmental hazards, risk factors, animal reservoirs of infectious disease, and vector distribution (Declich and Carter, 1994). With adequate and timely information, action can be taken to contain or treat disease, and ultimately to reduce morbidity or mortality. There are different types of surveillance systems, one is passive and the other active. The passive system is much more dependent on voluntary reporting by clinics or laboratories while the other is much more expensive and involves active solicitation of reports of new cases (Potter and Tauxe, 1997). Declich and Carter (1994) state that “the collection of data is the most costly and difficult component of a surveillance system and that the quality of a surveillance system is only as good as the quality of the data being collected.” Susser (1993) states “the effort for equity in health must be sensitive to the dynamic nature of health and disease through time. The antecedents of disease are shadowy enemies, changing their shapes as society changes its form. To contain these enemies requires the
epidemiological capacity to measure and monitor both the performance of health services and states of health.”

Data Standards

One final area where the field of public health has focused on data is in the development of data standards. Coffey and colleagues (1997) used an example of fire hoses to describe the dilemma of standardization with health data. The great Baltimore fire of 1904 lasted for two days and destroyed more than 1500 buildings. Although there was an abundance of water and fire fighters, their efforts were impeded when few hoses fit the available hydrants. “Embarrassed that their hoses lacked standard couplings, the Bureau (National Bureau of Standards) investigated and found over 600 variations in fire hose couplings across the country.” With the aid of Federal funding, fire associations finally agreed upon hose coupling standards. Yet, 13 years later only a handful of cities in the U.S. had complied with these standards (Cochrane, 1966). Like the initial hoses and fire hydrants, there are no single standards for health data. In August of 1996 the Health Insurance Portability and Accountability Act (HIPAA) was enacted and included groundbreaking provisions for the development of a national health information system through the establishment of standards (Coffey et al., 1997).

HIPAA, best known for its health insurance reform, guarantees portability of health insurance between jobs and restricts denials of coverage based on preexisting conditions. Less known are the provisions for “administrative simplification”, to improve the “efficiency and effectiveness of the health care system, by encouraging the
development of a health information system through the establishment of standards and requirements for the electronic transmission of certain health information (HIPAA, 1996).” The secretary of the U.S. Department of Health and Human Services was directed “to adopt standards for financial and administrative transactions and associated data elements, including code sets for clinical nomenclature.” Examples of the transactions to be standardized included health care claims, enrollment, disenrollment, payments, and health claim status. Additionally, to promote information sharing, the secretary was also directed to adopt standards for a “unique health identifier for each individual, employer, health plan, and health care provider.” Also, for security purposes, “to protect the integrity and confidentiality of information and to protect against unauthorized uses and disclosures (HIPAA, 1996).”

The development of standards has been seen as a way to reduce the administrative costs of health care reimbursement, improve competition through information disclosure, enhance performance measurement, and improve the quality of care. Additionally, the development of standards may also lead to the development of a national health information system. This system may be “inhibited by the lack of standards for defining data elements, coding data, defining data file structures, and exchanging data electronically (Coffey et al., 1997).” Some standards do exist and numerous groups have been involved in the development of more consistent health care data. These groups include the National Committee on Vital and Health Statistics, the National Uniform Billing Committee, and the American National Standards Institute Committee (ANSI). Yet, these standards often had “different rules for different
organizations and were changed often to meet local needs. There has been no single standard electronically (Coffey et al., 1997).

Coffey and colleagues (1997) completed a study to compare the content of 10 state data organizations and two statewide hospital associations. The twelve statewide data systems were said to be atypical. They represented some of the most advanced and largest inpatient data systems in the country. The authors found that key information on gender, ZIP code, race, and ethnicity were recorded in various ways by statewide systems. For example, a rather straightforward and simple element, like gender, included five different coding schemes across the 12 states. The 12 states also varied considerably in their collection of race and ethnicity data. Nine states collected race data while the states of Arizona, Illinois, and Washington did not. Interestingly, only five states collected data on Hispanic ethnicity despite the rapid growth of the Hispanic population in the U.S. The authors state that clearly a standard needs to be adopted since “failing to collect such information leaves us unable to detect inter-group differences in treatments or incidence of disease or to identify strategies to care for underserved populations (Coffey et al, 1997).”

As a response to HIPAA, the Centers for Disease Control and Prevention and the Agency for Toxic Substances and Disease Registry (CDC/ATSDR) have begun to develop a proposal for data standardization for use in health information and surveillance systems. CDC and ATSDR are anticipating the adoption of a national HIPAA standard and as such, are working to ensure that the standards meet public health needs. In their draft proposal entitled, Common Data Elements (CDE) Implementation Guide
www.cdc.gov/data/index.htm), the authors include standards for variable selection, standardization of numeric fields, and an “approved” list of common data elements. The approved data elements studied and approved include date, country, age, sex, region, name, and address. Issues, which are still under discussion, include missing values, marital status, ethnicity, race, and occupation (U.S. Dept of Health and Human Services, 1999). In summary, there are no current data standards, yet it seems likely that the U.S. government will adopt a health care data standard in concordance with HIPAA. Coffey and colleagues (1997) make suggestions in their paper for the successful adoption of standards and proven ways to benefit from such standards. The following statement made by Coffey and colleagues describes the potential dilemma for adopting such standards, “the story of the fire hoses teaches us that standards alone do not ensure their universal adoption” (Coffey et al., 1997).” Despite the needs for health data standards, health information has been collected for numerous years utilizing many different methodologies.

Methodologies for Data Collection

Data Transcription and Acquisition

In many research studies, both in epidemiology and clinical medicine, researchers may follow an individualized protocol for the collection and recording of data. Typically, data are recorded on a paper collection form, but sometimes data can be directly input into an electronic record. In most instances, researchers key data directly into a computer file using the hard copy of the data and a computer software program
If not done previously, coding or classifying the study data must also take place. This coding is done for the purposes of translating the data into an appropriate format for statistical analysis. There are several places in the data transcription and acquisition steps which are predisposed to weakness and error. First, research assistants, interviewers, or the subjects themselves must complete the information correctly on a form to prevent data transcription errors. Errors may occur when, for example, individuals do not record the correct response to a question or when a question is inadvertently skipped. A second shortcoming occurs when there is a time lag between the review of paper forms and actual data entry. In other words, data entry does not occur until months later after the initial data collection has occurred. When and if errors are located, corrections can be rather difficult and often impossible to complete as study subjects or research assistants are no longer available. A third drawback relates to the limited amount of data that data entry workers can key. The Data Entry Management Association (1990) estimated that an experienced operator can key 12 megabytes of data per year at 8000 strokes per hour while an inexperienced operator would perform at a much lower capacity (as cited in McFadden et al, 1995). The timeliness of keyed data then becomes a crucial factor and is related to the individual who keys the data, the number of forms to process, and the density of the data being collected. In response to the aforementioned transcription and acquisition issues, researchers have developed and tested several different methodologies with the hopes of improving the accuracy, efficiency, and timeliness of data acquisition. To date, most of this research has been completed in clinical research settings. Although data collection in
the public health field is just as crucial as in the clinical setting, lack of resources makes such research difficult (Gerzoff et al., 1996). Yet results from many of these studies can be easily generalized to the public health field.

**Computer Assisted Data Collection**

Computer assisted data collection (CADC) has improved the process of data entry by removing the paper-recording phase of data collection. Since the late 1970’s, CADC has mostly been applied to telephone surveys (Harlow, 1985). CADC involves entering data directly into a computer using pre-designed computer screens as subjects or patients are interviewed. The major benefit of this methodology is that one opportunity for error is eliminated as the transcription step is removed. Data collection and data entry are merged into one step (McFadden et al., 1995). Additionally, less paper handling is necessary, storage requirements of forms are lessened, and there is a shorter time frame in which to ascertain errors. The CADC system can also be set up to automate skip rules, to enforce the completion of required data fields, and to perform rapid calculations (Christiansen et. al, 1990). Concerns of using this type of methodology include: the possibility that participants or research assistants will dislike the system or that data collection will become more difficult or time-consuming, the elimination of hard copy source documents for backup and auditing purposes, the potential for less flexibility when revisions are made to a questionnaire, and the additional funding necessary for computer hardware, software, and programming (Christianson, 1990; McFadden, 1995).
Christiansen and colleagues (1990) completed a pilot study (five staff members and 16 volunteers) to determine the reaction of study participants to a CADC system, to ascertain whether staff and participants preferred CADC over a paper system, and to compare CADC with a paper system for quality and time requirements. The results from the study indicated 1) that all staff members had a preference for CADC over paper collection, 2) that the staff cited faster and more accurate data entry, and 3) that they were less likely to enter a field incorrectly or skip an item. None of the participants in the study had any problems with CADC and most had no preference for using either system. The time required for data collection was similar, with a total of 28 minutes for the paper forms (this did not include time for keying data) and 31.5 minutes for CADC. The paper forms required an additional 5-7 minutes to key the data from the forms, which was 8% longer than CADC. There were also some differences with respect to "suspicious" values. Twenty-three of 861 (2.7%) values collected on paper were suspicious as compared to 2.0% for the CADC. Yet, the majority of the suspicious values (21 of 25) were resolved at the time of collection compared with only one of 23 from the paper system. When data were collected with both methods for the same participant, 11 of the 27 errors were due to collection errors in completing the paper forms, the other 16 discrepancies could not be determined since the participants had left prior to keying the data. Thus, "the CADC system yielded relatively clean data and reduced the need for further error resolution." The costs of CADC were estimated to be $20,000 in increased personnel cost for development and $36,000 for additional workstations. This was compared to the reduction of data entry staff needed for the
study, an estimated cost of $400,000 for their six-year study. The authors state that the major problem with CADC involved time delays due to the programming time needed to modify the data entry screens (Christiansen et al, 1990).

**Double Data Entry**

Duplicate data entry (DE) is a process where data from one form is keyed into the computer twice. Two separate computer files are created, usually by two different operators, and then compared and crosschecked for data entry error. DE has been utilized for data entry since the late 1970's as computers became more accessible for data entry. Probably the most accepted methodology for data entry in the clinical trial setting has been DE (Reynolds-Haertle et al., 1992; Bagniewska et al., 1986; DuChene et al., 1986). DE has also been utilized by the World Health Organization (Gibson et al., 1994). DE reportedly yields low error rates, yet involves much time. The introduction of other methods like CADC, has created a debate as to the continued need and value of utilizing DE. Because of the cost of DE, particularly when used in large multi-center clinical trials, researchers began to study its value. For example, what amount of error should be expected when using DE and does DE significantly lower data entry error as when compared to single data entry (SE), see Table One (next page).
Table 1. Summary of Duplicate Data Entry Methods, Ordered by Year

<table>
<thead>
<tr>
<th>Study</th>
<th>Authors, Year</th>
<th>Type of Data Checking</th>
<th>Number of forms or fields</th>
<th>Errors per 10,000 fields</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coronary Artery Surgery Trial (CASS)</td>
<td>Kronmal R.A. et al., 1978</td>
<td>DE, paper keypunch system</td>
<td>51,000 fields</td>
<td>5.9</td>
<td>*substantial cost savings close to $200,000 using an electronic system with DE</td>
</tr>
<tr>
<td>Systolic Hypertension in the Elderly Program (SHEP) Pilot Study</td>
<td>Bagniewska A. et al, 1986</td>
<td>SE with range and consistency checks</td>
<td>19,285 fields</td>
<td>20.0</td>
<td>*accuracy can be enhanced by computerized editing procedures</td>
</tr>
<tr>
<td>Multiple Risk Factor Intervention Trial (MRFIT)</td>
<td>DuChene A.G., et al. 1986</td>
<td>DE</td>
<td>497 forms, 43,299 fields</td>
<td>6</td>
<td>*error rates varied by time of entry (initial vs subsequent entry) and by type of form *fields were fairly short, an average of three characters in length.</td>
</tr>
<tr>
<td>Hypertension Prevention Trial (HPT)</td>
<td>Prud'homme G.J et al., 1989</td>
<td>DE</td>
<td>57,000 forms</td>
<td>69 initial 5 final</td>
<td>*initial errors detected and corrected using DE. *5 errors found after final audit</td>
</tr>
<tr>
<td>Multiple Risk Factor Intervention Trial (MRFIT) -continued next page</td>
<td>Neaton J.D. et al., 1990</td>
<td>DE with extensive data checking SE with extensive error checking SE with minimal error checking</td>
<td>3,496 forms, 435,212 fields 264 forms, 64,416 fields 346,104 forms, 3,114,936 fields</td>
<td>3.5 9.5 26.1</td>
<td>*Without verification expect 4 to 5 times more fields in error. *Higher rates appear for alphabetic fields and for long fields like Soc. Sec. Num. *Extensive logic checking may provide error rates similar to DE.</td>
</tr>
</tbody>
</table>

DE = Duplicate Data Entry, SE = Single Data Entry
Table 1. Continued, Summary of Duplicate Data Entry Methods, Ordered by Year

<table>
<thead>
<tr>
<th>Study</th>
<th>Authors, Year</th>
<th>Type of Data Checking</th>
<th>Number of forms or fields</th>
<th>Errors per 10,000 fields</th>
<th>Conclusions</th>
<th></th>
</tr>
</thead>
</table>
| Treatment of Mild Hypertension Study (TOMHS) | Neaton J.D. et al., 1990          | SE with extensive error checking | 602 forms, 143,213 fields | 24.6                     | -continued from previous page *
|                                            |                                    |                       |                           |                          | Selective verification of fields may be an efficient error checking method |  |
| Cardiac Arrhythmia Suppression Trial (CAST) | Reynolds-Haertle R.A and R. McBride, 1992 | DE vs SE             | 821 forms, 42,278 fields | 15                       | *32% reduction in error by DE *
|                                            |                                    |                       |                           |                          | DE significantly lowered error rates (p=0.001) |  |
|                                            |                                    |                       |                           |                          | DE took 37% longer than SE |  |
|                                            |                                    |                       |                           |                          | an error rate of 10 errors per 10,000 was set as an attainable goal |  |
| Continuous Hyperfractionated Accelerated Radiotherapy (CHART) | Gibson D.G. et al, 1994 | DE vs SE             | 448 forms, 16,277 fields | 13                       | *Use trained data managers for data entry, who exercised great care and accuracy. *
|                                            |                                    |                       |                           |                          | Use well constructed consistency checks without DE. |  |
|                                            |                                    |                       |                           |                          | DE is not necessary |  |
In summary, several studies have assessed the effectiveness and timeliness of duplicate data entry procedures while comparing DE to different data entry methods. Single data entry procedures were compared to duplicate data entry systems. Additionally, procedures involving data consistency checks and other extensive error checks were also compared. Error rates for DE ranged from 3.5 to 15 errors per 10,000 fields (Kronmal et al., 1978; DuChene, et al., 1986; Prud'homme et al., 1989; Neaton et al, 1990, Reynolds-Haertle et al., 1992; Gibson et al., 1994). SE yielded higher error rates, ranging from 9.5 to 24.6 errors per 10,000 fields (Bagniewska et al., 1986; Neaton et al., 1990; Reynolds-Haertle et al., 1992; Gibson et al., 1994). Although DE yielded lower error rates, DE was associated with higher costs and greater time involvement (37% longer than SE) (Reynolds-Haertle et al., 1992). Interestingly, lower error rates could be obtained with SE by using extensive data checking, consistency checking, and trained data managers (Neaton et al., 1990; Gibson et al., 1994).

Many of the aforementioned researchers also discussed the types of fields being checked. For example, shorter fields had less error while longer and alphabetic fields had greater error. This can be attributed to the fact that alphabetic fields usually have more characters than numeric fields and that the response set for alphabetic fields is larger (26 characters as compared to 10 characters) (DuChene et al., 1986; Neaton et al., 1990). For example, a field like social security number had a higher error rate of 117.7 per 10,000 as compared to a field like exam year, 1.7 per 10,000 fields (Neaton et al., 1990). Ultimately, error rates will vary depending on the methodology used, however, an error rate of 10 or fewer errors per 10,000 fields can be a realistic goal (Reynold-Heartle et al.,
In addition to the type of data being collected, Hosking (1995) states that as the volume of data increases, it becomes more difficult to ensure accurate and quality data. Neaton and colleagues (1990) suggest the use of optical scanners as an efficient system for working with large amounts of data and converting data to a "machine-readable format". Scanners are now readily available and can easily be attached to personal computers. "Such devices have the potential both for reducing error and for saving time (Neaton et al., 1990)."

**Automated Data Capture**

Automated data capture (ADC) is a general term used to describe a variety of advanced data collection technologies. By utilizing one of several techniques, this technology allows for the direct acquisition of data into a computer and thus, bypasses the data-keying process. These advanced technologies include scanners, digital notepads, bar-coders, swipe/smart cards, electronic patient diaries, radio signals, and voice recognition. For example, the scanning of one bar code can call up an entire patient record (Kasten et al., 1992). ADC is very appealing since it seems to promise immediate access to data, a decrease in data entry personnel, a reduction in labor costs, the potential to collect more data, and the promise of high quality data (Bish, 1996; Jilovec, 1996; Kasten et al., 1992). ADC is readily and successfully used in grocery stores, department stores, banks, the transportation industry, hospitals, libraries, and shipping industries. More recently, researchers have begun to assess the use of this technology in the health field. In addition to the prospect for timely and accurate data, health-related projects are
growing in size. Projects produced by Federal government agencies, multicenter field trials, clinical research centers, and multinational investigations, have become much larger in scale and consequently, include the collection of enormous and extensive data sets (Arndt et al., 1994). Arndt and colleagues (1994) state that “the sheer size of data sets can influence data quality.” It seems reasonable that a technology like ADC could greatly improve the acquisition of high-quality and high-quantity data, ultimately providing the means to guide timely public health programs and decisions (Susser, 1993). To date, several types of ADC have been used and tested in the medical field. The understanding, accuracy, and use of these techniques will be explored.

**Bar Coding**

ADC may appear to be a new and upcoming technology, yet ADC has been around for several decades. Bar coding is a form of ADC and has been one of the greatest alternatives to data entry (Kasten, 1992). Bar coding can be described as the placement of a code or pattern of bars and spaces placed upon a package or label. Bar codes can be viewed on the back of cereal boxes, express mail packages, coupons, or magazines. The bars and spaces represent “one of several common symbologies or languages that enable codes to speak directly to computers through special scanners.” Each symbology is made up of specific codes or “identifiers that use various bar and space configurations to represent numbers, letters, and other symbols (Hakanson, 1986).” These codes are read by a special scanner, which uses an intense light to read the code. The code is then converted to an electrical signal which is decoded and transmitted as digital information to a computer (Hakanson, 1986). Use of bar coding dates as far back as the 1960's when
the railroad industry used bar coding to track the location of railroad cars in North America. Additionally, supermarkets utilized bar coding for automatic inventory control and checkout processing. It wasn’t until 1973 that the industry aborted some unsuccessful attempts at bar coding and chose a new standardized code, the Universal Product Code (UPC) for use in North American stores (Kasten, 1992). In industry and manufacturing, bar coding is continually used, i.e. goods are packed, invoiced, weighed and labeled; information is instantly captured; and inventory levels and accounts receivable systems are automatically updated (Jilovec, 1996). Interestingly, where data integrity is paramount, i.e. financial institutions, bar coding has been successfully utilized and accepted for use with electronic commerce and financial data (Kellock, 1994). Even so, supermarket consumers may feel differently. Many consumers reported high error rates at store checkout scanners. More than half of them, nearly 22,000 responses, noted occasional discrepancies between prices marked on shelves and the prices that were scanned (Consumer Reports, 1997).

Rappoport (1984, 1985) discusses the use of this technology in the medical setting. The title of one of his papers, “If bar code works in supermarkets, it should be great for medicine”, gives recognition to the fact that ADC can aid data collection efforts in the health field. In 1983 the Health Industry Bar Code Council was formed. One of the council’s missions was to examine and evaluate all available bar code methods, to compare them with other ADC techniques, and to establish a standard for the industry (Rappaport, 1985). A specific standard, called Code 39, was accepted for use within the entire health care field because of its flexibility, variable length, and ability to read both
alpha and numeric data. In 1985, Rappaport emphasized that patient-generated materials needed to be included in machine readable identification systems. These materials included forms, records, laboratory and blood specimens, x-rays, EKG tracings, drug labels, and central supplies. Consequently, bar coding has been successfully utilized in hospitals and medical laboratories.

For example, clinical laboratories can now positively and permanently identify specimens in a timely manner. “The ideal specimen has an unmistakable ID from the moment of acquisition to receipt of the final report (Kasten, 1992).” Bar coding has also been used to identify conference attendees, surgical instruments, and insurance forms (Jilovec, 1996). Since its inception bar coding has gained in popularity and many positive aspects have been described. Bar coding is: flexible, almost any object can be marked and scanned inexpensively; dependable, scanning can occur even if the bar code is partially gone; rugged, bar coding can be successfully used in hostile environments; and extremely accurate, one substitution error in three million characters (Hakanson, 1986). Additionally, the technology can make life easier for lab staff; many hospitals have found that bar coding actually boosts employee morale; laboratories derive satisfaction from knowing they are working with leading-edge technology; and shrunken budgets have forced laboratories to use “ingenuity” in their quest for efficiency and productivity (Kasten et al., 1992). However, Kasten and colleagues (1992) state that the process of utilizing bar codes in the medical setting was not without problems. The problems included: not placing the codes on at the time of acquisition, thus posing the potential for identification error; space constraints on small specimen tubes; and poor quality printers
to print the bar code labels. Since then, many of these problems have been addressed and the current technology has worked out many of these problems (Kasten, 1992). Consequently, several studies have assessed the accuracy of using bar code technology in the medical field (Lerou et al., 1988; Chau et al., 1993).

One such study assessed the accuracy of bar codes as compared to handwriting during videotaped trauma resuscitations. The investigators developed a bar code system to provide time-stamped entry of patient demographic information, vital signs, procedures performed, laboratory and radiology orders, and medications given during resuscitation. A 24 by 36-inch bar code template was prepared containing bar code elements for all the aforementioned items. Each label in the system corresponded to a particular clinical event or data point. The system consisted of a bar code scanner, bar code labels, a bar code printer, a personal computer and a software decoder program. Data were entered by sweeping the scanner over the bar code labels. The scanner then converted the dark bars and light spaces to an analog signal, which was then digitized and converted to an ASCII code. For example, the entry of the heart rate “85” required an individual to sweep the scanner over three bar codes: one bar represented the words “heart rate”, one represented the number eight, and one represented the number two. The final entry read “HR: 82/minute”.

Four videotapes of trauma resuscitations, occurring in the emergency room, were selected for review by one investigator. A time-coded master list of all events was prepared by the investigator and used as a gold standard. Twenty-four emergency nurses, without prior experience, were allowed to familiarize themselves with the bar code
system minutes before watching the videotapes. Each nurse viewed the videotapes during a single uninterrupted session. The nurses recorded two cases by handwritten entry and two cases by bar code entry. The order of the viewing and recording methods were randomized. Forty-eight hand written records and 48 bar-coded records were generated. The handwritten and bar-coded records were compared with the master list of events and the number of errors counted. The mean number of errors per record for bar code was 2.63 compared with 4.48 for handwriting \((p < 0.001)\). The total number of omission-type errors was 108 for bar-coded records compared with 175 for handwritten records. The total number of inaccuracy-type errors was 18 for bar coding versus 40 for handwritten records. Nursing experience, 5 months to 19 years, had no significant effect on total number of errors. Thus, the authors concluded, that bar-coded data entry resulted in significantly fewer errors compared with conventional handwriting (Chau et al., 1993). Similar findings have been ascertained in pharmacy departments (Scott et al., 1996; Kanmaz et al., 1997). However, Kanmaz was able to document the cost per pharmacist for a manual system, $414.84, versus a bar code system, $450.19. Although, the costs per system were not statistically significant, the bar code system was more accurate with an error rate of 1.7% as compared to 5.8% for the manual entry (Kanmaz et al., 1997). Although bar coding technology has been available for several decades, it has taken a number of years to successfully implement this technology in the medical field. Likewise, the study and use of other ADC techniques have been very minimal and limited.
Automated Forms Processing and Scanning Technology

Another type of ADC involves the computerized scanning of images, forms, or pictures. This type of ADC has been described in a number of different ways including optical character recognition (OCR), optical mark scanning technology, discrete optical marking, document image processing, optical scanning, image character recognition, and intelligent character recognition (Denwood, 1996; Shiffman et al., 1997; Smyth et al., 1997; Puskar et al., 1996; Hammer et al., 1993; Titlestad G., 1995; Schumaker et al., 1998). This process first involves the actual capture or snapshot of images, forms, or marks utilizing a computer scanner and scanning software. The second step in the process involves the actual recognition of scanned images and the conversion of the information into a readable and usable format. Interestingly, OCR can be traced back to an individual named David Shephard, who is acclaimed by his peers as the “father” of OCR. His initial production of OCR was adapted and used by Reader’s Digest back in 1954 to read data from member subscriptions. This system, named GIZMO, is currently located at the Smithsonian Institute in Washington D.C. (Schantz, 1996).

Use of scanning technology to collect data has been achieved in two discernible ways. First, with the use of discrete optical marking (DOM) or optical mark technology, individuals are asked to simply fill in “bubbles” or marks as choice responses to questions on a specialized form. An optical scanner then reviews the completed forms “using a discrete set of read components to determine mark density at specific response positions on a document (Denwood, 1996).” The scanner then uses various read levels to distinguish light marks, dark marks, smudges, and erasures on the completed document.
Because of this, the scanner can make very accurate readings even if the document is filled out poorly. DOM is viewed as very accurate, efficient, cost effective, and is most effective with categorical or numerical data. DOM is used widely at many colleges and universities for standardized examinations and course evaluations (Davidson et al., 1996). The major disadvantages are: 1) the possibility of a time lag before scanning, as forms may need to be visually checked for stray marks that could adversely affect the accuracy of the scanner, 2) the limited types of data that can be collected and scanned, and 3) the need to use preprinted compatible forms. The major advantage of DOM is the ability to handle massive amounts of data relatively quickly (Davidson et al., 1996). According to Denwood (1996) “where there is relatively high volume and requirement for high accuracy without human intervention, there is no faster, more accurate, and cost effective way to collect data.”

The second way to use scanning technology, in the data collection step, is to use a document or image scanner. The image scanner uses a camera type device to produce an image of the document in pixels. A pixel is a “picture element” or a part of the original document that coincides with a spot at a given moment. The resolution of the image is determined by the number of pixels per square inch. Image scanners can be used with software to collect data, i.e. the scanning software compares the images with known patterns and makes a judgement based upon the comparison of what character is present. The recognition can be as limited as just recognizing characters printed by machine (Optical Character Recognition) or can be as sophisticated as reading handwritten characters. The reading of hand written characters is called intelligent character
recognition (ICR) and requires much processing power. With ICR, the scanner sends the processor tremendous amounts of data in the form of pixels where complex algorithms are used to make a judgement as to what characters are present. The scanner can not distinguish mark density, i.e. the scanner can only see a mark as present or not present and will not distinguish between marks, smudges, erasures, or dirt (Denwood, 1996). Image scanning technology has recently been used in nursing, pediatric clinics, hospital surveillance, mental health screening, psychiatry, cancer registration, and physician practices (Davidson et al., 1996; Denwood, 1996; Nolan et al., 1997; Shiffman et al., 1997; Smyth et al., 1997; Puskar et al., 1996; Hammer et al., 1993; Titlestad G., 1995; Schumaker et al., 1998). Although, the aforementioned studies used scanning technologies, many studies did not assess the accuracy of such techniques (Hammer et al., 1993; Davidson et al., 1996; Puskar et al., 1996; Nolan et al., 1997; Shiffman, 1997; Schumaker et al., 1998). Several applicable studies will be reviewed.

Research was conducted at the Hines Department of Veterans Affairs (VA) in Chicago and Minnesota to examine multiple data collection systems. The system was to aid in the transition from inpatient care to ambulatory care, to provide data for the evaluation of quality of patient care, and to provide the necessary information for third-party billing. The researchers needed forms that could be easily and inexpensively designed and modified, printed on plain paper, preprinted with patient identification data, and scanned on both sides in one pass. Initially the investigators utilized a manual, clerical data entry system to determine the type of information to be collected. Denwood (1996) described the differences with many ADC technologies. The main goal
of the project was to compare the cost, accuracy, flexibility, training requirements and possibility of theft of all the data collection systems. Unfortunately, there was no description or documentation as to how the comparative study was completed. They state that "after our evaluations, we determined that OMR met our criteria and also offered the most cost effective alternative in terms of equipment purchase, relocation and replacement." The cost of the system was $8000 for the hardware equipment. Accordingly, "2000 forms per hour could be processed with 99% accuracy." Denwood states "an exciting process has begun, based on little bubbles on a piece of paper and mature technology (Denwood, 1996)."

Shiffman and colleagues (1997) developed and implemented a system for the structured collection and electronic capture of data for pediatric health at a clinic located at Yale University in Connecticut. The system called SEURAT (scanning for evaluation, utilization review, analysis, and training) was created to use paper-based, electronically scannable forms to meet several goals. The goals included the facilitation of efficient, legible, and complete documentation of patient encounters; to enhance compliance with health maintenance guidelines; to simplify quality assessment and reporting; to standardize documentation of immunizations for reporting; and to document compliance with Medicaid’s early and periodic screening, diagnosis, and treatment program; and to identify areas requiring quality improvement. The authors hypothesized that the use of structured forms would lead to more thorough documentation of patient encounters and health maintenance activities, that the effect would persist over time, and that user acceptance rates would be high. The new system was to replace paper-based records,
which were seen as illegible and incomplete of necessary information. In the past, physicians documented patient encounters on handwritten forms and relied on memory, personally prepared notes, manuals, and charts to assist in age-appropriate care for pediatric patients. A set of 13 forms was developed using the software program Teleform (Cardiff Software, San Marcos, CA) to collect all necessary information. A total of 388 structured items were identified to assess and document the health maintenance activities for the children. Teleform was used to manage document scanning, data capture, and verification activities. Although, the physicians were encouraged to write in the “white space” of the form, only structured data elements (categorical variables and bubbles) were captured by the system. Thirty-four patients were compared using the new scanning structured system (ST) versus the unstructured system (UNST). Batches of the forms were scanned within twenty-four hours and data were validated using Teleform’s verification process. Verification involved the review of forms by a specially trained staff person to view the electronic image and edit any data items “flagged by the software.” The accuracy of the system was not studied or discussed. The results of the study indicated that the physicians in the ST group documented more data elements or health maintenance activities per encounter than did those in the UNST group, a mean of 22.5 items per visit versus 10.3 items per visit (p<0.001). Additionally, thirty-four pediatricians completed a user satisfaction survey. Additionally, ninety percent of the residents expressed a preference for the ST forms over the UNST system. A follow-up was completed one year later, and the improved documentation was maintained. Thus, “the implementation of the system for structured data collection has resulted in an
increase in the number of health maintenance activities documented and a high degree of user satisfaction.” Additionally, the authors comment that “now data are available to assist with the decision making about individual patients and for aggregation for quality assurance and quality improvement activities (Shiffman et al., 1997).”

Puskar and colleagues (1996) also used Teleform in their paper entitled, “High touch meets high tech: Distance mental health screening for rural youth using Teleform.” Using Teleform, the authors designed eleven instruments to examine the health status of teenagers in rural Pennsylvania. The use of Teleform was said to address some key issues in dealing with adolescents. First, data entry with Teleform assured confidentiality, as students were told that the computer would read the questionnaires and that only summary information would be reported. Second, screening could be completed on large numbers of students who resided far from areas where service was provided. Third, Teleform could provide accurate and up to date information in a timely manner. The authors state, “with such increased communication capabilities, nurses can collect information and assess needs from a distance.” In this way, the information gathered from Teleform was used to create screening tools to reveal where “high-touch” interventions were necessary. A final key component mentioned for using Teleform was that, when nurses spend time working on computers, “they are far removed from the people they care for.” The new system was seen as a way to provide more timely interventions with “at-risk” adolescents. In this research project, 445 students volunteered for the study. Within two days of the screening, any student who scored high in regard to depressive symptomatology was identified, seen, and assessed by a
skilled nurse. The authors state that Teleform “allows for increased ease of data entry and more accurate data entry”, yet there was no assessment or information provided on accuracy rates (Puskar et al, 1996).

Davidson and colleagues (1996), have also discussed at length the use of Teleform in the nursing field. The authors provide an overview of their positive experience with Teleform and document how Teleform works. Examples on how the software is utilized within their school are provided. Although no data are presented, some key suggestions are made for using the software. For example, they state that the “time necessary to verify data depends on how well the forms are filled out.” They also find that the use of black felt pens worked better than the use of light pencils when filling in forms. Additionally, they report that “sloppily made numbers and letters require more time to correct than those that are neatly printed.” The authors suggest providing an example of how to make numbers and letters with questionnaires to overcome this problem. Finally, Teleform “is an effective application of technology that has enhanced productivity of faculty and staff (Davidson et al., 1996). A similar paper was written by Nolan and colleagues (1997) who discuss the application of scanner technology for the collection of quality data for nursing executives. The one additive comment was that Teleform allows for photocopying of scanner forms on standard white paper. The authors comment that the quality of the copy is an important consideration. Copies that are darkened by a poor quality photocopy machine may be misread by the scanner as having marks made by the respondent completing the form (Nolan et al., 1997).
A study completed by Smyth and colleagues (1997) assessed the accuracy of an optical scanning system. The authors were investigating the use of this ADC system for infection surveillance efforts in hospitals in Belfast, Ireland. The investigators stated that surveillance requires a customized database where, previously, most data had to be hand entered. Further, manual data entry is "labor intensive, tedious, slow, prone to human error, and introduces constraints on how widely surveillance can be applied." The requirements for the new system were a technology that was easy to use, adaptable, and something that would provide good value for the money. A system call Formic (Formic Ltd., London, UK) was utilized to design the surveillance system. The software is similar to Teleform in that Formic is a comprehensive questionnaire design and automated data-capture system. Their system was networked and the cost of the software was $10,500 (May, 1994). A surveillance form was created pertaining to surgical wound infection. Data from 100 completed forms were scanned and validated using the ADC system. The data from the same 100 forms were entered manually into a computer database and were then validated manually. The survey included 31 questions and sub-questions that produced a possible 59 response options. The patient identifier was the only data entered manually for the scanning system since the version of the software could not read alphanumeric characters. All response options were completed by marking an "X" in the appropriate boxes using a pen containing black ink. The questionnaires were scanned and validated by rescanning using a "validate by re-scanning" software option. The images recorded at scanning and rescanning were compared by the software and any discrepancies were flagged. The manually entered surveys were compared with the
original surveillance questionnaires by two independent observers, and discrepancies noted. Error rates were calculated by tallying errors detected at validation, dividing the number of errors by the number of response options, and expressing the figure as a percentage. The authors found that the amount of time to design the automated data entry form/database was not different from the design time of the manual system (7.58 hours for both). The total data entry time for the automated system was significantly less than the manual system (0.41 hours versus 9.09 hours). In terms of the accuracy of the automated system, there was one discrepancy error recorded as compared to 72 in the manual process (this was out of a total of 5900 response options, 59 responses per form x 100 surveys). Overall, the automated system had an accuracy rate of 99.98%, compared to the manual system 98.76%. In the final comparison with the original surveys there were still seven discrepancies in the manually entered data and none in the automated data. A final comparison was made to determine the estimated costs savings of the automated system. The costs were estimated for the processing of 10,000 surveillance questionnaires. The clerical costs for automated processing were $0.03 per questionnaire as compared to $0.66 for manual entry. The total cost of the automated system was $298.48 as compared to $6,617.52 for manual data entry with error validation and correction (costs were based on a personnel cost of $7.28 per hour). Interestingly the costs of the software and hardware were not included in the cost comparison. However, the authors state that “automated data entry is indispensable. The expected gains from instituting automated data entry are considerable savings in time, associated cost savings,
and potential savings associated with reductions in hospital acquired infection rates (Smyth et al., 1997)."

Titlestad (1995) reports accuracy rates in a short paper reviewing the use of document image processing for cancer registration. The cancer registry, located in Norway, receives 80,000 annual reports concerning new and supplementary information on cancer patients. At the time of their paper, the investigator was comparing the new scanning technology with a traditional paper-based registration system. Titlestad states that the first part of this study compared three different character readers (CGK, XDR, and Nestor). Results of the testing showed that over 90% of the numeric and 70% of the character letters could be correctly recognized by the system. No information was given as to the study design or methodology behind their testing. Although it was reported that feedback was given back to clinicians to improve the quality of the handwriting received on the forms. An XDR Network Reader was put into place for the rest of the upcoming study. The upcoming study promises to assess whether the new system produces the expected high quality data, how the new system compares to the traditional system, what percentage of clinicians would use the new application, and whether the new system would be cost effective (Titlestad, 1995).

A study completed by Jorgenson and Karlsmose (1998) provides the most informative look at automated forms processing with Teleform. The authors comment that there is a lack of empirical evidence to support the notion that Teleform saves time and money and improves data accuracy. Accordingly, "a search for the terms: automated-forms-processing or data entry or OCR in the nine most relevant databases,
revealed no previously published reports on the validation of automated forms processing for research purposes.” To validate the use of Teleform, the investigators used 401 consecutive questionnaires, randomly selected from approximately 2000 questionnaires, filled in by 195 physicians. The surveys were completed on patients with musculoskeletal disease in Denmark. The investigators utilized a newer version of Teleform (version 5.3) which automatically interpreted handwritten and machine printed text (OCR/ICR technology), shaded or checked optical mark fields, bar codes, and image zones. Their four-page questionnaire contained 29 fields, 42 variables, and 69 characters. Text and alphanumeric data were labeled as “constrained print fields” (CPF). The authors utilized an “always review” function for one CPF and a confidence threshold technique for the other nine print fields. The confidence threshold allowed for CPF to accept a character only if the confidence of recognizing a character exceeded the defined level (80% or 99%). Prior to data entry one of the investigators went through all the forms to check for obvious mistakes and to recode various information, which involved filling in six of the CPF. Data entry was completed in four different ways: 1) manual data entry by a commercial provider (single entry), 2) manual data entry by a skilled secretary performing double data entry, 3) automated forms processing using Teleform with a confidence threshold level at 80% for the CPF, and 4) automated forms processing using Teleform with the form confidence level at 99% for the CPF. The times spent on manual data entry, checking forms, preparation for scanning, feeding the scanner, and time spent verifying forms in Teleform were all recorded. Four separate data files were
created and were compared for differences among the fields in the separate files. The number and types of mistakes were recorded for each data file.

The study found different results depending on the type of data entry technique, the person filling in the forms, and the type of fields on the forms. For print fields, duplicate data entry resulted in a significantly lower error rate than all other techniques \((p<0.01)\). Duplicate data entry yielded an error rate of 2.0 characters per 10,000 characters; single data entry by commercial firm at 5.3/10,000; Teleform with 80% confidence level at 13.8/10,000; and Teleform with 99% confidence level at 9.8/10,000. The error rates for choice fields were 1.0 for duplicate data entry, 17.7 for single data entry, 2.4 for Teleform at 80%, and 4.0 for Teleform at 99%. For choice fields, the Teleform error rates were significantly lower than that for single data entry \((p<0.005)\) but comparable to duplicate data entry \((p<0.1)\). The overall error rates for all fields combined were 1.5 for duplicate data entry, 10.8 for single data entry, 8.7 for Teleform at 80%, and 7.2 for Teleform at 99%. Overall, Teleform data entry resulted in statistically higher error rates than duplicate data entry \((p<0.005)\). The lower error rate for the 99% Teleform was not statistically lower than the 80% setting \((p>0.50)\). The authors also found a difference in error rate depending on who filled in the questionnaires. If the researcher completed the questionnaire, the error rate was significantly lower than that of respondent doctors for CPF fields \((2.3 \text{ versus } 39.5 \text{ for Teleform at } 80\%, p<0.05)\). The error rate for CPF filled in by researchers was comparable to that of double data entry \((2.3 \text{ versus } 1.1, p>0.50)\).
The investigation of the time for processing the forms was compared for the four techniques. The overall time for the single entry process was 287 minutes, for duplicate data entry 375 minutes, for Teleform at 80% 142 minutes, and Teleform at 99% 164 minutes. Manual data entry took greater than 200% of the time used with automated forms processing, resulting in considerable cost differentials. The price for processing 10,000 characters was $68 for single data entry by commercial provider, $44 for duplicate data entry, $16 for Teleform at 80%, and $19 for Teleform set at 99%. Values for statistical differences were not provided for these comparisons.

In summary, the authors found the error rates for all techniques to be very low. Overall, Teleform and automated forms processing performed better than single data entry but poorer than duplicate data entry. “The main weakness of AFP (automated forms processing) was the recognition of numeric characters.” There were two explanations given for the errors. The first explanation being that Teleform misinterpreted a poorly written character and accepted it without asking for verification. The second possibility was that Teleform presented an incorrect “best guess character” that the operator accepted by mistake. Because there was a trend for fewer errors at the higher confidence setting, the authors state that the first explanation was most plausible. Additionally, there was much greater error for fields filled in by respondents than by the researcher. Although recommendations were provided to the respondents on completing the forms, the authors believed that these recommendations were not followed. The authors recommended avoiding numeric fields for research purposes, when it cannot be assured that respondents will adhere to directions. Although
Teleform reduced processing time to about one half to one third of that of manual data entry the authors felt that a very large number of forms would need to be processed in order to recover their initial investment ($5300 for computer and software, $14,000 for scanning hardware). Additionally, the authors state that considerable amounts of time and computer expertise were required to implement the automated processing (Jorgensen and Karlsmose, 1998).

Summary and Conclusions

In summary, some of the key data-related issues facing the field of public health have been reviewed. These include the assimilation and accessibility of health data, assessing the needs for health and health services data, the collection of adequate and appropriate data, and the development of data standards. Several methodologies for data collection were also described including computer-assisted data collection, double data entry, and automated data capture techniques. Available studies were presented; many of the studies assessed accuracy, costs savings, time involved, and user satisfaction. A review of the data collection techniques revealed error rates of less than 1%, fewer than 100 errors per 10,000 fields (see Table 2, page 49). The ramifications of error rates are hard to discern. One investigator states, “I think it is very easy to get over-compulsive in this kind of work... What is of interest is not the error rates themselves but whether the conclusions differ before and after correcting the errors (Neaton et al., 1990).” Arndt and colleagues (1994) addressed this same issue, and discovered 2.4% error in a large multicenter medical study. The authors state, “these errors would have affected the
data's reliability, decisions based on the study, and possibly the choice of analysis (Arndt et al., 1994).” The story of the “teenage widows” also corroborates the need for data monitoring, as numbers in rare categories were greatly exaggerated (Coale and Stephan, 1962). Likewise, the Census Bureau estimated the cost of entering an incorrect item to be 200 times that of entering a correct item (O'Reagan, 1969). Consequently, an error rate of less than 10 per 10,000 fields is possible and should be attainable (Reynold-Heartle et al., 1992; Glassman et al., 1995). Techniques like duplicate data entry and computer assisted data collection have been extensively reviewed. The use of “high-tech” methods like ADC and automated forms processing, i.e. Teleform software and scanning technology, has been less studied. Early reports on the use of Teleform state that the technology saves time and money while providing accurate data entry (Puskar et al., 1994; Davidson et al., 1996). It seems likely that this new technology could greatly benefit public health programs. Yet, only one study adequately describes accuracy rates for Teleform (Jorgensen and Karlsmose, 1998). However, this study was completed in a highly controlled environment, making its results less generalizable to a public health setting. The current study attempts to determine the accuracy and optimal conditions for Teleform in an environment similar to a public health setting.
Table 2. Summary of Accuracy Rates by Data Entry Methods

<table>
<thead>
<tr>
<th>Data Entry Method</th>
<th>Author</th>
<th>Year</th>
<th>Errors per 10,000 fields</th>
<th>Accuracy Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Assisted Data Collection</td>
<td>Christianson et al.</td>
<td>1990</td>
<td>200</td>
<td>98%</td>
</tr>
<tr>
<td>Single Data Entry, minimal checking</td>
<td>Neaton et al.</td>
<td>1990</td>
<td>26.1</td>
<td>99.7%</td>
</tr>
<tr>
<td>Single Data Entry with error checking</td>
<td>Bagniewska et al.</td>
<td>1986</td>
<td>20</td>
<td>99.8%</td>
</tr>
<tr>
<td></td>
<td>Neaton et al.</td>
<td>1990</td>
<td>9.5</td>
<td>99.91%</td>
</tr>
<tr>
<td>Duplicate Data Entry</td>
<td>Kronmal et al.</td>
<td>1978</td>
<td>5.9</td>
<td>99.94%</td>
</tr>
<tr>
<td></td>
<td>DuChene et al.</td>
<td>1986</td>
<td>6</td>
<td>99.94%</td>
</tr>
<tr>
<td></td>
<td>Prod'homme et al.</td>
<td>1989</td>
<td>5</td>
<td>99.95%</td>
</tr>
<tr>
<td></td>
<td>Reynolds-Haerlte and McBride</td>
<td>1992</td>
<td>15</td>
<td>99.85%</td>
</tr>
<tr>
<td></td>
<td>Gibson et al.</td>
<td>1994</td>
<td>13</td>
<td>99.87%</td>
</tr>
<tr>
<td>Bar Code</td>
<td>Kanmaz et al.</td>
<td>1997</td>
<td>170</td>
<td>98.3%</td>
</tr>
<tr>
<td>Optical Mark Technology</td>
<td>Denwood et al.</td>
<td>1996</td>
<td>100</td>
<td>99%</td>
</tr>
<tr>
<td>Scanning Technology, Choice responses only</td>
<td>Smyth et al.</td>
<td>1997</td>
<td>1.7</td>
<td>99.98%</td>
</tr>
<tr>
<td>Scanning Technology, Teleform</td>
<td>Jorgenson and Karlsomose</td>
<td>1998</td>
<td>7.2</td>
<td>99.93%</td>
</tr>
</tbody>
</table>
Chapter 3

METHODOLOGY

Volunteers

Fourteen volunteers were asked to take part in the study. Employees located at the University of Connecticut Health Center’s Claude Pepper Older Americans Independence Center and the Balance and Gait Enhancement Laboratory were asked to participate. Additionally, colleagues and friends from outside the university were also recruited. Volunteers included those with both data management and research experience (Researchers = R, n = 6) and those without experience (Non-researchers = NR, n = 8). There were no specific gender or age requirements.

Procedures

The volunteers were given three separate manila envelopes, labeled time one to time three. Each envelope contained five forms and one different type of writing device. The writing devices included a blue pen (Bic Round Stic; Medium Point), a black pen (Sanford Uni-Ball Onyx; Fine Point), and a pencil (Eberhard Faber American; 2 (HB)). The order of the writing devices was randomized within the numbered envelopes. The volunteers were instructed to complete the forms on three separate occasions, once per day, over a three-day period. A total of 15 forms were to be completed by each individual.
Only written instruction was provided for how to complete the forms as presented below:

Answer every question by marking the answer as indicated. If you are unsure about how to answer a question, please give the best answer you can. Please fill in one appropriate box for each question. The box should be filed in as such: ■ or □.

Teleform

Teleform (Teleform version 5.0, Cardiff Software, INC.; San Marcos, CA; $1500.00) was used to design five data collection forms and corresponding data files. The forms included three forms previously designed for use at the Claude Pepper Older Americans Independence Center and two newly designed forms (See Appendix A). Data were collected on the five forms and then scanned using a Fujitsu ScanPartner 10c scanner (Fujitsu America, Inc.; Anaheim, CA: $1500). Following scanning, a process of “verification” was completed to assure data accuracy and completeness. The forms were set up to automatically save corrected data to a Microsoft Excel file format.

Several field formats were included on each of the forms and included text fields, numeric fields, and choice fields. A Teleform enhancement, entitled “character recognition confidence threshold”, was utilized for both the text and number fields. A 95 percent confidence threshold was set to accept or reject the data. In other words, Teleform would assign a confidence rating to each character on a form. The character would be accepted and saved when the confidence rating was 95 percent or greater. If the
confidence rating was less than 95 percent, the character and field would be rejected and the field would be marked for review and held for verification. Additionally, fields that were incorrectly completed or mismarked were also held for verification. During the verification step, errors were corrected by comparing the paper form with the computerized image of the form. After verification, the data were saved to the Excel file.

**Form Description**

The five forms included a total of 81 text and numeric fields, with 451 character items, and 120 choice fields. Text and numeric fields were combined under the label of “constrained print fields” (CPF) for purposes of analysis (see Figure 1, page 54). Four CPF variables were created:

1. number of fields to verify
2. number of fields to correct during verification
3. number of field substitution errors in data after verifying
4. percent of field errors in data set.

Figure 2 (page 54), provides a description of choice fields. Similar to CPF fields, four choice field variables were created:

1. number of choice fields to verify
2. number of choice fields to correct during verification
3. number of choice field errors in data set
4. percent of field errors in data set.
Statistical Analysis

Statistical analysis was completed using SPSS for Windows (version 10.0). To test for significant demographic differences between the research and non-research groups, an independent-samples t-test was used to compare age, hours spent working on a computer at home, and hours spent working on a computer at work. Fisher’s Exact Chi-square Test was used to test for significant differences for categorical variables. Repeated measures analysis of variance (ANOVA) was utilized to determine if there were differences between the three writing devices and the two groups of volunteers (R versus NR). The Least Significant Difference test (LSD) was used in conjunction with the ANOVA to determine which writing instruments differed and to adjust for multiple comparisons. Additionally, a separate comparison was made, looking at those with and without Teleform experience. Because of small group sizes, a Kruskal-Wallis one-way ANOVA was performed.
Figure 1. Constrained Print Fields Include Both Text and Numeric Fields

Figure 2. Example of Choice Fields

☐ American Indian or Alaskan Native
☐ Asian/Oriental or Pacific Islander
☐ Black/African American
☐ White/Caucasian
☐ Other

☐ Less than $20,000
☐ $20,000-$39,000
☐ $40,000-$59,000
☐ $60,000-$79,000
☐ $80,000 or more
Chapter 4

RESULTS

Volunteer Demographic Information

Fourteen volunteers (6 females and 8 males) between the ages of 22 and 66 years (mean age = 32.6 ± 14.8 years) completed the study between August, 1997 and November, 1998. Demographic data describing individual characteristics (age, gender, and computer experience) were obtained from the Demographic Information form (Appendix A). To test for significant differences between the research and non-research groups, an independent-samples t-test was used to compare age, hours spent working on a computer at home, and hours spent working on a computer at work. Levene's test was used to test for equality of variances. Fisher's Exact Test was used to test for significant differences for the following categorical variables: number of each gender, have a computer at home and/or work (yes or no), and ever used Teleform software. Descriptive characteristics and results of the analysis are presented in Table 3. The research group had significantly more experience with Teleform software than non-researchers (p=0.02). Otherwise, there were no significant differences between the two groups for age, number of men or women, number with computers at home, number with computers at work, and hours of computer use at home or work. The Levene's test for equality of variance revealed that one variable, the number of hours spent on
computer at home, was not normally distributed ($p=0.031$). However, the results for
the unequal-variance $t$ value revealed no significant differences between the two groups in
this variable.

Table 3. Volunteer Demographics by Type of Research Experience

<table>
<thead>
<tr>
<th>Age in years (mean ± std. dev.)</th>
<th>Researcher (n=6)</th>
<th>Non-Researcher (n=8)</th>
<th>Statistical Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>29.0 ± 8.7</td>
<td>35.3 ± 18.2</td>
<td>0.46</td>
</tr>
<tr>
<td>Number with computers at home</td>
<td>3 Female, 3 Male</td>
<td>3 Female, 5 Male</td>
<td>1.0</td>
</tr>
<tr>
<td>Number use computers at work</td>
<td>5</td>
<td>3</td>
<td>0.14</td>
</tr>
<tr>
<td>Ever use Teleform software</td>
<td>6</td>
<td>4</td>
<td>0.09</td>
</tr>
<tr>
<td>Hours spent on computer at home per week</td>
<td>4.3 ± 4.6</td>
<td>1.3 ± 2.1</td>
<td>0.18</td>
</tr>
<tr>
<td>Hours spent on computer at work per week</td>
<td>24.2 ± 12.4</td>
<td>9.4 ± 14.4</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Independent-samples $t$-test for age, hours on computer home and work.
Fisher’s Exact Test for gender, computer at home/work, familiar Teleform.

Forms

Five separate forms were utilized to test the accuracy, optimal conditions and
software settings for Teleform. Descriptive characteristics of the forms are presented in
Table 4, next page. The forms contained a total of 81 constrained print fields, 120 choice
fields and 451 characters. A total of 199 forms were returned out of an expected 210 (14
volunteers by five forms by three writing instruments). This was due to one particular
individual who did not complete all forms. Only those with complete information were
used in the following analysis ($n=13$).
Table 4. Form Demographics

<table>
<thead>
<tr>
<th>Form Name</th>
<th>Number of Constrained Print Fields</th>
<th>Number of Choice Fields</th>
<th>Number of Characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic Form (1 page)</td>
<td>17</td>
<td>9</td>
<td>166</td>
</tr>
<tr>
<td>NCA (1 p)</td>
<td>9</td>
<td>3</td>
<td>45</td>
</tr>
<tr>
<td>Social Support For Exercise (2 p)</td>
<td>6</td>
<td>25</td>
<td>36</td>
</tr>
<tr>
<td>SF-36 (3 p)</td>
<td>7</td>
<td>37</td>
<td>42</td>
</tr>
<tr>
<td>Health History Form (5 p)</td>
<td>42</td>
<td>46</td>
<td>162</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>81</strong></td>
<td><strong>120</strong></td>
<td><strong>451</strong></td>
</tr>
</tbody>
</table>

Constrained Print Fields

Text and numeric fields were combined for analysis and entitled constrained print fields. The results of the four summary variables are reported in Tables 5 through 8, i.e. Table 5) number of constrained print fields to verify, 6) number of fields to correct during verification, 7) number of field substitution errors after verification, and 8) percent of field errors in data set. The General Linear Model Repeated Measures ANOVA was used to determine whether the three writing instruments and the two research groups differed. The Least Significant Difference Test (LSD) was used when the overall ANOVA was significant. The ANOVA was used to assess differences for each individual's use of the three writing instruments and to adjust for multiple comparisons.

After forms were scanned, the Teleform verifier marked forms needing review. A form would be marked for verification if any of the characters and corresponding fields received a confidence rating of less than 95%. Thus, the first step in the process was to assess differences in the number of constrained print fields to verify between the
two groups and three writing instruments. If for example, a higher number of fields need verification, more time would also be necessary to review each flagged item. Results of the repeated measures ANOVA revealed a significant difference for the writing instrument effect (Table 5, p=0.001). However, there were no significant differences for researchers versus non-researchers (p=0.072) or for the interaction of research groups and writing instruments (p=0.599). The LSD test revealed that the use of a black pen required significantly less fields to verify than blue pen (p=0.001) or pencil (p=0.003). In addition to the summary table, Figure 3 displays the mean results of the two groups and three writing instruments (page 60).

Table 5. Mean Number of Constrained Print Fields to Verify
(Values listed are Means ± Standard Deviations)

<table>
<thead>
<tr>
<th>Group</th>
<th>Black Pen</th>
<th>Blue Pen</th>
<th>Pencil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher (n=6)</td>
<td>25 ± 8.8</td>
<td>39.8 ± 8.9</td>
<td>43 ± 8.8</td>
</tr>
<tr>
<td>Non-Researcher (n=7)</td>
<td>32.7 ± 32.7</td>
<td>46.7 ± 9.5</td>
<td>42.7 ± 10.1</td>
</tr>
<tr>
<td>Total</td>
<td>29.2 ± 8.1*</td>
<td>43.5 ± 9.5</td>
<td>42.8 ± 9.1</td>
</tr>
</tbody>
</table>

*Use of black pen versus blue pen (LSD, p=0.001) and pencil (p=0.003).
After Teleform flags specific character fields to verify, a certain number of those fields will have been assigned a "best guess character". If a best guess character is correct, re-keying the data is not necessary and less time is needed to resolve the flagged item. Yet, a certain number of best guess characters and unread characters will need to be corrected within the constrained print field. Table 6 (page 61) provides the results of the Repeated Measures ANOVA for the number of constrained print fields needing correction during verification. There was a significant difference for the writing instrument effect ($p<0.001$). However, there were no significant differences for researchers versus non-researchers ($p=0.062$) or for the interaction between the research groups and writing instruments ($p=0.550$). The LSD test revealed that the use of a black
pen required significantly less fields to correct than blue pen (p=0.002) and pencil (p<0.001). Figure 4, below, displays a graphical representation of the group and writing instrument differences.

Table 6. Mean Number of Fields to Correct During Verification
(Values listed are Means ± Standard Deviations)

<table>
<thead>
<tr>
<th>Group</th>
<th>Black Pen</th>
<th>Blue Pen</th>
<th>Pencil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher (n=6)</td>
<td>14.3 ± 5.4</td>
<td>25.3 ± 8.0</td>
<td>30.8 ± 6.5</td>
</tr>
<tr>
<td>Non-Researcher (n=7)</td>
<td>20.7 ± 4.8</td>
<td>29.9 ± 9.6</td>
<td>32.4 ± 6.0</td>
</tr>
<tr>
<td>Total</td>
<td>17.8 ± 5.9*</td>
<td>27.8 ± 8.9</td>
<td>31.7 ± 6.0</td>
</tr>
</tbody>
</table>

*Use of black pen versus blue pen (p=0.002) and pencil (p<0.001).

Figure 4. Mean Number of Character Fields to Correct During Verification
Once the data were scanned and verified, data were saved to a Microsoft Excel table, the data were then visually checked to determine the number of substitution errors that were saved to the file. Substitution errors would occur if Teleform accidentally assigned a confidence rating of 95 percent or greater to a character. These characters and corresponding fields would not have been flagged for verification. Thus, the software accepted an incorrect character within the field, resulting in a substitution error. Results for the number of character field substitution errors are displayed in Table 7 and Figure 5. The writing instrument effect was significantly different (p < 0.001). However, there were no significant differences for researchers versus non-researchers (p = 0.774) or for the interaction between the research groups and writing instruments (p = 0.822). The LSD test revealed that the black pen and pencil were significantly different (p = 0.003). The use of a blue pen was also significantly different from pencil (p = 0.002). There were no significant differences for black pen and blue pen (p = 0.658).

Table 7. Number of Character Field Substitution Errors After Verification
(Values listed are Means ± Standard Deviations)

<table>
<thead>
<tr>
<th>Group</th>
<th>Black Pen</th>
<th>Blue Pen</th>
<th>Pencil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher (n=6)</td>
<td>8.3 ± 2.6</td>
<td>9.5 ± 5.5</td>
<td>18.3 ± 11.7</td>
</tr>
<tr>
<td>Non-Researcher (n=7)</td>
<td>8.9 ± 4.8</td>
<td>9.0 ± 4.4</td>
<td>20.7 ± 8.5</td>
</tr>
<tr>
<td>Total</td>
<td>8.6 ± 3.8*</td>
<td>9.2 ± 4.7*</td>
<td>19.6 ± 9.7</td>
</tr>
</tbody>
</table>

*Black Pen and Pencil (p = 0.003)
†Blue Pen and Pencil (p = 0.002)
In order to assess error rates for the constrained print fields, the number of substitution errors detected were tallied, divided by the total number of character fields in the forms (n=81), and finally expressed as a percentage. Results from the ANOVA are displayed in Table 8. Similar to the number of constrained print field errors, there was a significant difference for the writing instrument effect (p<0.001). However, there were no significant differences for researchers versus non-researchers (p=0.774) or for the interaction between the research groups and writing instruments (p=0.822). Since there were no group differences, accuracy rates were determined for each writing instrument by subtracting the substitution error percentage from 100. Thus, the use of black pen was 89.4% accurate, for blue pen 88.6%, for pencil 75.8%.
Table 8. Percent of Character Field Substitution Errors in Data Set
(Values listed are Means ± Standard Deviations)

<table>
<thead>
<tr>
<th>Group</th>
<th>Black Pen</th>
<th>Blue Pen</th>
<th>Pencil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher (n=6)</td>
<td>10.3 ± 3.2</td>
<td>11.7 ± 6.8</td>
<td>22.6 ± 14.4</td>
</tr>
<tr>
<td>Non-Researcher (n=7)</td>
<td>10.9 ± 5.9</td>
<td>11.1 ± 5.4</td>
<td>25.6 ± 10.5</td>
</tr>
<tr>
<td>Total</td>
<td>10.6 ± 4.7*</td>
<td>11.4 ± 5.8†</td>
<td>24.2 ± 12.0</td>
</tr>
</tbody>
</table>

*Black pen versus pencil (p=0.003)
† Blue pen versus pencil (p=0.002)

Choice Fields

Choice fields require one of several boxes to be selected, i.e. choosing a yes or no response. Once a box is selected the corresponding code associated with the answer will be saved to a Microsoft Excel spreadsheet after scanning and verifying. The results of the four summary variables are reported in Tables 9 through 12, i.e. Table 9) number of choice fields to verify, 10) number of choice fields to correct during verification, 11) number of choice field errors after verification, and 12) percent of field errors in data set. As done previously, the General Linear Model Repeated Measures ANOVA was used to determine whether the three writing instruments and the two research groups were significantly different. The Least Significant Difference test (LSD) was also used to assess differences within the three writing instruments.

After forms were scanned, the Teleform verifier marked any choice fields and corresponding forms needing review. A choice field would be marked for verification if any of the choices had been mismarked (i.e. an individual not placing a check directly in the box) or incorrectly completed (i.e. two boxes selected within the same field, with one being erased). Thus, the next step was to assess differences in the number of choice fields
to verify between the two groups and three writing instruments. Results of the repeated measures ANOVA are displayed in Table 9 and Figure 6 (below). There were no significant differences between groups (p = 0.488), writing instruments (p = 0.202), or the interaction between research groups and writing instruments (p = 0.804).

Table 9. Number of Choice Fields to Verify
(Values listed are Means ± Standard Deviations)

<table>
<thead>
<tr>
<th>Group</th>
<th>Black Pen</th>
<th>Blue Pen</th>
<th>Pencil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher (n=6)</td>
<td>18.5 ± 25.3</td>
<td>37.2 ± 29.4</td>
<td>28.2 ± 23.7</td>
</tr>
<tr>
<td>Non-Researcher (n=7)</td>
<td>30.3 ± 32.3</td>
<td>39.4 ± 24.2</td>
<td>38.0 ± 14.0</td>
</tr>
<tr>
<td>Total</td>
<td>24.8 ± 28.7*</td>
<td>38.4 ± 25.5</td>
<td>33.5 ± 18.9</td>
</tr>
</tbody>
</table>

*Black pen versus blue pen (p = 0.073)

Figure 6. Mean Number of Choice Fields to Verify

Writing Instrument
Like the constrained print fields, once specific choice fields are flagged for verification, a certain number of those fields will have been assigned a "best guess choice". If the best guess choice is correct, selecting a new choice from the available options is not necessary, saving additional time during verification. Yet, a certain number of best guess choices will need to be corrected within the constrained print field. Table 10 provides the results of the Repeated Measures ANOVA for the number of choice fields needing correction during verification. The writing instrument effect was significantly different (p=0.017). However, there were no significant differences for researchers versus non-researchers (p=0.338) or for the interaction between the research groups and writing instruments (p=0.477). The LSD test revealed that the use of a pencil required significantly fewer choice fields to correct than black pen (p=0.017) and blue pen (p=0.04). There were no differences between black pen and blue pen (p=0.544). No graphical representation is provided due to the small difference between groups and writing instruments.

<table>
<thead>
<tr>
<th>Group</th>
<th>Black Pen</th>
<th>Blue Pen</th>
<th>Pencil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher (n=6)</td>
<td>0.5 ± 0.5</td>
<td>1.8 ± 2.2</td>
<td>0 ± 0</td>
</tr>
<tr>
<td>Non-Researcher (n=7)</td>
<td>1.0 ± 0.8</td>
<td>0.3 ± 0.5</td>
<td>0.1 ± 0.4</td>
</tr>
<tr>
<td>Total</td>
<td>0.8 ± 0.7</td>
<td>1.0 ± 1.7</td>
<td>0.1 ± 0.3*</td>
</tr>
</tbody>
</table>

*Pencil versus black pen (p=0.017) and blue pen (p=0.04)

Similar to constrained print fields, once the data from choice fields were scanned, verified, and saved, the choice field data were then visually checked to determine the
occurrence of choice field errors. Choice field errors would occur if Teleform accidentally accepted a choice field as correct when in fact there was a mistake. For example, if a check mark did not completely fill the choice field or was a very light marking, Teleform could accept the field as missing (an omission type error). Additionally, if some other mark on the page accidentally crossed one of the choice boxes, the choice may have been accepted as a correct answer (substitution type error). Thus, the software may have accepted an incorrect choice for the field or accepted the field as missing, resulting in a choice field error. Results for the number of choice field errors are displayed in Table 11 and Figure 7. The writing instrument effect was significantly different (p=0.016). There were no significant differences for researchers versus non-researchers (p=0.405) or for the interaction between the research groups and writing instruments (p=0.796). However, the use of black pen had significantly less error than pencil (p=0.028). While use of a blue pen was not significantly different from the use of black pen (p=0.079) or pencil (p=0.10).

Table 11. Number of Choice Field Errors in Data Set
(Values listed are Means ± Standard Deviations)

<table>
<thead>
<tr>
<th>Group</th>
<th>Black Pen</th>
<th>Blue Pen</th>
<th>Pencil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher (n=6)</td>
<td>3.5 ± 7.2</td>
<td>11.3 ± 17.0</td>
<td>18.2 ± 31.8</td>
</tr>
<tr>
<td>Non-Researcher (n=7)</td>
<td>8.7 ± 11.4</td>
<td>19.4 ± 24.4</td>
<td>31.3 ± 28.4</td>
</tr>
<tr>
<td>Total</td>
<td>6.3 ± 9.7*</td>
<td>15.7 ± 20.9</td>
<td>25.2 ± 29.5</td>
</tr>
</tbody>
</table>

*Black pen versus pencil (p=0.028).
In order to assess error rates for choice fields, the actual percent of field errors was determined by dividing the total number of choice field errors, across all forms, by the total number of choice fields (n=120). Means and standard deviations from the ANOVA are displayed in Table 12. The results revealed that the writing instrument effect was significantly different (p=0.016). However, there were no significant differences for researchers versus non-researchers (p=0.405) or for the interaction between the research groups and writing instruments (p=0.796). Thus, error rates were significantly better for black pen versus blue pen (p=0.037) and pencil (p=0.017), and for blue pen versus pencil (p=0.006). Accuracy rates were determined for each writing
instrument by subtracting the substitution error percentage from 100. Thus, accuracy rates for the use of black pen was 94.7%, for blue pen 86.9%, and for pencil 79%.

Table 12. Percent of Choice Field Errors in Data Set
(Values listed are Means ± Standard Deviations)

<table>
<thead>
<tr>
<th>Group</th>
<th>Black Pen</th>
<th>Blue Pen</th>
<th>Pencil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher (n=6)</td>
<td>2.9 ± 6.0</td>
<td>9.4 ± 14.2</td>
<td>15.1 ± 26.5</td>
</tr>
<tr>
<td>Non-Researcher (n=7)</td>
<td>7.3 ± .9.5</td>
<td>16.2 ± 20.3</td>
<td>26.1 ± 23.7</td>
</tr>
<tr>
<td>Total</td>
<td>5.3 ± 8.1*</td>
<td>13.1 ± 17.4†</td>
<td>21.0 ± 25.0</td>
</tr>
</tbody>
</table>

* Black pen vs blue pen (p=0.037), black pen vs pencil (p=0.017)
† Blue pen versus pencil (p=0.006)

Comparison of Those with Teleform Experience and Those Without

Because there were no significant differences between the research group and non-research groups, a separate analysis was completed looking at those individuals with Teleform experience (n=4) and those without (n=10). This might add additional information regarding the notion that previous Teleform training may enhance the accuracy of scanning. Due to the positive findings in the previous analysis, only forms completed with the black pen were analyzed. Additionally, all 14 individuals had data available for this comparison. Because of the small and unequal group sizes, a Kruskal-Wallis one-way analysis of variance was used. The results of the analysis are provided in Table 13 (page 70). Those with Teleform experience had significantly less character fields to verify and correct, less choice fields to verify, and less choice field errors than those without Teleform experience.
Table 13. Comparison of Those With and Without Teleform Experience Using Black Pen (Kruskal-Wallis One-Way Analysis of Variance)

<table>
<thead>
<tr>
<th></th>
<th>With Teleform Experience (n=4)</th>
<th>Without Teleform Experience (n=10)</th>
<th>Statistical Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td># of character fields to verify</td>
<td>19.8 ± 5.0</td>
<td>32.6 ± 5.8</td>
<td>0.004</td>
</tr>
<tr>
<td># of character fields to correct</td>
<td>11.0 ± 1.8</td>
<td>19.8 ± 5.0</td>
<td>0.011</td>
</tr>
<tr>
<td># of character field errors</td>
<td>8.3 ± 1.7</td>
<td>9.0 ± 4.3</td>
<td>0.477</td>
</tr>
<tr>
<td>% of character field errors</td>
<td>10.2 ± 2.1</td>
<td>11.1 ± 5.3</td>
<td>0.477</td>
</tr>
<tr>
<td># of choice fields to verify</td>
<td>3.5 ± 5.1</td>
<td>33.8 ± 28.3</td>
<td>0.019</td>
</tr>
<tr>
<td># of choice fields to correct</td>
<td>0.3 ± 0.5</td>
<td>0.9 ± 0.7</td>
<td>0.124</td>
</tr>
<tr>
<td>% of choice field errors</td>
<td>0 ± 0</td>
<td>9.1 ± 9.9</td>
<td>0.034</td>
</tr>
</tbody>
</table>

Comparison of Constrained Print Fields and Choice Fields

Finally, a comparison was made between the constrained print fields and choice fields to ascertain whether one particular type of field yielded significantly better results than the other. Because the number of fields were different, 81 for constrained print fields and 120 for choice fields, the results for each field was converted to a percentage for comparative purposes. A paired-samples t-test was used to assess statistical differences. Only forms completed in black pen were compared in this analysis. Results of the analysis are presented in Table 14 (below). Choice fields required significantly less fields to verify, and less fields to correct.

Table 14. Comparison of Constrained Print Fields with Choice Fields Using Black Pen (Paired-Samples T-Test)

<table>
<thead>
<tr>
<th></th>
<th>Constrained Print Fields (n=14)</th>
<th>Choice Fields (n=14)</th>
<th>Statistical Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of fields to verify</td>
<td>44.2 ± 11.2</td>
<td>17.5 ± 20.2</td>
<td>0.001</td>
</tr>
<tr>
<td>% of fields to correct</td>
<td>27.1 ± 8.8</td>
<td>0.4 ± 0.4</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>% of field errors</td>
<td>10.8 ± 4.5</td>
<td>5.4 ± 7.8</td>
<td>0.069</td>
</tr>
</tbody>
</table>
In summary, the accuracy, optimal conditions for enhancing accuracy, and the feasibility of using Teleform in an environment similar to a public health setting were explored. Fourteen men and women between the ages of 22 and 66 years took part in the study. Individuals with research experience and those without experience were similar in age, access to computers at home or work, and time spent on a computer at home or work. ($p=0.09$ to $p=1.0$). However, researchers had significantly more experience with Teleform than non-researchers ($p=0.02$). All individuals completed five forms on three separate occasions with three different writing instruments. Each set of the five forms contained 81 constrained print fields and 120 choice fields. Ninety five percent of the forms were completed and returned for further analysis (199 out of 210 expected).

Four summary variables were created from the constrained print field data. This included the number of constrained print fields to verify, number of fields to correct during verification, number of field substitution errors, and percent of field errors in the final data set. There were no significant differences for researcher versus non-researchers in any of the four constrained print field variables ($p=0.062$ to $p=0.774$). However, use of a black pen required significantly less fields to verify than blue pen (29.2 versus 43.5, $p=0.001$) or pencil (29.2 versus 42.8, $p=0.003$). Black pen also required significantly less fields to correct during verification than blue pen (17.8 versus 27.8, $p=0.002$) and pencil (17.8 versus 31.7, $p<0.001$). Likewise, the number of substitution errors for black pen was significantly less than pencil (8.6 versus 19.6, $p=0.003$) but not for blue pen (8.6
versus 9.2, p=0.0658). However, the use of a blue pen also resulted in fewer substitution errors than pencil (9.2 versus 19.6, p=0.002). Overall, error rates for constrained print fields were 10.6 percent for black pen, 11.4 percent for blue pen, and 24.2 percent for pencil.

Four summary variables were also created for the choice fields. Similar to constrained print fields, this included the number of choice fields to verify, number of choice fields to correct during verification, number of choice field substitution errors, and percent of choice field errors in the final data. Like the constrained print fields, there were no significant differences for researcher versus non-researchers in any of the four choice field variables (p=0.338 to p=0.448). Unlike the constrained print fields, there were no significant differences between the writing instruments for the number of fields to verify (p=0.073 to p=0.447) despite the fact that the mean values appear to be different (24.8 black, 38.4 blue, 33.5 pencil). Interestingly, use of a pencil required significantly less choice fields to correct during verification than black pen, although the differences appear to be small (0.8 versus 0.1, p=0.017). Even so, once the data were checked for errors, black pen had significantly less substitution errors than pencil (6.3 versus 25.2, p=0.028) but not for blue pen (6.3 versus 15.7, p=0.079). Overall, the analysis revealed 5.3 percent error for black pen, 13.1 percent error for blue pen, and 21 percent error for pencil, all writing instruments were significantly different (p=0.006 to 0.037). Additionally, all fields were converted to percentages for comparative purposes. For those forms completed with black pen, choice fields required significantly less fields to verify (17.5% versus 44.2%, p=0.001) and less fields to correct (0.4% versus 27.1%,
p < 0.001) than constrained print fields. Although the percent of field errors was less for choice fields (5.4% versus 10.8%) the differences were not significantly different. Figures 8 through 10 (pages 73 to 74) compare the results of the constrained print fields with choice fields and the three writing instruments.

Figure 8. Percent of Character Fields and Choice Fields to Verify

![Figure 8. Percent of Character Fields and Choice Fields to Verify](image)

Figure 9. Percent of Character Fields and Choice Fields to Correct

![Figure 9. Percent of Character Fields and Choice Fields to Correct](image)
Finally, because there were no significant differences between researchers and non-researchers, one last comparison ascertained differences between individuals with Teleform experience and those without. For those forms completed with black pen, individuals with Teleform experience had significantly less character and choice fields to verify, less character fields to correct, and less choice field errors than those without Teleform experience.
Chapter 5

DISCUSSION

This study determined the accuracy, optimal conditions for enhancing accuracy, and the feasibility of using Teleform in a public health setting. The specific aims of the study were to determine whether: 1) Teleform software provides an accurate methodology for data entry and whether data checking is necessary, 2) The use of different writing instruments influence accuracy, 3) The type of person and prior research experience influence accuracy, and 4) Teleform software provides a viable alternative to data entry in the public health field.

Accuracy of Teleform Software and Scanning Technology

Accurate, relevant, and timely data can provide government and public health authorities with crucial information regarding disease outbreaks, morbidity and mortality rates, health care costs, and health care utilization. Without such data, disease outbreaks would become unrestrained, the quality of medical treatment would be impossible to ascertain, changes in disease morbidity or mortality would be unknown, and public health policy would be based on speculation rather than fact. Even so, once data are collected the accuracy of large data sets are not usually ascertained or reported. The study of the “teenage widows” from the 1950 census report revealed how a simple and small coding error may greatly exaggerate numbers in rare categories (Coale and Stephan, 1962). Although there are no standards for data collection or for accuracy in
the field of Public Health, several research studies within clinical research settings have assessed accuracy for a variety of data collection techniques. These techniques include computer assisted data collection, single data entry with and without error checking, duplicate data entry, and more recently, automated techniques for data capture (Christianson et al., 1990; Neaton et al., 1990; Bagniewska et al., 1986; Kronmal et al., 1978; Kanmaz et al., 1997; Denwood et al., 1996; Smyth et al, 1997; Jorgensen and Karlsmose, 1998).

In many clinical research settings duplicate data entry (DE) has been perceived as the “gold standard” and most acceptable methodology for data collection (Reynolds-Haertle et al., 1992). This support is due to the very low error rates reported in the clinical literature. Studies have documented the number of errors from DE to be between 5.9 and 15 errors per 10,000 fields. This translates into error rates of 0.06% to 0.15% (Kronmal et al, 1978; Duchene et al, 1986; Prod’homme et al., 1989; Reynolds-Haertle and McBride, 1992; Gibson et al., 1994). However, the need for accuracy must be weighed against the high costs and the increased time necessary for DE. When time was measured between methods, duplicate data entry required 37% more time than the single data entry method (Reynolds-Haertle and McBride, 1992), 2,250% more time than scanning technology with choice fields (0.4 hours versus 9.1 hours, Smyth et al, 1997), and 264% more time than scanning technology with both choice and constrained print fields (142 minutes versus 375, Jorgensen and Karlsmose, 1998). Additionally, the costs for DE were 58% more than single data entry and 277% more than scanning technology (Reynolds-Haertle and McBride, 1992; Jorgensen and Karlsmose, 1998).
Consequently, DE may not be a viable alternative in public health settings where budget declines are more prevalent than not (Gerzoff et al, 1996). Take for example the setting where infection control and surveillance programs are essential, there may be little time to key data twice. Yet, timely and accurate data are necessary. As Declich and Carter (1994) state “the quality of a surveillance system is only as good as the quality of the data being collected.” Thus, a system, which involves automated data capture, can be very appealing to many health departments. The promises of ADC include immediate access to data, a decrease in personnel, reduction in labor costs, and high quality data (Bish, 1996; Jilovec, 1996; Kasten et al, 1992).

A small number of studies have assessed accuracy rates for ADC techniques. Error rates for bar coding have been reported as 1.7% (Kanmaz et al, 1997), for optical mark technology 1% (Denwood et al., 1996), scanning technology with choice responses only 0.02% to 0.04% (Smythe et al., 1997; Jorgensen and Karlsmose, 1998), 0.14% for constrained print fields with an 80% Teleform confidence threshold (Jorgensen and Karlsmose, 1998), and 0.1% for constrained print fields with a 99% Teleform confidence threshold (Jorgensen and Karlsmose, 1998). In the current study, the accuracy of Teleform software was not only dependent on the type of data field completed (choice field versus constrained print field), but also on the type of writing instrument utilized and whether the individual had prior Teleform experience.
Accuracy and Scanning Differences Between Different Writing Instruments

Constrained Print Field Results

In the current study, three different writing instruments were provided to the 14 volunteers in a randomized manner. After forms were scanned using Teleform, most forms were held for review when any characters and corresponding constrained print fields received a confidence rating of less than 95%. A form, which requires many fields to verify or correct, will take more time to review than a form needing verification of a few items. When constrained print fields were compared, the number of fields to verify for black pen (29.2 out of 81 total) was significantly less than that for blue pen (43.5, p=0.001) and pencil (42.8, p=0.003). Following verification, a certain number of characters and corresponding fields needed to be corrected. In most instances, Teleform assigned a “best guess character” for any item in question. If the best guess was correct, no change was necessary and less time was needed to resolve the specific item. Like the number of fields to verify, use of the black pen (17.8) required significantly less fields to correct than blue pen (27.8, p=0.002) and pencil (31.7, p<0.001). After scanning and verification, the final data set was visually checked to ascertain the number of substitution errors. Substitution errors occurred when, for example, Teleform accepted a character as correct at the 95% confidence threshold, when in fact the character was in error. The number of constrained print field errors were 8.6 for black pen, 9.2 for blue pen, and 19.6 for pencil. Black and blue pens had significantly fewer errors than pencil (p≤0.003). Yet, there were no significant differences for black versus blue pen (p=0.658). The overall error rates for constrained print fields were 10.6% for black pen,
11.4% for blue pen, and 24.2% for pencil. Thus, to reduce the number of constrained print fields to verify and the number of fields to correct, the use of a black pen performed better than blue pen or pencil. This would suggest that less time is necessary when scanning and verifying forms completed with black pen. And finally, to reduce the number of errors in the final data set, one should use a black or blue pen.

**Choice Field Results**

Similar to constrained print fields, once forms were scanned, any questionable choice fields were held for review. Results of the repeated measures ANOVA, revealed no significant differences between the number of choice fields to verify for the three writing instruments: 30.3 for black pen, 38.4 for blue pen, and 33.5 for pencil (p=0.073 to 0.447). Following verification, the number of choice fields needing correction was also compared for each writing instrument. Although the number of fields to correct was small, the use of a pencil required significantly less choice fields to correct than black pen (0.1 versus 0.8, p=0.017). There were no significant differences for the use of blue pen (1.0) when compared to black pen or pencil.

Following scanning and verification the final data set was then visually checked against the original forms to ascertain the number of substitution or omission type errors within the choice fields. Interestingly, the number of choice field errors for pencil was significantly higher than that for black pen (25.2 for pencil versus 6.3 for black pen, p=0.028) even though the use of a pencil required significantly less fields to verify. The use of blue pen did not differ significantly from the use of black pen (6.3 versus 15.7, p=0.658). The overall error rates for choice fields were 5.3% for black pen,
13.1% for blue pen, and 21% for pencil. Thus, the results of this analysis indicated that even though the use of a pencil required a smaller number of choice fields to verify and correct than black pen, the use of the pencil resulted in a much higher error rate than black or blue pen. Although the difference between substitution errors and omission errors were not reported in this study, upon visual examination, the final choice field errors appeared to be more of the omission type than the substitution type when using a pencil.

Other Research and Summary

To date, no other known ADC study has looked at differences in scanning technology when using different writing instruments. Smythe and colleagues (1997) state, “ensuring that a distinct mark is made on the questionnaire, using a pen containing black ink eliminates the problem of indistinct marks being undetected.” Yet, no data are presented to support this statement. In summary, when completing forms containing constrained print fields, the use of a black pen significantly reduced the number of fields to verify and the number of fields to correct than blue pen or pencil. These findings would suggest that less time would be needed for scanning and verification when using a black pen to complete forms containing constrained print fields. For choice fields, on the other hand, there were no differences in the number of fields to verify for any of the writing instruments. Yet the use of a pencil required less fields to correct then black pen. This difference, however, was not meaningful once the data were checked for accuracy. The number of final errors was substantially higher when using a pencil than when using a black or blue pen. Finally, for both constrained
print fields and choice fields, the number of final substitution errors in the data set were reduced when forms were completed in black or blue pen. Overall, the use of a black pen resulted in less constrained print fields to verify and correct than blue pen and pencil. Black pen also resulted in less error than pencil (for both types of fields) while black pen and blue pen were not significantly different. Thus, when using forms containing both constrained print fields and choice fields, more favorable results should be obtained with the use of a black pen.

Comparison of Those With and Without Research Experience

In the current study, individuals with research experience were compared to those without such experience. The hypothesis would be that those with more research experience would have better training, better accuracy, and would follow instruction better than those without research experience. There were six individuals with research experience and eight without. There were no differences between the two groups in age, number of males or females, number with computers at home or work, or the amount of time spent on computers at home or work. However, four researchers had prior Teleform experience and this was significantly different than non-researchers (p=0.02). One individual in the non-research group did not complete all three sets of forms and thus, the final comparison between research and non-research groups included 13 individuals. Unexpectedly, there were no significant differences between the two research groups for number of fields to verify, number of fields to correct, or the amount of error. This was true for both the constrained print field variables and
choice field variables. Because there were no differences, a separate analysis compared those with Teleform experience \((n=4)\) with those without \((n=10)\).

The hypothesis would be that an individual with Teleform experience would have better training and accuracy than one without such experience. The results of this comparison might also indicate that research experience alone does not guarantee better form completion or better Teleform performance. Due to the aforementioned findings, this analysis focused on forms completed with black pen. Additionally, using the results from the black pen increased the sample size to 14. The results of the Kruskal-Wallis ANOVA found that those with prior Teleform experience had significantly fewer constrained print fields to verify \((19.8 \text{ versus } 32.6, p=0.004)\), significantly fewer constrained print fields to correct \((11.0 \text{ versus } 19.8, p=0.011)\), significantly fewer choice fields to verify \((3.5 \text{ versus } 33.8, p=0.019)\), and significantly fewer choice fields errors in the data set \((0 \text{ versus } 9.1, p=0.034)\). Additionally, these findings show that those with Teleform experience had 100% accuracy for choice fields. Thus, the results from this analysis suggest that prior training and knowledge of Teleform software greatly enhance the overall performance of Teleform as well as the choice field accuracy rate. In the current study, very minimal instruction was given as to how to complete the forms to enhance scanning and accuracy. This was done to test the ability of Teleform and scanning software to work in a less structured setting as compared to a more rigid clinical research setting. In summary, when using Teleform software, prior training on how to complete both constrained print fields and choice fields is strongly recommended.
To date, the study by Jorgensen and Karlsmose (1998) is the only study which compared the accuracy of Teleform software between different types of individuals. In this study, the accuracy of constrained print characters (not constrained print fields) was compared between one of the authors (a researcher) and a group of physicians. One form containing 10 numeric constrained print fields was utilized for the comparison. The researcher filled in six of the 10 constrained print fields on all 401 forms used in the study while the remaining four fields were completed by the physicians. Similar to the current study, the authors state that, “there was an important difference in error rate depending on who filled in the questionnaires.” For the fields completed by the researcher the constrained print field error rates were 0.02% for the 80% confidence threshold and 0.01% for the 99% confidence threshold. While the physicians obtained 0.4% error at the 80% confidence threshold and 0.3% for the 99% confidence threshold (Jorgensen and Karlsmose, 1998). Although differences were reported between the researcher and physicians, the authors did not ascertain whether these error rates were significantly different from each other. This was mainly due to the fact that the authors were focusing on determining significant differences between different data entry methods (DE, single data entry, and Teleform). Yet, the authors state that

“numeric characters filled in by physicians resulted in substantially more errors than characters filled in by the researcher. The reason for this is that Teleform was sensitive to whether or not the characters were written in accordance with the recommendations given to enhance recognition, and some respondents did not follow these recommendations.”
Comparatively, the reported error rates by Jorgensen and Karlsmose (1988) are substantially less than those found in the current study for constrained print fields, for those with Teleform experience (10.2%) and for those without (11.1%). However, several differences should be noted between the studies, 1) in the current study the constrained print fields included both text and numeric characters thereby increasing the opportunity for error, 2) there were a total of 81 constrained print fields as compared to their 10 fields also increasing the opportunity for error, and 3) the variability seen for one trained researcher completing constrained print fields should be significantly less than that for a group of individuals. Further discussion and comparison of accuracy rates can be found in the next section. Interestingly, Jorgensen and Karlsmose (1998), state that “when using AFP (Teleform) for research purposes, it is advisable to avoid numeric fields if it cannot be assured that respondents will adhere to the recommendations on how to write characters to enhance recognition.” In summary, the findings of both the current study and the study by Jorgensen and Karlsmose (1998) suggest that prior training on how to complete both constrained print fields and choice fields is paramount. When instructions are followed, the overall performance and accuracy of Teleform can be greatly enhanced. Additionally, another option would be to limit the number and type of individuals who complete the questionnaires. However, this might only be possible in a setting where questionnaires are administered by trained interviewers.
Overall Accuracy of Teleform Software

Error Rates for Constrained Print Fields

Because the use of a black pen performed better than blue pen and pencil in most instances, the following discussion will focus on the resulting error rates for black pen. When reviewing the results for constrained print fields (confidence threshold of 95%), the number of errors for all individuals combined was 8.6 per 81 fields. This would translate to 1,061 errors per 10,000 fields or an error rate of 10.6%. To date, the study by Jorgensen and Karlsmose (1998) is the only comparative study which assessed the accuracy of constrained print fields using Teleform. The current study’s constrained print field error rates are reportedly much higher than the 0.1% (99% confidence) and 0.14% (80% confidence) overall error rates for constrained print fields reported by Jorgensen and Karlsmose (1998). However, several differences are noted between the two studies in relation to constrained print fields.

First, the total number of constrained print fields in their study was small, 10 fields (with 34 characters) out of a total of 29 total fields. In the current study, there were 81 constrained print fields (with 451 characters) out of a total of 201 fields, yielding a slightly higher percentage of constrained print fields, 40% in the current study versus 34%. Additionally, in the study by Jorgensen and Karlsmose (1998), constrained print fields included only numeric data. While the current study included both numeric and text data in the constrained print fields. Neaton et al. (1990) reports that higher error rates are seen for alphabetic as compared to numeric fields. This is attributed to the fact that 1) alphabetic fields usually have more characters than numeric
fields, and 2) the response set for alphabetic fields is larger (26 characters) than for numeric fields (10 characters) (Neaton et al., 1990).

The second reason contributing to higher error rates in the current study, may be related to the form review process prior to scanning. In the Jorgensen and Karlsnose study (1998) all the forms were reviewed prior to scanning by one of the authors to “check for obvious mistakes and to recode various information.” Also, as mentioned previously, six of the ten constrained print fields were then filled in by the one author. Only four constrained print fields were completed by different physicians who participated in the project. In the current study, there were no differences in error for constrained print fields for researchers versus non-researchers (10.3% versus 10.9%), nor for those with Teleform experience versus those without (10.2% versus 11.1%). Still, it seems likely, that one individual with Teleform experience who completes 60% of the constrained print fields, would have less error than when a variety of individuals complete forms.

The third reason for higher error rates in the current study may be related to the small number of individuals involved in the study. There were 13 individuals who completed the study with five forms ranging from one to five pages. The study by Jorgensen and Karlsnose (1998) included 401 questionnaires completed by 195 different physicians. In clinical research studies sample size estimation is critical to demonstrating significant differences between treatment groups. Power analysis consists of determining how large a sample is required to detect actual differences of some meaningful size (Dawson-Saunders and Trapp, 1994). Type II errors, missing a
significant difference when one actually exists, occur more frequently when the sample size is too small to detect actual differences. Yet, the small sample in the current study did yield some significant findings. However, in a small sample it may also possible to have one or two individuals who influence the mean in one direction or another. For example, if two individuals with really poor handwriting had much higher error rates than everyone else in the sample, the overall error rate could be inflated. Yet, the number of constrained print field errors were normally distributed within this sample (Shapiro-Wilk statistic, $p=0.737$). Still, it may be possible that this small sample was not an accurate representation of the general population. Thus, conducting this study on a larger scale might yield error rates more comparable to reported findings by Jorgensen and Karlsmose (1998).

A final consideration when comparing error rates between the two studies, might be related to the confidence threshold setting. In the current study all constrained print fields were set to a 95% confidence threshold. Yet, the error rates reported by Jorgensen and Karlsmose (1998) for the 80% confidence threshold (0.14%) was still lower than the error rate reported in this study. The study by Jorgensen and Karlsmose (1998) also showed slight differences between the 80% and 99% confidence threshold (0.1% versus 0.14%), but the two accuracy rates were not significantly different. Although these differences don't appear to be great, it may be possible that changing the confidence threshold to 99% or 100% might also reduce error rates. The downfall of using the 100% confidence threshold would be that the number of fields to verify would significantly increase, as well as, the overall time for verification.
Error Rates for Choice Fields

In addition to error rates for constrained print fields, error rates were also ascertained for choice fields. The results of this study determined that the number of errors for choice fields, for all individuals, was 6.3 per 120 fields. This would translate to 525 errors per 10,000 fields or an error rate of 5.3%. However, a small group of individuals with Teleform experience had no choice field errors in the data set or an error rate of 0%. To date, only two studies have determined accuracy rates for choice fields using Teleform or similar software (Smyth et al., 1997; Jorgensen and Karlsmose, 1998). When looking at the current study's overall choice field error rate, the values are higher than the 0.02% reported by Smyth and colleagues (1997) and the 0.02% error rate reported by Jorgensen and Karlsmose (1998). While the error rate for those with Teleform experience (0%) was comparable to the two studies.

When comparing the overall choice field error rate found in the current study with those of the aforementioned studies, the higher error rate may be related to one of several factors. First, the total number of choice fields were different between the studies, Smyth and colleagues (1997) findings were based on 31 choice fields (two page form), while Jorgensen and Karlsmose (1998) reported on only 19 choice fields (four page form). In the current study, there were 120 choice fields per person (total of 12 pages), nearly four to six times more than the aforementioned studies. Additionally, in a previous study by Neaton and colleagues (1990), the authors found that error rates were higher for “long forms” (908 fields) when using single data entry with extensive error checking (Neaton et al., 1990). Although a comparison was not made between
form lengths (one page versus five page forms) in the current study, it seems possible that higher error rates might occur for each individual when more options for error are available.

Second, differences in study methodology may also be a contributing factor to the higher error rate in the current study. In the study by Jorgensen and Karlsmose (1998) a visual review of all data was performed by one of the authors prior to scanning. The author states that all obvious mistakes were corrected, as well as, the “recoding of various information (Jorgensen and Karlsmose, 1998)” In the current study there was no review or corrections made to any forms prior to scanning. Again, this was done to field test Teleform in an unrestricted setting. For the second study, the authors state that a “validate by re-scanning” technique was performed. In other words, a technique similar to duplicate data entry was performed using the scanning technology (Smyth et al, 1997). It seems likely that scanning the same form twice and then using a compare function would also result in fewer choice field errors, and that rates similar to DE would be attainable. However, the application of this validation technique would require additional technical expertise, something that might not be possible in a public health setting.

Although the overall error rates for choice fields were higher than the reviewed studies, an error rate of 0% was obtained in the current study for those with prior Teleform experience. This suggests that prior training on how to complete choice fields might greatly improve the overall accuracy of Teleform. Thus, two reasonable suggestions for improving overall accuracy would be 1) the inclusion of prior Teleform
training and 2) a review of all forms and correction of obvious mistakes prior to scanning.

**Comparison of Constrained Print Fields and Choice Fields**

A final comparison was made between constrained print fields and choice fields to determine whether one type of field performed better than the other. The number of fields to verify and the number of fields to correct were converted to percentages for comparative purposes. Results of the paired-samples t-test revealed that choice fields required significantly less fields to verify (17.5% needed verification versus 44.2%, p=0.001) and significantly less fields to correct (0.4% versus 27.1%, p<0.001). However, the difference in accuracy between the two fields was not significantly different (5.4% versus 10.8%). When looking at the results for constrained print fields, it seems likely that the length and type (numeric versus text) of constrained print fields would affect the overall performance of scanning technology. The use of choice fields requires much less information to process than constrained print fields. Neaton and colleagues (1990) reported that error rates are higher for alphabetic fields such as name and address and for long fields like social security number. Thus, to improve the overall performance of Teleform technology one might try to limit the number or length of constrained print fields.

**What is an Acceptable Error Rate?**

In addition to assessing the accuracy of Teleform software, the first aim questions whether data checking is necessary following the scanning and verification steps. In the current study, the accuracy of Teleform software was dependent on the type of data
field completed (choice field or constrained print field), the type of writing instrument used, and whether the individual had prior Teleform experience. The question might then be phrased, how much error is acceptable and how much error is too much? More specifically, is the amount of error in this study acceptable? If so, a considerable amount of time would be saved from further data checking. A review of many different data collection techniques revealed errors of less than 100 per 10,000 fields or an error rate of less than 1% (Kronmal et al., 1978; Bagniewska et al., 1986; DuChene et al., 1986; Neaton et al., 1990; Reynolds-Haerlte and McBride, 1992; Smyth et al., 1997; Jorgensen and Karlsmose, 1998). In clinical research settings duplicate data entry has been deemed the gold standard for data collection with error rates between 0.06% to 0.15% (Kronmal et al., 1978; Duchene et al., 1986; Prod'homme et al., 1989; Reynolds-Haerlte and McBride, 1992; Gibson et al., 1994). Only two known studies have assessed the accuracy of form scanning technology. Both of these studies report error rates comparable or better than those for duplicate data entry with error rates for choice fields between 0.02% to 0.04% and error rates for constrained print fields between 0.14% and 0.1% (Smythe et al., 1997; Jorgensen and Karlsmose, 1998). These reported error rates are admittedly very low and the only similar error rate found in this study was for choice field data, completed by four individuals with prior Teleform experience. However, both studies used more stringent methodologies than what was used in the current study. One study used a duplicate scanning technique (Smythe et al., 1997) and the other reviewed all forms prior to scanning with many fields being completed by one Teleform-experienced individual. The current study was designed in a less controlled manner, so as to
ascertain the accuracy of scanning technology with minimal time and effort on the part of the research setting. The goal was to imitate a small public health setting, such as a local health department, where minimal staff, time, and resources are available. Consequently this study did not provide detailed instruction to the volunteers in the study on how to complete the forms. Most likely then, the results in this study are based on a “worst-case scenario.” Marinez and colleagues (1984) state that “Ideally, one seeks an error-free study rather than merely control of the errors. In most situations, however, the resources required for an error-free study are unreasonable.” Yet, many clinical researchers believe that an error rate of 10 or fewer errors per 10,000 fields (0.1% or less) is possible and that this rate should be set as an achievable goal for research studies (Neaton et al, 1990; Reynold-Haertle et al., 1992; Glassman et al., 1995). In response to these guidelines, several authors have raised the question as to whether or not the conclusions of the study would differ before and after correcting for errors (Neaton et al, 1990; Arndt and Woolson, 1993; Arndt et al., 1994; Day et al., 1998). For example, would an error rate of 2% significantly change the results of a study as compared to an error rate of 0.1%?

To address the aforementioned question, Arndt and colleagues (1994) assessed the effect of errors in a multicenter medical study. The authors assessed the number of errors at each of seven steps during the data entry and checking process. The steps involved a broad range of techniques including single data entry, duplicate data entry, visual review of forms, range checking, a review by an error-checking software program, and duplicate checks. Following single data entry the authors found 2.4%
error, after duplicate data entry the error rate dropped to 0.098% and so on. The authors state that their 2.4% error rate may initially appear to be reasonable. However, they state that all but 14 of the 688 forms processed were affected by error. Additionally, the impact of these errors would have significantly changed the data's reliability, the study's overall conclusions, and possibly the choice of analysis (Arndt et al., 1994).

Additionally, some discussion has focused on the type or size of the specific error found in the data set. A study by Day and colleagues (1998) discussed the type of changes made to a data set when a variety of types of errors are introduced. For example, when a systolic blood pressure value of 125 mmHg has one of three digits changed, several types of error may occur. If the five is erroneously recorded as four, leading to a value of 124 mmHG, the resulting error may be considered minor. Yet, if the one is erroneously recorded as an eight, creating the wrong value of 825 mmHG, the resulting error would be substantial. However, the authors state that “the most dangerous types of errors may be those which alter a true value of say, 125 to 145 or 165. Such errors may materially affect the conclusions of a study but would be very difficult to check except with the aid of double data entry (Day et al., 1998).” Likewise Arndt and Woolson (1993) demonstrated that it is “not so much the number of mistakes, but the shape and location of the errors' distribution relative to the distribution of correct values.” Mistakes in their study had the greatest effects when the variable was highly informative or when an incorrect item forced following items to be skipped (Arndt and Woolson, 1993). Thus, the aforementioned studies support the
notion that final conclusions can differ significantly before and after correcting for errors. Consequently, it appears that when applying these findings to the current study's reported error rates, that some form of data checking should be required following scanning and verification. Additionally, this recommendation should hold until other types of correction techniques are evaluated. For example, could the overall error rate be reduced by reviewing all forms prior to scanning, adding range and consistency checks, providing training on form completion, minimizing the number of constrained print fields, or using trained Teleform individuals for completing interviews? Until further notice, when using Teleform as described in the current study, data checking should be completed following the scanning and verification process.

Feasibility of Using Teleform in a Public Health Setting

According to the Webster's Dictionary (1984), feasible is defined as 1) the ability to be accomplished, possible, or 2) appropriate and suitable. Although a considerable amount of public health data has been collected and made accessible, public health agencies need to collect and “make available information regarding the health of the community, including statistics on health status, community health needs, and epidemiologic and other studies of health problems (Institutes of Medicine, 1988).” Likewise, policymakers want information on a variety of issues including the availability of health services, the costs and quality of services, patient outcomes, the health status of populations and subpopulations, levels of health in different regions, and access to care for special populations (Feinleib, 1993a; Roos et al., 1996; Musser, 1996).
Further, to obtain additional funding or grant monies for public health programs, some type of data collection and analysis will most likely need to be completed. Yet, many public health agencies have experienced budget declines (Gerzoff et al., 1996). Due to lack of resources a simple, fast, and accurate system for data entry would be a valuable asset to any program. The cost of Teleform software and the scanner, used in this study, was $3,000. This is a much more affordable methodology than something like duplicate data entry where the costs were 58% more than single data entry and 277% more than scanning technology (Reynolds-Haertle and McBride, 1992; Jorgensen and Karlsmose, 1998). Likewise, the amount of time saved by scanning was overwhelming when compared to duplicate data entry. In one study, the time for scanning was 0.4 hours as compared to 9.1 hours for DE (Smythe et al., 1977). However, it should be noted that a training period should be allowed when implementing a system like Teleform. At the Claude Pepper Older Americans Independence Center, there were some problems with installation and setup that needed to be resolved by a trained individual. Additionally, staff needed to be trained on how to design Teleform forms and how to use the scanner. A trained data manager would most likely have fewer problems and need less training than someone without such experience. Even so, the costs and time savings associated with Teleform should override any initial inconvenience of learning the automated technique. Finally, the accuracy of Teleform should be considered. Most error rates in the current study were higher than anticipated but were obtained under less controlled conditions. However, those with Teleform experience reached 100% accuracy for choice field data. Additionally, other
researchers have achieved low error rates for both choice fields and constrained print fields using scanning technology, and the reported rates met the established standard of less than 10 errors per 10,000 fields (Smythe et al., 1997; Jorgensen and Karlsmose, 1998). Thus, the costs and time savings associated with this automated system as well as the potential to obtain almost perfect accuracy, make Teleform a feasible and appropriate tool for public health agencies.

Summary and Conclusions

The primary purpose of this study was to determine the accuracy, optimal conditions and software settings for the automated forms processing program, Teleform. A second purpose was to determine the feasibility of utilizing Teleform in public health research. The first objective of the study was to determine whether Teleform software provided an accurate methodology for data entry and whether data checking was necessary. In the current study the accuracy of Teleform was dependent on several factors including the type of writing instrument utilized, whether the individual had prior Teleform experience, and what type of data field was completed (constrained print field or choice field). When reviewing the overall results for black pen, the error rate was 10.6% for constrained print fields and 5.3% for choice fields. Interestingly, the error rate for choice fields for those individuals with prior Teleform experience was 0%. Although the overall error rates appear to be somewhat low, previous research in a more structured setting reported error rates under 0.1% for both constrained print fields and choice fields (Smyth et al., 1997; Jorgensen and Karlsmose, 1998). Additionally, research has shown that an error rate of less than 0.1% is attainable.
and acceptable in the research setting (Neaton et al., 1990; Reynold-Heartle et al., 1992; Glassman et al., 1995). Thus, when scanning conditions are similar to those reported in the current study, data checking is necessary.

The second objective of the study was to determine whether there were differences in accuracy due to the kind of writing instrument used to complete forms. In the current study, three different writing instruments were provided to the volunteers in a randomized manner. Overall, the use of black pen resulted in fewer constrained print fields to verify (29.2 black pen, 43.5 blue pen, 42.8 pencil) and correct (17.8 black pen, 27.8 blue pen, 31.7 pencil) than blue pen and pencil. Black pen also resulted in fewer errors than pencil for both constrained print fields (10.6% black pen, 11.4% blue pen, 24.2% pencil) and choice fields (5.3% black pen, 13.1% blue pen, 24.2% pencil) while black pen and blue pen were not significantly different. Thus, when using forms containing both constrained print fields and choice fields, the use of black pen should provide more favorable results.

The third objective of the study was to determine whether there were differences in accuracy due to the type of person completing forms (researcher versus non-researcher). No significant differences were found between the two research groups for number of fields to verify, number of fields to correct, or the amount of error. This was true for the two types of variables. Because there were no differences, a separate analysis compared a small group of individuals with Teleform experience versus those without such experience. Interestingly, those with prior Teleform experience had significantly fewer constrained print fields to verify (19.8 versus 32.6), fewer constrained
print fields to correct (11.0 versus 19.8), as well as, a significantly lower choice field error rate (0% versus 7.6%). These results suggest that prior training may improve the overall scanning performance and the choice field accuracy rate of Teleform. Thus, training and/or educational materials should be provided to individuals on how to complete forms designed in Teleform software.

The final objective of the study was to determine whether Teleform software provides a viable alternative to data entry in the public health field. Although many different data sets have been made accessible to public health researchers, there is still a growing need to collect additional data within the field. Due to lack of resources, an inexpensive, fast, and accurate automated data entry system could greatly enhance the data collection process in public health settings. The cost of Teleform software and the scanner were $3000. When reviewing the literature for time and cost savings associated with scanning technology, the cost of duplicate data entry was 277% more than the cost for Teleform (Jorgensen and Karlsnose, 1988) and duplicate data entry required a 22-fold increase in the amount of time needed for scanning and verification (Smythe et al., 1997). Although, an initial setup and training period is required when first using Teleform, the costs and time savings associated with the system should override any initial inconvenience. Although, the overall error rates found in the current study were higher than anticipated, lower error rates were obtained in those with Teleform experience and in other studies with more structured conditions (Smyth et al., 1997; Jorgensen and Karlsnose, 1998). Thus, the costs and time savings, as well as, the potential for extremely low error make Teleform a feasible and viable alternative to data
entry in the public health field. However, further research should be completed to determine the affects of any additional changes made to enhance Teleform accuracy.

Suggestions for Future Research

The current study monitored the number and types of errors made during data collection. Without such monitoring, the quality of the data would be unknown and conclusions drawn from the research project potentially flawed. Consequently, both clinical research and public health settings should establish methods to monitor and research the quality of data collected. Second, many ideas were suggested in the current study to enhance the overall accuracy and performance of Teleform. These suggestions included: 1) using a black pen to complete forms, 2) training and further educational materials to enhance form completion, 3) reducing the number and length of constrained print fields, 4) using range and consistency checks for individual fields, 5) raising the confidence threshold to 99 or 100%, 6) limiting the type and number of individuals who complete forms, and 7) checking all forms prior to scanning and correcting any noticeable errors. However, further research should document the potential influences of any of these described suggestions.
Appendix A

Data Collection Forms
THE CLAUDE PEPPER
OLDER AMERICANS INDEPENDENCE CENTER
Of the University of Connecticut Health Center
Successful Aging through Research

Demographic Information
Page 1 of 1

Volunteer ID
Study: TELFM

Date: __/__/1991

Acroic

Instrument Type
☐ Pencil
☐ Blue Pen
☐ Black Pen

First Name
Last Name
M/F

1. What is your date of birth? __/__/__ __/__/__ Age:

2. What is your current mailing address?

Street
Apt

City
State
Zip

3. What is your occupation?
Occupation

4. Do you have a computer in your home?
☐ Yes ☐ No ☐ Don’t Know

4a. If so, how many hours per week do you work on your computer at home?

5. Do you have access to a computer at work or your volunteer job?
☐ Yes ☐ No ☐ Don’t Know

5a. If so, how many hours per week do you work on your computer at work?

7. What kind of experience do you have with entering data into a computer?
☐ Much data entry experience ☐ Moderate data entry experience ☐ No data entry experience
☐ Very much data entry experience ☐ Some data entry experience

8. What kind of experience do you have with designing and using forms to collect data?
☐ Much form design experience ☐ Moderate data entry experience ☐ No form design experience
☐ Very much form design experience ☐ Some form design experience

9. Have you ever used a computer program called Teleform before?
☐ Yes ☐ No ☐ Don’t Know

10. Have you ever used a computer scanner before?
☐ Yes ☐ No ☐ Don’t Know

11. How confident do you feel in filling out forms of this nature?
☐ Extremely confident ☐ Very confident ☐ Moderately confident ☐ Somewhat confident ☐ Not at all confident
The Claude Pepper Older Americans Independence Center
Of the University of Connecticut Health Center
Successful Aging through Research

Form 15
NCA
Page 1 of 1

Volunteer ID: _______________________________ Study: ________________
Acro Disc: ________________________________

Ex 3/3/20_________________________ Interviewer: ________________

Volunteer Time Point
☐ screening  ☐ 18 month
☐ baseline  ☐ 21 month
☐ 3 month  ☐ 24 month
☐ 6 month  ☐ 30 month
☐ 9 month  ☐ 36 month
☐ 12 month  ☐ other
☐ 15 month

1a) Do you smoke now, or have you in the past?
☐ yes ☐ no, I stopped ☐ no, I never smoked

If stopped, when? (indicate year) ______________________

1b) If YES, how much do you smoke now? ____________ (cigarettes per day)

2) Thinking about the past month, estimate how many drinks containing alcohol you had on an average week?

NOTE: 1 drink is equal to:

1 (1.5 oz) shot of liquor

OR

1 (5 oz) glass of wine

OR

1 (12 oz) can of beer

During an average week, I have ___________ drink(s).

Have you cut down on your drinking? ☐ yes ☐ no

3) Thinking about the past month, estimate how many drinks containing caffeine you had on an average day.

NOTE: 1 drink is equal to:

1 cup of coffee/tea

OR

2 (12 oz) cans of caffeinated soda

On the average day, I had ___________ drink(s).
Below is a list of things people might do or say to someone who is trying to exercise regularly. If you are not trying to exercise, some of the questions may not apply to you but please read and give an answer to every question. The following questions refer to any exercise that you do.

Please rate each question twice. For family, rate how often anyone living in your household has said or done what is described during the last 3 months. For friends, rate how often your friends, acquaintances, or co-workers have said or done what is described during the last 3 months. Please fill in one appropriate box for each question. The box should be filled as such: [ ] or [ ]

During the past three months, my family (or members of my household) or friends:

<table>
<thead>
<tr>
<th></th>
<th>Not at All</th>
<th>Rarely</th>
<th>A Few Times</th>
<th>Often</th>
<th>Very Often</th>
<th>Does Not Apply</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Exercised with me.</td>
<td>Family</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Gave me encouragement to stick with my exercise program</td>
<td>Family</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Changed their schedule so we could exercise together.</td>
<td>Family</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Offered to exercise with me.</td>
<td>Family</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Gave me helpful reminders to exercise (&quot;Are you going to exercise tonight?&quot;).</td>
<td>Family</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Planned for exercise on recreational outings.</td>
<td>Family</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
During the past three months, my family (or members of my household) or friends:

<table>
<thead>
<tr>
<th></th>
<th>None</th>
<th>Rarely</th>
<th>A Few Times</th>
<th>Often</th>
<th>Very Often</th>
<th>Does Not Apply</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. Discussed exercise with me:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Talked about how much they liked to exercise:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Helped plan activities around my exercise:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Asked me for ideas on how they can get more exercise:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Took over chores so I had more time to exercise:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Made positive comments about my physical appearance:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. In general, would you say your health is:
   - Excellent
   - Good
   - Poor
   - Very good
   - Fair

2. Compared to one year ago, how would you rate your health in general now?
   - Much better now than one year ago
   - Somewhat better now than one year ago
   - About the same as one year ago
   - Somewhat worse now than one year ago
   - Much worse now than one year ago

3. The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

   a. Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports
      - Yes, Limited A Lot
      - Yes, Limited A Little
      - No, Not Limited At All

   b. Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling or playing golf
      - Yes, Limited A Lot
      - Yes, Limited A Little
      - No, Not Limited At All

   c. Lifting or carrying groceries
      - Yes, Limited A Lot
      - Yes, Limited A Little
      - No, Not Limited At All

   d. Climbing several flights of stairs
      - Yes, Limited A Lot
      - Yes, Limited A Little
      - No, Not Limited At All

   e. Climbing one flight of stairs
      - Yes, Limited A Lot
      - Yes, Limited A Little
      - No, Not Limited At All

   f. Bending, kneeling, or stooping
      - Yes, Limited A Lot
      - Yes, Limited A Little
      - No, Not Limited At All

   g. Walking more than a mile
      - Yes, Limited A Lot
      - Yes, Limited A Little
      - No, Not Limited At All

   h. Walking several blocks
      - Yes, Limited A Lot
      - Yes, Limited A Little
      - No, Not Limited At All

   i. Walking one block
      - Yes, Limited A Lot
      - Yes, Limited A Little
      - No, Not Limited At All

   j. Bathing or dressing yourself
      - Yes, Limited A Lot
      - Yes, Limited A Little
      - No, Not Limited At All
4. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?
   a. Cut down on the amount of time you spent on work or other activities
      □ Yes □ No
   b. Accomplished less than you would like
      □ Yes □ No
   c. Were limited in the kind of work or other activities
      □ Yes □ No
   d. Had difficulty performing the work or other activities (for example, it took extra effort)
      □ Yes □ No

5. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?
   a. Cut down on the amount of time you spent on work or other activities
      □ Yes □ No
   b. Accomplished less than you would like
      □ Yes □ No
   c. Didn't do work or other activities as carefully as usual
      □ Yes □ No

6. During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups?
   □ Not at all □ Quite a bit
   □ Slightly □ Extremely
   □ Moderately

7. How much bodily pain have you had during the past 4 weeks?
   □ None □ Moderate
   □ Very mild □ Severe
   □ Mild □ Very severe

8. During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?
   □ None at all □ Quite a bit
   □ A little bit □ Extremely
   □ Moderately

9. These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks:
   a. Did you feel full of pep?
      □ All of the time □ Some of the time
      □ Most of the time □ A little of the time
      □ A good bit of the time □ None of the time
   b. Have you been a very nervous person?
      □ All of the time □ Some of the time
      □ Most of the time □ A little of the time
      □ A good bit of the time □ None of the time
   c. Have you felt so down in the dumps that nothing could cheer you up?
      □ All of the time □ Some of the time
      □ Most of the time □ A little of the time
      □ A good bit of the time □ None of the time
   d. Have you felt calm and peaceful?
      □ All of the time □ Some of the time
      □ Most of the time □ A little of the time
      □ A good bit of the time □ None of the time
<table>
<thead>
<tr>
<th>Form 7</th>
<th>The MOS 36-Item Short-Form Health Survey (SF-36)</th>
<th>Page 3 of 3</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>No. 10. During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)?</th>
</tr>
</thead>
<tbody>
<tr>
<td>None of the time</td>
</tr>
<tr>
<td>A little of the time</td>
</tr>
<tr>
<td>Most of the time</td>
</tr>
<tr>
<td>All of the time</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No. 11. How TRUE or FALSE is each of the following statements for you?</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. I seem to get sick a little easier than other people</td>
</tr>
<tr>
<td>Definitely True</td>
</tr>
<tr>
<td>Mostly True</td>
</tr>
<tr>
<td>Mostly False</td>
</tr>
<tr>
<td>Definitely False</td>
</tr>
<tr>
<td>Don't Know</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>b. I am as healthy as anybody I know?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitely True</td>
</tr>
<tr>
<td>Mostly True</td>
</tr>
<tr>
<td>Mostly False</td>
</tr>
<tr>
<td>Definitely False</td>
</tr>
<tr>
<td>Don't Know</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>c. I expect my health to get worse?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitely True</td>
</tr>
<tr>
<td>Mostly True</td>
</tr>
<tr>
<td>Mostly False</td>
</tr>
<tr>
<td>Definitely False</td>
</tr>
<tr>
<td>Don't Know</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>d. My health is excellent?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitely True</td>
</tr>
<tr>
<td>Mostly True</td>
</tr>
<tr>
<td>Mostly False</td>
</tr>
<tr>
<td>Definitely False</td>
</tr>
<tr>
<td>Don't Know</td>
</tr>
</tbody>
</table>
### Personal Data

1. What is your birthdate?
   - 
   - 
   - 

2. Which of the following best describes your racial background?
   - American Indian or Alaskan Native
   - Asian, Oriental or Pacific Islander
   - Black or African American
   - White or Caucasian
   - Other

3. Are you of Spanish or Hispanic origin or ancestry?
   - No
   - Yes

4. What is the highest grade you completed in school?
   - 8th grade or less
   - Some highschool
   - Highschool graduate
   - Some college
   - College graduate
   - Any post-graduate work

5. Which of the following categories best describes your household’s total income last year before taxes? Please include income from all sources, such as salaries and wages, Social Security, retirement income, investments, and other sources.
   - Less than $20,000
   - $20,000-$39,000
   - $40,000-$59,000
   - $60,000-$79,000
   - $80,000 or more

### Lifestyle

6. How would you describe your cigarette smoking habits?
   - Currently smoke (go to question 7)
   - Used to smoke (go to question 8)
   - Never smoked (go to question 9)

7. (Currently smoke)
   a. How many cigarettes per day?
   - 
   - 
   -
   b. How many years you smoked?
   - 
   - 
   -
### Health History Form

#### page 2 of 5

8. (Used to smoke)  
   a. How many years has it been since you smoked on a fairly regular basis?  
      ![Blank space for years]  
      ![Blank space for months]
   
   b. During the last two years that you smoked, how many cigarettes did you smoke per day, on average?  
      ![Blank space for cigarettes per day]

9. How many years have you been exposed to secondhand smoke (that is, smoke from people around you) in the home or workplace? (If none enter ‘0’).  
   ![Blank space for years]

10. How many alcoholic beverages do you consume in a typical week? Write the number of each type that you drink.  
   a. bottles or cans of beer (if none, enter ‘0’)
   ![Blank space]
   b. glasses of wine (if none, enter ‘0’)
   ![Blank space]
   c. wine coolers (if none, enter ‘0’)
   ![Blank space]
   d. mixed drinks or shots of liquor (if none enter ‘0’)
   ![Blank space]

12. During the past year, how many times per week did you usually eat/drink the following foods? (enter ‘0’ if you did not usually eat a food at least 1 time per week)

   a. cottage cheese
   ![Blank space]
   b. a glass of milk
   ![Blank space]
   c. yogurt, frozen yogurt, or ice cream
   ![Blank space]
   d. cereal (hot or cold) with milk
   ![Blank space]
   e. an egg
   ![Blank space]
   f. canned fish with bones, such as mackerel, sardines, or salmon
   ![Blank space]
   g. cheese or cheese dishes such as macaroni and cheese
   ![Blank space]
   h. meat (red or white), poultry or fish
   ![Blank space]
   i. calcium-enriched orange juice
   ![Blank space]
   j. unenriched orange or grapefruit juice
   ![Blank space]
   k. green salad
   ![Blank space]
   l. mustard greens, turnip greens, kale, or collards
   ![Blank space]
   m. beans, such as pinto beans, kidney beans, baked beans, or black eyed peas
   ![Blank space]
   n. caffeinated drinks, such as coffee, tea, or soft drinks
   ![Blank space]

11. Have you ever been a heavy drinker or treated for an alcohol problem?  
   No [ ]  
   Yes [ ]
13. For each of the following (a-f), select the category that best describes your level of physical activity in the past year. (Some examples are given in parentheses)

<table>
<thead>
<tr>
<th>a. Physical activity you get from your occupation or physical routine</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ None</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>□ None</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>□ None</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>□ None</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>□ None</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>□ None</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

b. Amount you walk each day including to and from your car, the bus, or the train

| □ None | □ Mild (less than 2 blocks) | □ Moderate (more than 3 blocks) | □ Strenuous (more than three miles) |
| | | | |

c. Amount of housework you do in a week

| □ None | □ Mild (light dusting) | □ Moderate (dusting, vacuuming) | □ Strenuous (heavy cleaning) |
| | | | |
| □ None | □ Mild | □ Moderate | □ Strenuous |
| | | | |
| □ None | □ Mild (walking) | □ Moderate (aerobics or swimming about 3 times/week) | □ Strenuous (jogging more than 3 times/week) |
| | | | |
| □ None | □ Mild (golf) | □ Moderate (tennis-doubles, volleyball) | □ Strenuous (tennis-singles, racquetball) |
| | | | |

d. Amount of your own yardwork you do in a week

| □ None | □ Mild | □ Moderate | □ Strenuous |
| | | | |
| | | | |
| □ None | □ Mild | □ Moderate | □ Strenuous |
| | | | |
| □ None | □ Mild (walking) | □ Moderate (aerobics or swimming about 3 times/week) | □ Strenuous (jogging more than 3 times/week) |
| | | | |
| □ None | □ Mild (golf) | □ Moderate (tennis-doubles, volleyball) | □ Strenuous (tennis-singles, racquetball) |
| | | | |

14. For each of the following members of your family, indicate whether they have had angina, a heart attack, heart failure, or any other symptom of heart disease:

| Mother | □ No | □ Yes | □ Don't Know | □ Not Applicable |
| | | | | |
| Father | □ No | □ Yes | □ Don't Know | □ Not Applicable |
| | | | | |
| Sister | □ No | □ Yes | □ Don't Know | □ Not Applicable |
| | | | | |
| Brother | □ No | □ Yes | □ Don't Know | □ Not Applicable |
| | | | | |
| Daughter | □ No | □ Yes | □ Don't Know | □ Not Applicable |
| | | | | |
| Son | □ No | □ Yes | □ Don't Know | □ Not Applicable |
| | | | | |

15. For each of the following members of your family, indicate whether they have had any signs of osteoporosis (such as broken bones from minor injuries, hip fracture, marked obvious rounding of shoulders, loss of adult height):

| Grandmother | □ No | □ Yes | □ Don't Know | □ Not Applicable |
| | | | | |
| Grandfather | □ No | □ Yes | □ Don't Know | □ Not Applicable |
| | | | | |
| Mother | □ No | □ Yes | □ Don't Know | □ Not Applicable |
| | | | | |
| Father | □ No | □ Yes | □ Don't Know | □ Not Applicable |
| | | | | |
| Sister | □ No | □ Yes | □ Don't Know | □ Not Applicable |
| | | | | |
| Brother | □ No | □ Yes | □ Don't Know | □ Not Applicable |
| | | | | |
| Daughter | □ No | □ Yes | □ Don't Know | □ Not Applicable |
| | | | | |
| Son | □ No | □ Yes | □ Don't Know | □ Not Applicable |
# Health History Form

## page 4 of 5

16. During the past 4 weeks, how would you describe any back pain or leg pain (sometimes called sciatica) you have had when at rest?

- [ ] Very Mild
- [ ] Severe
- [ ] Mild
- [ ] Very Severe
- [ ] Moderate

17. During the past 4 weeks, how many days did you cut down on things you usually do because of back pain or leg pain (sometimes called sciatica)?

<table>
<thead>
<tr>
<th>Number of days</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ] 1</td>
</tr>
<tr>
<td>[ ] 2</td>
</tr>
<tr>
<td>[ ] 3</td>
</tr>
<tr>
<td>[ ] 4</td>
</tr>
<tr>
<td>[ ] 5 or more</td>
</tr>
</tbody>
</table>

If you are 18 years old or younger, skip to question 28.

18. After age 18, how many times have you broken or fractured each of the following?

<table>
<thead>
<tr>
<th>Bone</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4 or more</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Hip</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>b. Wrist</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>c. Back</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>d. Other bones</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
</tbody>
</table>

(please specify)

19. During the past 4 weeks, how would you describe any aches in your joints you have had?

- [ ] Very Mild
- [ ] Mild
- [ ] Moderate
- [ ] Severe
- [ ] Very Severe
- [ ] I have not had any aches in my joints

20. Have you ever had or been treated for and eating disorder (such as anorexia or bulimia)?

- [ ] No
- [ ] Yes, currently

(If yes, indicate the year in which the disorder ended)

[ ] [ ] [ ] [ ] year

(If you are 21 years old or younger, skip to question 28.)

21. How tall were you when you were 21 years old?

[ ] feet [ ] inches

22. How much did you weigh when you were 21 years old?

[ ] pounds

23. How would you describe your body frame?

- [ ] Small
- [ ] Medium
- [ ] Large

24. List all non-prescribed medications you are currently taking. (Include medicines such as vitamins or multivitamins, calcium supplements, iron, aspirin, antacids, laxatives, etc.)

a. ____________________ e. ____________________

b. ____________________ f. ____________________

c. ____________________ g. ____________________

d. ____________________ h. ____________________

## Health History Form

### page 5 of 5

25. Have you ever taken any of the following medications for more than six months?

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>Steroid inhaler</td>
<td>□ No □ Yes</td>
</tr>
<tr>
<td>b.</td>
<td>Oral steroids, such as prednisone, hydrocortisone, Medrol, Decadron</td>
<td>□ No □ Yes</td>
</tr>
<tr>
<td>c.</td>
<td>Thyroid hormone, such as Prolact, S-P-T, Cytomel, Levoxine, Synthroid, Ethroid</td>
<td>□ No □ Yes</td>
</tr>
<tr>
<td>d.</td>
<td>Medications for epilepsy, such as Dilantin, Tegretol, phenobarbital</td>
<td>□ No □ Yes</td>
</tr>
<tr>
<td>e.</td>
<td>Water pills (diuretics), such as hydrochlorothiazide, Lasix, Dyazide, Duril</td>
<td>□ No □ Yes</td>
</tr>
<tr>
<td>f.</td>
<td>Birth control pills, such as Ortho-Novum, Triphasil, Lo/Ovral</td>
<td>□ No □ Yes</td>
</tr>
<tr>
<td>g.</td>
<td>Other hormones, such as testosterone, calcitonin, tamoxifen</td>
<td>□ No □ Yes</td>
</tr>
<tr>
<td>h.</td>
<td>Fluoride, Didronel (etidronate)</td>
<td>□ No □ Yes</td>
</tr>
</tbody>
</table>

26. Are you currently or have you ever taken or used estrogens (pills, patches, vaginal creams, injections, implants), such as Primarin, Estrace, Estraderm or Ogen?

- □ No
- □ Yes, I am currently taking estrogen
- □ Yes, but I do not take them now
REFERENCES


National Archive of Computerized Data on Aging. (1998). *Data Collections from the National Archive of Computerized Data on Aging (NACDA)*. Inter-university Consortium for Political and Social Research, Ann Arbor, MI.


