Methodology Report

Draft version: 7/21/2004

SHAPE 2002

SHAPE 2002: Methodology Report

This report was prepared by:
Urban Landreman, M.S., M.B.A.,
Hennepin County Community
Health Department

Yingmei Ding, M.D., M.S., Hennepin County
Community Health Department

Sheldon Swaney, B.A., Hennepin County
Community Health Department

Timothy Zimmerman B.A., Hennepin County
Community Health Department

Advisory Group

Suggested Citation

SHAPE 2002 Project Staff

Project Director: Margaret Hargreaves,
Hennepin County Community Health
Department

Project Leadership Team
Todd Monson, Mark Lee, Jodell O'Connell
and Margaret Hargreaves, Hennepin County
Community Health Department; David Doth,
Gretchen Musicant and Pat Harrison,
Minneapolis Department of Health and Family
Support; and Karen Zeleznak, Bloomington
Division of Public Health

Project Task Leaders
George Bowlin, Michelle Chiezah, Yingmei
Ding, Urban Landreman, Sheldon Swaney
and Timothy Zimmerman, Hennepin County
Community Health Department; and Elizabeth
Songalia, Bloomington Division of Public
Health

Survey Firm
Todd Rockwood, Colleen King and Keith
Onken, Health Services Research and Policy
Survey Center, University of Minnesota
School of Public Health

Project Team Members
George Bowlin, Michelle Chiezah, Yingmei
Ding, Peter Fish, Suzanne Gaines, Gayle
Geber, Margaret Holm, Urban Landreman,
Diane Loeffler, Sheldon Swaney, Jessica
Trujillo, Gizaw Tsehai and Timothy
Zimmerman, Hennepin County Community
Health Department; Patty Bowler, Megan
Ellingson, Pat Harrison, David Johnson, Gopal
Narayanan, David Rak and Abbey
Sidebottom, Minneapolis Department of
Health and Family Support; Eileen O'Connell
and Elizabeth Songalia, Bloomington Division
of Public Health; Michelle Anderson and
Nancy Leegard, Hennepin County
Public Affairs; Noemi Alcocer, Zoua Chang
and Mahdi Jama, Hennepin County
Multicultural Center; Ann Kinney, Minnesota
Department of Health; Michael Finch,
UnitedHealthcare; Michael Davern, SHADAC
Center, University of Minnesota; Assata
Brown, HER; and Kathy Graves, Parenteau
Graves

Acknowledgements

SHAPE 2002: Methodology Report
Table of Contents

SHAPE 2002: Methodology Report
Introduction

There is a saying often used in reference to large projects that almost everyone agrees that the concept is wonderful, but that "the devil is in the details". It is a saying that seems quite appropos for the Survey on the Health of Adults, the Population, and the Environment (SHAPE) 2002 project.

SHAPE is an important and ongoing health surveillance project that monitors the health of adults in Hennepin County. The project uses a holistic perspective by taking a comprehensive look at the health status, health behaviors and health care practices of residents, as well as other social and environmental factors that influence health. The survey has been conducted twice — first in 1998 and again in 2002.

SHAPE 2002

In 2002, the Hennepin County Community Health Department, in collaboration with the Minneapolis Department of Health and Family Support and the Bloomington Division of Public Health, conducted the SHAPE 2002 survey. Data were collected through interviews conducted by the Survey Research Center at the Division of Health Services Research and Policy, part of the University of Minnesota's School of Public Health. The SHAPE 2002 survey results build on the information collected in the 1998 survey, identify health changes and maintain the project's countywide and community-level focus. The SHAPE 2002 data enable local-level analyses of key health indicators, providing public health officials and policymakers an opportunity to identify emerging issues. Local officials can also use the data to evaluate the effectiveness of population health initiatives.

The project began with conversations in 2000 and the final data was collected in late summer of 2002. In between, a large number of local and national experts were consulted, options were considered, and choices were made.

This report documents the methods used at each step of the project. This documentation is important so that the steps taken by the many people involved with the project are not forgotten, but also so that others can learn from our experience.

Project Organization

The SHAPE 2002 project set up a number of work teams to handle the multitude of tasks needed to accomplish its goals. Some of the teams are as follows:

1. **Project Leadership Team**. The Project Leadership Team handled all the top-level administrative tasks. These tasks included, in part, funding, policy development, coordination among all the participating agencies, and media relations.

2. **Community Involvement Work Group**. The Community Involvement Work Group networked with community organizations, groups, and individuals to solicit their input in the development and implementation of the survey and dissemination of the survey results.

3. **Survey Content Work Group.** The Survey Content Work Group developed the actual content of the survey. The group conducted literature reviews, reviewed existing surveys, sought input from local and national topic-specific experts, and decided on the final wording of the questions.

4. **Marketing Work Group.** The Marketing Work Group developed materials and programs to market the SHAPE survey to the general public and to specific subpopulations within the county.

5. **Methodology Work Group.** The Methodology Work Group examined Census and population data to develop the sampling plan for the survey. The group also spent much time examining the phone numbers acquired for the survey to ensure optimal coverage.

6. **Data Coding Work Group**. The Data Coding Work Group reviewed the answers given to open-ended questions and developed coding strategies to convert string data into numberic codes so those answers could be analyzed in future work.

7. **Data Cleaning Work Group.** The Data Cleaning Work Group reviewed the raw dataset which came from the Survey Research Center and developed rules for which which variables needed to be modified, which new variables needed to be created, and which cases needed to be dropped. The resulting file became the SHAPE 2002 research database.

8. **Weighting Work Group.** The Weighting Work Group developed the mathematical procedures to compute the appropriate weights to use for various types of analyses performed by researchers.

The steps taken to acquire and prepare the SHAPE 2002 dataset, after getting the crucial political and financial support to move forward, were not strictly sequential, but for simplicity sake are listed sequentially in this report. These steps can be listed as the following:

1) **Questionnaire Design and Content**. Decide what questions to ask and the order in which to ask them.

2) **Request for Proposal.** Select the organization which would contact adults in Hennepin County for the survey.

3) **Sampling Design.** Decide how to develop a sampling plan to contact enough adults to satisfy all the geographic, racial and ethnic group, and other requirements of the project in a way to keep the costs within budget.

4) **Marketing Efforts.** Use formal and infomal means to promote the survey in communities throughout Hennepin County to increase the probability that people will agree to participate in the survey and that people will value the results once the data are available.

5) **Data Collection.** Contact adults in Hennepin County either by phone or other methods to obtain the information wanted in the survey.

6) **Completion Status.** Compute and analyze how the response and participation rates varied across geographic areas.

7) **Management of SHAPE Data.** Explore the raw data to see which variables should be recoded based on the value of other variables, which cases should be excluded from the final dataset used for analysis, and which new variables to create for use in subsequent analyses.

8) **Weighting the Data.** Perform the computations to assign weights to each case so that the results of the survey can be representative of various geographic areas within the county or of racial or ethnic populations within the county.

9) **Public-Use File**. Prepare a file consisting of a subset of the variables for all the cases so that outside researchers can use the SHAPE data for projects of interest and value to the larger community.

10) **Lessons Learned.** Take time to reflect on and evaluate the process used so that future efforts include steps that ended up being valuable, do not repeat mistakes, and are of a higher quality.

Detailed descriptions of each of these steps are included in the following sections of this document along with many of the underlying documents, spreadsheets, and maps which were used at numerous points in the project.

It is hoped that by documenting the steps taken during SHAPE 2002, others can benefit from our experience and improve future similar surveillance efforts either within Hennepin County or elsewhere.

1. QUESTIONNAIRE DESIGN AND CONTENT

The Survey of the Health of Adults, the Population, and the Environment (SHAPE) is an ongoing public health surveillance and assessment project. The overall objectives for SHAPE 2002 are two folds: first, to assess the health level and trends among Hennepin County adults, at, not only the countywide level, but also at the community level, as well as for major racial and ethnic populations; second, to understand the factors and conditions that relate to the health of county residents, by geographic area, by race and ethnicity and by socio-economic status. High priority is given to collecting data items which community members have expressed to be of particular interest or usefulness to them.

To make sure that the survey questionnaire was designed to reflect SHAPE's purpose, serve the survey's objectives and meet community data needs, the SHAPE 2002 Project Team undertook the following six stages of questionnaire development activities:
- Preparation
- Community consultation
- Survey questionnaire construction
- Review and field test
- Finalization
- Translation

The SHAPE 1998 questionnaire served as the basis for development of the SHAPE 2002 questionnaire. In order to monitor the trend of major health indicators, most 1998 questions were either kept or modified. Inputs gathered through consultation guided which questions from SHAPE 1998 to keep, to modify, and to drop, and which questions to add to reflect the emerging data needs of community members, especially the data needed by racial and ethnic communities and groups.

1.1. Preparation

Establishing working group and develop work plan

The SHAPE 2002 Survey Content Work Group as well as the Community Involvement Work Group were created. Each of the groups consists of staff from the collaborating health departments: Hennepin County Community Health Department, Minneapolis Department of Health and Family Support and Bloomington Division of Public Health. Ann Kinney, a consultant from Minnesota Department of Health, and Michael Finch, a contracted consultation from UnitedHealth, were also members of each group. Members of work groups conducted research, gathered resource materials, consulted with communities, and met regularly to discuss survey items, review and evaluate community inputs, different resource materials and survey tools for the development of the SHAPE 2002 questionnaire.

A work plan for community consultation and survey content development (Appendix F) was developed. It outlined the purpose, scope, process, work products, working structure as well as proposed time line. Each stage of activities and products was reported to the SHAPE 2002 Leadership Team for review, and direction.

Guiding Principles For Survey Item Selection

A set of guiding principles was established for evaluating survey questions and selecting items. The principles were used to design a survey that served its purpose, and to guide the cutting process - trim *the survey we would like*, a survey that contains all the important and needed survey items, but that takes hours to complete, to *a survey we can do*, a 30-minute telephone survey.

Guiding principles for general survey questions:
The questions need to
- Measure a health issue or concern of public health significance;
- Serve the purpose of the survey objectives;

- Be measureable across the population;
- Be suitable to be gathered via a telephone interview;
- Be useful for planning, programming and policy development;
- Be not currently available at the local level; and
- Have benchmark data available.

Guiding principles for survey questions that are needed by racial and ethnic communities:
The questions need to
- Measure health conditions or factors related to health that are of particular concern to the racial and ethnic groups;
- Be useful for planning, programming or policy development initiated by either of the community group, city, county or the state on behalf of the racial and ethnic group; and
- Be related to the priorities in reducing the health disparity of racial and ethnic groups at the city, county or state level.

Guiding principles for finding the proper balance of survey content that reflect broad public health perspective:
The survey need to
- Include questions that address health issues at both the individual level, as well as at the structural or socio-environmental level
- Include questions that address all major domains health with broad public health perspective, including the following:
    - Overall physical and mental health
    - Health access and utilization
    - Community support and social environment factors
    - Healthy lifestyles and behaviors
    - Demographic

Generating consultation materials

To facilitate community consultation and involvement, the Community Involvement Work Group developed a set of materials including the following:
- SHAPE II Project Overview (Appendix G)
  A fact sheet that covers the most commonly asked questions about the project. The fact sheet was made available on the SHAPE 2002 webside and was included in the consultation package and was is the handout at consultation meetings;
- Guiding Principles for SHAPE II Survey Item Inclusion (Appendix H)
  The guiding principles as described above, along with key background information and references for SHAPE 2002's objectives, definitions of public health and public health surveillance, and some principles of public health were presented in a fact sheet format to be included in the consultation package or during consultation meetings;
- SHAPE II Survey Contents – Proposed Topics and Sample Question (Appendix I)
  The single sheet document arranged the proposed survey content, based on SHAPE 1998 into major health domains and topics. Only sample question under each survey topic were included. This sheet was used to provide a very quick overview of SHAPE 2002's scope. It was used mostly in consultation with local officials and policy makers, and with racial and cultural communities;
- SHAPE II Survey Content – Topic areas and Proposed Questions (Appendix J)
  The six page document lists all the SHAPE 1998 survey questions by six topic areas. It was used for the topic-specific consultations.

1.2. Community Consultation

Community involvement was essential in order that SHAPE 2002 collected and reported data that reflects the needs and interests of local communities and racial and ethnic populations, and that the survey results will be actually be of value and used for policy and program development. Partnerships with communities that are established or enhanced through the project will serve as a foundation for continued cooperation between the County's community health agencies and local communities, and should lead to improving the health status of all residents with reduced health disparities among county residents.

Comprehensive consultations with project stakeholders and communities were made to reach the following objectives:
- The data collected will answer the research questions and survey objectives,
- The primary health data needs among stakeholders and community partners, especially cultural groups, are met,
- The cultural groups reach consensus on primary health data needs
- The data collected is valid,
- Survey questions are delivered in culturally appropriate and sensitive manner,
- Communities get involved in the data collection, dissemination and utilization, and
- The community health departments develop and enhance partnerships with community groups and organizations.

Five consultation components were planned. The following sections summarize the activities that were undertaken for each of the five consultation components.

1.2.1 Consultation with local public health agencies

Leadership and staff of SHAPE 2002's primary stakeholders, the three health departments who sponsored the survey, including Hennepin County Community Health Department, Minneapolis Department of Health and Family Support and Bloomington Division of Public Health, were among those first sought for input. They were invited to evaluate whether or not any of the SHAPE 1998 questions should be asked again in SHAPE 2002, how 1998 question should be revised, and what new data was needed. A consultation package (Appendix K) was prepared.

The feedback from Hennepin County Management Team and staff is summarized in Appendix L. Communication e-mails on inputs from Minneapolis Department of Health and Family Support, Bloomington Division of Public Health are kept in file.   All inputs were all carefully reviewed and discussed by the Content Work Group in the respective survey topic specific review meetings.

1.2.2. Consultation with local official and policy makers

The support and commitment from local official and policy makers for SHAPE 2002 were vital to make the project successful. Consultations with health boards, community health service advisory committees, and citizen advisory committees were made via presentations at their regular meeting. The presentation and handout includes:
- SHAPE 2002 project overview fact sheet (Appendix G)
- A 10-20 minutes powerpoint presentation on SHAPE project (Appendix M)
- SHAPE 2002 survey content proposed topics and SAMPLE questions (Appendix I).

Committees were invited for input for many aspects of the survey (Appendix N), including inputs on each of the proposed survey topic areas. Furthermore, they were asked specific questions about their roles as a committee, such as: Will the data from SHAPE 2002 be useful to you as a committee? Will the data from SHAPE 2002 be useful to people or organizations that you serve?

The following city and county leadership groups were visited by SHAPE 2002 project team members for consultation:
- Minneapolis Public Health Advisory Committee
- Minneapolis Latino Advisory Committee

- Minneapolis Senior Citizen Advisory Committee
- Advisory Board of Health, City of Edina
- Advisory Board of Health, City of Richfield
- Advisory Board of Health, City of Bloomington
- Hennepin County Board /Management Team (???)
- Hennepin County Community Health Service Advisory Committee
- North Minneapolis Health Advisory Committee
- Hennepin County Northwest Human Service Council

The inputs across these consultations had common themes (see Appendix O and P).   Committee members would love to see the demand for health data from cultural communities to be addressed by SHAPE 2002 and would like to see some action follow.

1.2.3. Consultation with racial and ethnic communities

To consult with the racial and ethnic communities in Hennepin County, a two-phase plan was conceived. The first was the consultation interview with racial and ethnic communities, to identify health issues and data needs that are unique to their communities. The second part was holding focus groups to take information gained from the consultation interviews and reach consensus on the top three health issues for each racial and ethnic groups to include in the survey.

Assata Brown, representing her company, HER, A Personal Development Company, was selected to conduct two phases of racial and ethnic outreach for the project.

Consultation Interviews

The primary consultation interview was to identify priority data needs from each of the community organizations representing racial and ethnic communities. The process was also designed to learn from communities how to get adults from those racial and ethnic groups to participate in the survey, to determine the best methods to conduct in-person survey, and to solicit advice on the ways to promote the findings of the survey. Interview questions were developed to meet these objectives (Appendix Q).

The Community Involvement Work Group worked with racial and ethnic group key contact persons to identify community organizations and individuals as candidates to be invited for consultation interview. A contact list containing the names of 100 organizations and individuals that represent the following racial ethnic groups was developed (Appendix R):
- African American (U.S.-born Blacks)
- African-born Blacks
- American Indians
- Hispanics /Latinos
- Southeast Asian
- Cross-cultural organizations

Assata Brown made the first selection from the list of contacts for each racial and ethnic groups for interviews.   Calls were made to each organization in an effort to explain the project and ask for an interview at their most convenient time and location. Messages were left on voice mail system whenever no one is available. If organization was not reached after three calls were placed at different time, that organization was removed from consideration and another organization was selected.

For those organizations that agreed to be interviewed, a letter was sent (Appendix S) which described the project, highlighted the importance of community involvement, described the purpose of the interview, and confirmed the scheduled interview appointment.

For most of the consultation interviews, a staff from one of the three health departments would present along with Assata Brown.   The health department staff started the introduction (Appendix T), and Assata Brown conducted the interview.   A minimum of five organizations from each of identified racial and ethnic group, and cross-cultural group were interviewed.   A total of 34 interviews were conducted between May 15 and June 8, 2001.   Appendix U lists the organizations that participated in interviews.

Assata Brown reported the process and results of the consultation interviews in the following reports:

1. SHAPE II Racial and Ethnic Community Outreach: Interviews and Focus Groups Report. Hennepin County Community Health. August 2001. Prepared by HER, A Personal Development Company

2. SHAPE II Racial and Ethnic Community Outreach: Interview Notes. Hennepin County Community Health. August 2001. Prepared by HER, A Personal Development Company

3. SHAPE II Racial and Ethnic Community Outreach: Quotes to Use to Formulate Questions. Hennepin County Community Health. August 2001. Prepared by HER, A Personal Development Company

The priority health issues and data needs identified by communities are summarized in the spread sheet "Priority Health Issues and Data Needs of Racial /Ethnic Communities" (Appendix V).

Focus Groups

The list of major health issues from each consultation interview were summarized by five racial and ethnic groups. Five structured focus (mini) groups were scheduled, one for each of the racial and ethnic groups. The Primary purpose of the focus groups was to:
- Reach consensus on three issues to include in the survey; and
- Determine the kind of information to collect and types of question to ask

Other purposes of the focus groups included the following:
- Decide the best way to reach people by phone; and
- Decide if a donation to a local charity would increase the participation of adults in that racial or ethnic group.

The structured interview questions were developed to meet these purposes (Appendix W).

Strategies to recruit focus/mini groups included the following:
- Send a letter to all organizations that expressed willingness to participate in a focus group at consultation interviews, and ask them to call and confirm their attendance (Appendix X);
- Send a letter to the remaining organizations from the racial and ethnic community contact list, explaining SHAPE, inviting them to a focus group, and asking them to call and confirm their attendance (Appendix Y); and
- Make follow-up calls to all organizations to invite them to participate in a focus group.

Four focus/mini focus groups were conducted, one for each of the African American, Southeast Asian, American Indian, and Hispanic and Latino communities, between June 18, 2001 and June 22, 2001. A total of 19 people participated in the focus/mini groups. The focus group for African -born Black was not held due to lack of participants.

All focus/mini groups sessions were led by Assata Brown from HER and were audio-taped.   The sessions were hold during normal business hours and at locations that were close to the communities,

including North Community YMCA, Centro Cultural Chicano, and Minnesota Indian Women's Resource Center.   The names of the participants and organization they represented were held in confidence.

The process and results of the focus/mini groups were reported in in Assata Brown's reports listed above. The top three health issues and sample questions are summarized in the spread sheet "Priority Health Issues and Data Needs of Racial /Ethnic Communities (Appendix V).

1.2.4.   Survey content topic specific consultation

A comprehensive consultation with health professionals, community-based organizations and groups, faith communities, researchers, government agencies on SHAPE 2002 content were also made.   This component was primarily accomplished through group consultation meetings.

To facilitate the group consultations, survey topic areas were divided into six areas. The SHAPE 1998 questions were serve as the basis for review and discussion. The six topic areas are the following:
Topic area 1. Community support and social environment
Topic area 2. Nutrition, body weight and physical activity
Topic area 3. Alcohol, tobacco use
Topic area 4. Community safety and violence
Topic area 5. Health access and utilization
Topic area 6. Physical and mental health

Six consultation meetings, one for each of above topics, were scheduled (Appendix Z).   A contact list was developed to invite individuals or organizations for each of the topic areas. Individuals or organizations that have broader health focus are also listed (Appendix A1).

A little over 100 individuals or organizations were invited for consultation meetings by mails and/or e-mails (Appendices B1, C1). Calls were made to confirm the attendance.   The invitation package include an invitation letter, SHAPE II Project Overview, SHAPE II proposed survey topics and questions, consultation meeting schedule, and consultation questions.

Some consultation questions (Appendix D1) were developed to meet the following purposes:
- Evaluate SHAPE 1998 questions to identify any items to improve or eliminate and identify new data needs;
- Assess how communities have used SHAPE 1998, and how SHAPE 2002 data are going to be used; and
- Seek advice on how the project team can assist using SHAPE 2002 data.

Among those invited, about 60 persons attended the consultation meetings. Many suggestions and comments were gathered and summary notes were prepared (Appendices E1 to J1). The Survey Content Work Group reviewed each of the suggestions and comments.

For those who could not attend the consultation meetings, other ways of input were encouraged, such as in writing, via phone or e-mail or a scheduled office visit.   At least 10 individuals offered input via these other means.   Meg Hargreaves, on behalf of the SHAPE project, sent a thank you letter to all persons who participated in any of the sessions (Appendix K1).

1.2. 5.   Technical consultation with local and national experts on survey methods and special survey content question construction

Besides those who were contacted for survey content topic-specific consultations, many local and national experts were contacted for advice, suggestions and resources on specific health issues.   The Survey Content Work Group contacted national experts from CDC, Harvard School of Public Health, the Seattle and King County Public Health Department, and the University of Michigan's Institute of Social

Research for discrimination questions and the International Quality of Live Assessment, University of California for SF12 and Health Related Quality of Life scale.   Local experts on dental health, long term care, nutrition, and young adults issues were also contacted.   The list of persons contacted is shown in Appendix L1.

1.3. Survey questionnaire construction

The development of SHAPE 2002 survey questions took several stages to accomplish. It took nine months (April 10, 2001 to Nov. 14, 2001) of weekly or biweekly meetings to get an initial draft of the survey questions ready for community members to review.   Although the original guiding principles for SHAPE 2002 (Appendix H) directed and steered the process, these criteria alone did not lead to a survey of an appropriate length of 30 minutes or fewer. Other strategies were also used to trim the survey's length.

Fresh start

The Survey Content Work Group reviewed the SHAPE 1998 questionnaire, independent from community input, to evaluate each survey item and decide whether to either keep, modify, or eliminate that item. New data needed were also identified.

Evaluate community inputs from five consultation components

The inputs from racial and ethnic communities were reviewed specifically by the Survey Content Work Group.   The review discussions were reflected in the attached " Priority Health Issues and Data Needs of Racial /Ethnic Communities" (Appendix V).

After obtaining input from all the racial and ethnic communities, the inputs from all other consultation components along with the review results from the cultural groups were consolidated into the topic-specific areas listed below:
Topic area 1. Community support and social environment
Topic area 2. Nutrition, body weight and physical activity
Topic area 3. Alcohol, tobacco use
Topic area 4. Community safety and violence
Topic area 5. Health access and utilization
Topic area 6. Physical and mental health
Topic area 7. Demographic variables
Specific topic: Responsible sexual behavior

For each of the above topic areas, the Survey Content Work Group discussed each survey item, searched for other available survey tools, decided whether to include or modify that item, and deliberated whether the construction of our own survey question was needed.

Many local and national resources and survey tools were reviewed in seeking existing and pre-validated survey questions.   Efforts were also made to make sure that key health indicators were kept the same to monitor trends, and that data were gathered the same as state and national benchmark data did.

The last version of review sheets for each of the topic areas are listed in Appendix N1 to T1. Appendix U1 lists the major survey tools and other resources that were reviewed.

All survey question matrix-228 questions

In early September 2001, the Survey Content Work Group finished the discussion, review, and selection of topic-specific questions. All survey questions that were recommended for consideration for SHAPE 2002 were pulled into one file – SHAPE 2002 all question matrix dated 9/11/2001 (Appendix V1). The

matrix contains 228 survey questions (not counting many sub-set questions) and was far beyond what a 30-minites survey could accomplish.

Cutting the survey length

Cutting the survey's length was a very challenging process since every single survey item in the all question matrix was viewed as valuable based on community input from all five consultation components, document and resource tool reviews, and many discussions and evaluations. However, a test 30-minute interview got only halfway through the 228 questions.   The Survey Content Work Group had to do this most unpleasant job- cutting survey questions.

a) SHAPE 2002 Survey Item Selection Guiding Principles

The cutting process was guided under the survey item inclusion guiding principles (Appendix H). However, the significant data needs that are of public health significance are far beyond what a 30-minute telephone survey can accommodate.   Other strategies had to be employed to lead the challenging process.

b) The most-needed data vs. nice-to-have data

To further cuts the survey length, the work group took a new strategy – identifying the most needed data driven by proposed SHAPE 2002 reports.

Based on the consultation from the Project Leadership Team, the following reports from SHAPE data were expected:
- health report(s) on baby boomers and seniors,
- social conditions and health (racial/ethnic health disparities),
- mental health,
- alcohol and tobacco young, young adult health,
- physical activity, nutrition and body weight, and
- an overall comparison report

Members of the Survey Content Work Group were assigned to one or more teams to develop a hypothetical abstract for each of the above report. The abstract included a potential title, hypothesis, related research questions, data domain, and key data element needed for the report.   This exercise aimed at identifying those most-needed data, rather nice-to-know data. The abstract reports are kept in file. The most-needed data identified from this excise were summarized into the all survey question matrix (Appendix V1).

c). The collaborating health departments selecting 120 questions among the 228 questions that of their priority

Survey Content Work Group members of the three health departments were asked to identify the top 120 out of the 228 questions from the all survey question matrix. Then, each department's pick of 120 questions were tallied. The differences in the items selected were discussed and a consensus was reached as to what questions to keep in the first version of the survey questionnaire. (Appendix W1)

First version of questionnaire

Through a huge amount of work on cutting the survey's length and adjusting the balance of the content areas, community priority data needs, and the collaborating department priority data needs, the first version of SHAPE 2002 questionnaire was achieved in the mid of November 2001.

1.4 Review and Field Test the Questionnaire

At the end of September 2001, the first version of the SHAPE 2002 questionnaire was send to communities for review. The groups included the following:
- Individuals or organizations who represent cultural and racial/ethnic communities, especially those who participated in racial/cultural interviews or focus groups;
- At least 30 individuals or organizations who were invited to topic-specific consultation meetings and expressed an interested in reviewing the survey questionnaire;
- Some local experts; and
- Divisions of the Hennepin County Community Health Department and other departments in Hennepin County; and divisions and groups with the Minneapolis Department of Health and Family Support and Bloomington Division of Public Health.

Specific questions for the reviewers were:
> Do any of these questions need to be modified or improved? If so, how?
> Do you know, if there is any better questions, better instrument or scales to collect these data?
> Do you have any additional comments or suggestions?

The review comments were gathered and reviewed survey content work group to revise survey questions to a field test version (Appendix X1).

Staff from the Survey Research Center implemented the field test version of the questionnaire among adults of various cultural and racial and ethnic background. Surprisingly, it was found that the average length of time to complete the survey was under 30 minutes, so a few additional questions could be put back into the survey. Some other editing and ….. were made.

The final version of SHAPE 2002 questionnaire (Appendix A) was achieved on January…. 2002.

1.5. Final SHAPE 2002 questionnaire

The final survey questions

Some specification (dental insurance for age 18-39….).

1.6. Translations of the questionnaire in multiple languages

The SHAPE team contracted with Betmar Languages of Minneapolis to translate the survey into four different languages:
1) Spanish;
2) Hmong;
3) Vietnamese; and
4) Somali

Some of the questions had already been translated into Spanish by CDC. The questions which had already been translated were sent to Betmar to be used as is.

Betmar's approach was to have one person translate the document and then had a second person look over the work. The SHAPE team asked Hennepin County Multi-Cultural Center staff and persons from the community to review the translations (true?). Some issues were raised about a few of the questions (true?), and they were resolved with the Betmar staff before the final versions were released.

The versions of the survey in the four non-English languages are available in Appendices B to E.

1.7. Survey topics and questions and sources - a comparison between SHAPE 2002 and 1998

2. Request for Proposal

This section needs to be written.

3. Sampling Design

Guiding principles

The guiding principles for selecting the sample for the SHAPE 2002 survey were the following:

1) Have a sufficient number of persons sampled in each of a number geographic areas of interest within Hennepin County to allow for the reporting of results by those areas;
2) Have a sufficient number of persons sampled in each of a number of racial and ethnic population groups to allow for the reporting of results by those groups;
3) Keep the costs within the budget;
4) Be able to generalize the results to the general population

Geographic area requirements

A initial planning decision of the project was to be able to report survey results for 20 geographic areas as listed in Table 3.1.

Table 3.1 Geographic Reporting Areas

1 Hennepin County as a whole
2 The City of Minneapolis (including the Hennepin County portion of the City of St. Anthony)
3 Suburban Hennepin County

The 11 communities (planning districts) in Minneapolis
4 Calhoun-Isles
5 Camden
6 Central
7 Longfellow
8 Near North
9 Nokomis
10 Northeast (including the Hennepin County portion of the City of St. Anthony)
11 Phillips
12 Powderhorn
13 Southwest
14 University

The 3 Suburban Hennepin County Human Service Areas
15 Northwest
16 West
17 South

Suburban Hennepin County sub regional reporting areas
18 The area formed by the aggregate of the cities of Brooklyn Center, Brooklyn Park, and Osseo – all of which are in Northwest Hennepin County

19 The area formed by the aggregate of the cities of Bloomington, Edina, and Richfield – all of which are in South Hennepin County

20 The City of Bloomington

Source: SHAPE 2002
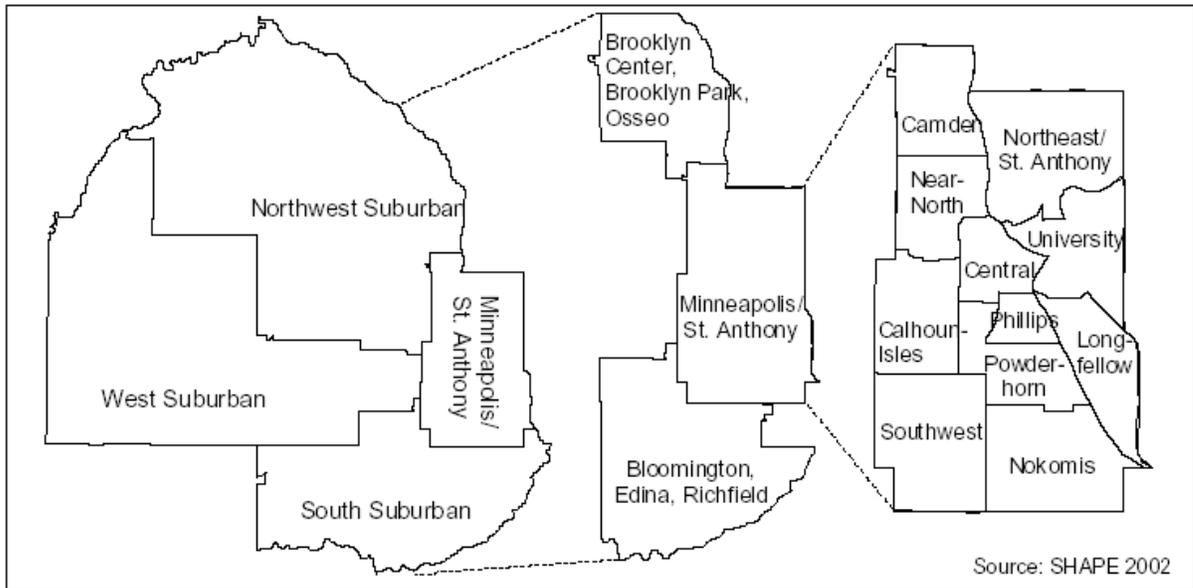
Racial and Ethnic group requirements

Direction was given by the SHAPE 2002 Project Leadership Team to identify the specific racial and ethnic populations for which statistically reliable survey results were needed. Six racial and ethnic groups within the county were selected based on their overall size and to focus on the health status and needs of recent immigrant groups to Hennepin County. The groups are shown in Table 3.2.

Table 3.2 Racial and Ethnic Populations of Interest
American Indian
U.S.-born Blacks
African-born Blacks
Southeast Asians
Hispanics or Latinos
Whites

As collected in the 2000 Census, a person's race and Hispanic ethnicity are considered two separate attributes and are asked as separate questions. Thus, the results of the 2000 Census are available by race (e.g., White, Asian, etc.), by Hispanic ethnicity (i.e., Hispanic), and, for some of the results, by a combination of race and ethnicity (e.g., White non-Hispanic).

The Public Law data (PL-94-171) from the 2000 Census, which was the primary census data available during the planning and early implementation of SHAPE 2002, presented results for the racial categories and for Hispanic ethnicity. (The primary purpose of PL-94-171 is for congressional redistricting.) PL-94-171, however, did not break down the results for subpopulations, such as Southeast Asians as a subgroup of Asians or African-born Blacks as a subgroup of African Americans.

Thus, it was possible to use the PL-94-171 data to locate geographic areas within the county where Hispanics, American Indians, Whites, African Americans, and Asians lived, but not specifically Southeast Asians, U.S.-born Blacks, and African-born Blacks.

The SHAPE methodology team used PL-94-171 information for the larger African-American, Asian, Native Hawaiian and Other Pacific Islander groups in conjunction with information from other sources,

such as birth records, social service program enrollment data, and personal knowledge, to estimate which geographic sub-areas were promising areas for over-sampling.

Since there are multiple definitions for whom to include under the title of Southeast Asian, the SHAPE implementation team chose to use the definition used by the Minnesota Department of Health (MDH). MDH classifies as Southeast Asian anyone who is Hmong or whose ancestry includes people born in Laos, Vietnam, Cambodia, Thailand, Malaysia, the Philippines, Burma, Singapore, or Brunei.

Sampling Strata

In order to report the results for both the required geographic areas plus the selected racial and ethnic groups in an affordable manner, some of the geographic areas were subdivided for sampling purposes based on information about the location of certain racial and ethnic population groups from the 2000 census. The areas that were subdivided were done so based on the 2000 census boundaries – most often based on census tracts, but occasionally on census block groups and census blocks.

The SHAPE project team used information gained from the 2000 Census to locate census tracts, block groups, and blocks that had higher concentrations of the racial and ethnic populations of interest. That information was used to subdivide several of the larger geographic reporting areas. Having multiple smaller areas made it more cost effective to locate sufficient numbers from the racial and ethnic populations of interest.

Sub-areas that had a high number of the racial and ethnic populations of interest were sampled at a higher rate than other areas. For example, approximately 15% of the households in the Richfield-A area were sampled compared to less than 1% of the households in the West Suburban Area.

The resulted in 29 geographic areas that were unique sampling strata; which are listed in Table 3.3.

Table 3.3 Geographic Sampling Strata
Minneapolis
1 Calhoun Isles
2 Camden
3 Central
Longfellow Total
        4 Longfellow-A
        5 Longfellow-B
        6 Longfellow-C
7 Near North
Nokomis Total
        8 Nokomis-A
        9 Nokomis-B
10 Northeast/St. Anthony
Phillips Total
        11 Phillips-A
        12 Phillips-Little Earth
        13 Phillips-B
Powderhorn Total
        14 Powderhorn-A
        15 Powderhorn-B
16 Southwest
University Total
        17 University-A
        18 University-Riverside Plaza
        19 University-B

Suburban Hennepin County

South Suburban Total
Bloomington, Edina, & Richfield Total
       Bloomington Total
              20 Bloomington-A
              21 Bloomington-B
       22 Edina
       Richfield Total
              23 Richfield-A
              24 Richfield-C
       25 Eden Prairie/Ft. Snelling
Northwest Suburban Total
Brooklyn Center, Brooklyn Park, & Osseo Total
       26 Brooklyn Center, Brooklyn Park, & Osseo-A
       27 Brooklyn Center, Brooklyn Park, & Osseo-B
28 Northwest Area Remainder
West Suburban Total
29 West Area

****

Sample from SSI

In preparation for the SHAPE 2002 survey, a set of residential phone numbers with addresses was purchased from Survey Sampling International (SSI), a large and well-respected survey and sampling firm based in Fairfield, Connecticut.

SHAPE purchased 100% of the known phone numbers from SSI for all the sampling strata within the city of Minneapolis, and the Bloomington-A, Richfield-A, and the Brooklyn Center, Brooklyn Park, & Osseo-A strata. A 78.7% sample of the phone numbers for the remaining strata was purchased. The figure of 78.7% was arrived at to maximize the number of available phone numbers given the project's budget.

SSI was able to select the sample using the 1990 Census boundary information only. Thus, the SHAPE implementation team needed to geocode the sample addresses, and assign the 2000 census geography after obtaining the sample data from SSI.

A total of 248,083 phone numbers were obtained from SSI.

Sample from addresses not in the SSI sample

Since 100% of the phone numbers for the City of Minneapolis were purchased from SSI, a spot check of several known, listed phone numbers was conducted. In so doing, it was discovered that several large apartment and condominium buildings in downtown Minneapolis were not included in the SSI list. A more extensive assessment was made to determine the extent of missing, listed phone numbers. It was discovered that not only some of the large multi-dwelling buildings in downtown Minneapolis were missing, but many multi-dwelling units throughout Minneapolis and Brooklyn Center, Brooklyn Park, Richfield, and Bloomington were missing, as well.

1.   SSI was contacted to determine the reason for the missing phone numbers. SSI explained that when selecting residential-only phone numbers, they must apply decision rules to determine whether a phone number is for a residence, or is for a business. These decision rules apparently excluded some, but not all, multi-dwelling buildings.

2.   To supplement our SSI-purchased list of phone numbers, the SHAPE Project Team developed a list of phone numbers and addresses missing from the SSI-purchases list by:

3. Using data from the Hennepin County Property Information System (PINS), and from the City of Minneapolis Assessor's Office, a list was developed of all residential buildings with 10 or more housing units in the cities of Minneapolis, and in the over-sampled areas of Brooklyn Center, Brooklyn Park, Richfield, and Bloomington.

4. Housing unit counts by census block from the 2000 Census were compared with the number of SSI-purchased phone number addresses geocoded to each census block. If there was a large mismatch between the two counts of housing units, aerial photos were used to determine the number of multi-unit dwellings, and PINS data was used to determine the address(es) of the missing building(s).

5. Using the langenberg.com internet resource, a reverse-directory search was conducted for each of the missing addresses. An electronic list of the resident names at those addresses was assembled. A total of **7,NNN** names and addresses were included in this supplemental database.

6. A random sample of names was selected from this list, and the phone numbers for those names was determined using the forward search capabilities at langenberg.com.

Sample of Student Dormitories

A supplemental sample of phone numbers for University of Minnesota dormitory rooms was added to strata University B strata as group quarter residences are not included in the SSI sample. This sample was randomly selected from the electronic phone directory at the University.

4. Marketing Efforts

One difficulty in implementing the SHAPE survey is that the populations which usually are of most interest to persons who set health policy and develop health programs are the very people who are the hardest to reach in a survey such as SHAPE. People who are in the lower economic groups or who have recently moved to Hennepin County from a foreign country are less likely to own a land-based phone, speak English fluently enough to talk on a telephone, and talk openly about private health matters to strangers, especially anyone related to the government.

This difficulty was well known to all the participating agencies involved with SHAPE 2002. It was decided early on, therefore, to develop marketing material and programs to partially counter those naturally-occurring phenomena which would limit the amount of useful data from the populations of special interest (i.e., Hispanics, Southeast Asians, U.S.-born Blacks, African-Born Blacks, and American Indians).

A Marketing Work Group was formed in September 2001. The group consisted of staff from the participating health agencies, Hennepin County's Office of Multi-Cultural Services (OMS), Hennepin County Public Affairs, and Kathy Graves from Perenteau-Graves, who had done much public relations work with the Minneapolis Department of Health and Family Support.

The Work Group focused on a number of areas to give the project an identity and to raise awareness and trust about the SHAPE 2002 project with populations around the county.

4.1 Logo

The Marketing Work Group wanted to give the SHAPE 2002 project an identify focus for all publicity material by developing a logo. After designing several options, the group settled on the following logo.

The logo was developed by asking staff from the Office of Multi-Cultural Services about symbols and colors that might be positive or negative with certain communities. We also discovered that the word survey didn't exist in Somali so to translate the meaning was too long to include the word as part of any logo development.

4.2 Flyers

The Marketing Work Group developed a series of information flyers and posters which could be distributed and posted throughout the county to raise awareness of the upcoming survey. The information flyers and posters were prepared in English, Spanish, Vietnamese, Somali, and Hmong. Betmar Languages did the translation. Appendices A2 to E2 show the five different information flyers. Appendices F2 to J2 show the five different posters.

4.3 Media

The staff from OMS offered advise about the best way to reach their communities and what forms of communication would work best. A contract was written with Total Market Coverage (TMC) to place ads in neighborhood and regional newspapers. TMC placed ¼ ads in selected newspapers in January 2002 and 1/8 inch ads in February 2002.

4.4 Community Events

The Marketing Work Group developed a calendar of upcoming community events where information about the SHAPE 2002 survey could be given out.

5. Data Collection

General approach

The SHAPE 2002 data were collected, for the most part, with telephone interviews using Computer-Assisted Telephone Interview (CATI) technology to manage the phone numbers to be called and the flow of the survey interview. The telephone interviews were augmented by some face-to-face interviews later on in the project. The University of Minnesota's Heath Services Research Center (SRC) managed the operations of the telephone and in-person interviews.

As was described in Section 2, SHAPE 2002 assigned 25 (?) geographic strata to accomplish the project's dual goals of covering the entire geographic area of the county while allowing for reporting on smaller areas and collecting sufficient numbers of specific racial and ethnic populations.

Interviewers, working with the CATI system, called phone numbers from the pool of numbers within each of the strata to fill the quota for that strata. As the data collection progressed, some modifications were made to assist in obtaining sufficient numbers of the racial and ethnic populations of interest.

Interviewer selection

The SRC used approximately xx full- and xx part-time interviewers during the project. Some of the interviewers had been with the SRC prior to SHAPE 2002, while others were hired just for this project.

The SHAPE survey was translated into Spanish, Somali, Hmong, and Vietnamese. Thus there was a need for persons fluent in those languages to administer the survey over the phone. For those surveys, the SRC used some bi-lingual interviewers who had been involved with previous SRC-based projects and did outreach to hire additional temporary workers who were bi-lingual. A total of xx bi-lingual interviewers worked on the project.

Training of interviewers

All new interviewers received some background material on the SHAPE 2002 survey and standardized training on how to work with the CATI system. The training consisted of … An experienced worker or supervisor monitored new interviewers for xx before that interviewer was allowed to make unsupervised calls.

Computer-Assisted Telephone Interviews

The SRC uses the xxx CATI system. SRC staff first programmed the CATI system using the English version of the SHAPE 2002 survey. Hennepin County CHD staff (others?) reviewed versions of the SHAPE 2002 CATI package as it was being developed. The main items reviewed were skip patterns, edit checks, and the wording of the questions. An additional item which was reviewed actively was the the programming of the income level breaks based on the household size.

After several iterations, SHAPE 2002 project staff approved the CATI system for use.

SRC staff developed a Spanish version of the SHAPE 2002 CATI system midway through the data collection period. The interviews administered in Somali, Hmong, and Vietnamese were recorded on paper. The results of those interviews were subsequently transcribed into the English CATI system.

Approximately xxxx interviews were administered to test out the CATI system. These interviews came primarily from geographic areas in Hennepin County which had small numbers of the racial and ethnic populations of interest.

Interviewing in multiple languages

As was stated above, the SHAPE 2002 survey was administered in English, Spanish, Hmong, Somali, and Vietnamese.

SRC staff used the following procedure with regard to the translated versions of the SHAPE survey.

1. A SRC interviewer would make the initial call to a phone number in English. If the survey could be accomplished in English, the interviewer would use English.

2. If the interviewer could not complete the survey in English, the interviewer would speculate, if possible, what the language spoken by the person on the telephone at the house was.

3. When an adequate number of households had been found whose residents had been speculated to speak a particular language, bi-lingual interviewers who spoke that language would call those households back and speak in the speculated language. If the residents in the household actually did speak the language which had been speculated, the interview would then follow the normal procedures. If the people in the household did not speak the language which the previous interview had speculated, the current interviewer would make a new speculation regarding language.

Surname screening

Several months into the project, the SHAPE 2002 steering group focused on the racial and ethnic groups which were below their expected counts for that point in the interviewing process. The group decided to do several activities to increase the likelihood of interviewing adequate numbers of all the racial and ethnic groups.

One technique was to take the list of phone numbers which had not been called yet and screen for names which looked like they were to racial and ethnic populations of interest. Those names were then added to the pool from which numbers were randomly chosen by the CATI system. This process was done for Hispanic/Latinos, African-born Blacks, and Southeast Asians.

Face-to-face interviewing at Little Earth and Riverside Plaza

A second technique used to increase the number of hard-to-reach populations was doing face-to-face interviews at two sites: Little Earth for American Indians and Riverside Plaza for African-born Blacks.

Little Earth is a 200-unit housing area in the Phillips neighborhood of Minneapolis and is home to a large American Indian population. SHAPE 2002 staff contacted Ellie xx, (title?) at Little Earth about doing face-to-face interviews at Little Earth. Ms. xx agreed to let SRC staff do the interviews on-site provided that some additional questions be asked that would be particularly helpful to the Little Earth community.

These questions were the following:

1.
2.
3.

SRC staff went to every household at Little Earth and was able to complete face-to-face interviews at 98 households beyond the xx households which had completed telephone interviews previously.

SHAPE 2002 staff also contacted the management at Riverside Plaza, a large highrise in the University community of Minneapolis, about doing face-to-face interviews there.

After receiving approval and the building management dropped off a flyer at each apartment in the building, interviewers completed 80 face-to-face interviews at the complex. Most of the interviews were with African-born Blacks.

Hiring, Training and Supervision of SHAPE Interviewers

The Survey Resource Center placed ads for part time interviewer positions on the U of M job website and in the Minnesota Daily. Applicants came into the Survey Center for personal interviews.

Training - Telephone:

SRC staff held two training sessions for the interviewing staff in addition to practice interviews with supervisors.

    a. 5 hour training - group training on interviewing techniques and a review of the survey instrument

    b. 4-5 hour training – group training on CATI (Computer Assisted Telephone Interviewing software) and a review of the survey instrument.

    c. Practice interviews were completed with supervisory staff.

Training – Face-to-Face

The training for the face-to-face interviewers (primarily Somali and English speaking – Little Earth) included the following:

    a.   Classroom training in face-to-face interviewing methods

    b.   Observation and certification of interviewers in the field prior to their interviewing actual respondents.

Supervision:

    a. Supervisors were present at all times survey interviews were being conducted.

    b. Supervisors used a silent listening monitor to monitor interviewer calls. About xx % of the total calls were monitored.

Monitoring the Survey Process

Operations

6. Completion Status

Completed Interviews and Survey Performance Rates

A total of 10,098 interviews were completed for the SHAPE 2002 project; 9920 interviews were completed by telephone and 178 were in-person interviews. Table 6.1 identifies the number of completed interviews in the 25 sampling strata and provides response, cooperation and refusal rates.

Table 6.1. Completed Interviews and Survey Performance Rates

Strata		Completed Interviews	Response Rate	Cooperation Rate	Refusal Rate
01	Calhoun Isles	415	70.6%	74.1%	24.7%
02	Camden	790	66.9%	73.8%	23.8%
03	Central	660	58.8%	74.2%	20.5%
04A	Longfellow	182	69.7%	80.2%	17.2%
04B	Longfellow	95	53.4%	68.8%	24.2%
04C	Longfellow	138	69.0%	80.7%	16.5%
05	Near North	716	68.8%	78.0%	19.4%
06A	Nokomis	238	65.7%	73.2%	24.0%
06B	Nokomis	167	57.8%	66.8%	28.7%
07	Northeast/St. Anthony	399	60.3%	68.2%	28.1%
08A	Phillips	516	77.6%	85.0%	13.7%
08B	Phillips	282	61.6%	76.0%	19.4%
09A	Powderhorn	838	67.9%	81.2%	15.7%
09B	Powderhorn	265	63.9%	78.9%	17.1%
10	Southwest	416	65.7%	75.1%	21.8%
11A	University	336	64.0%	81.4%	14.7%
11B	University	269	94.7%	95.4%	4.6%
12A	Bloomington, Edina, Richfield	892	64.1%	70.0%	27.5%
12B	Bloomington, Edina, Richfield	314	67.0%	71.5%	26.7%
12C	Bloomington, Edina, Richfield	105	66.9%	70.9%	27.4%
13	Eden Prairie/Ft. Snelling	173	63.8%	72.4%	24.4%
14A	Brooklyn Center, Brooklyn Park, Osseo	883	67.2%	73.5%	24.2%
14B	Brooklyn Center, Brooklyn Park, Osseo	279	72.7%	79.5%	18.8%
15	Northwest Suburban Area Remainder	176	63.8%	71.3%	25.7%
16	West Suburban Area	554	63.1%	72.5%	23.9%
	Total	**10,098**	**66.3%**	**75.4%**	**21.6%**

Standard definitions from the American Association of Public Opinion Research (AAPOR - *www.aapor.org*) were used to classify call disposition and to calculate survey performance rates. The disposition of all telephone numbers released for calling in SHAPE 2002 is included in Table 6.2.

Table 6.2. Final Call Disposition

Disposition	AAPOR Code	Count
Total Telephone Numbers Released		29,165
Not Eligible		
Quota Filled	4800	3,480
Group Quarters	4530	23
Institution	4520	22
Business/ Govt.	4510	461
Call Forwarding	4430	72
Cell Phone	4420	66
Disconnect	4320	2,535
Non-working Number	4310	6,724
Fax Number	4200	545
Unknown Eligibility, Non-Interview		
Technical Phone Problem	3150	191
Answering Machine	3140	627
No Answer	3130	165
Busy	3120	19
Call Back	3100	228
Eligible, Non-Interview		
Language Unable to Complete	2330	240
Physically/ Mentally Unable to Complete	2320	374
Refusal	2110	3,295
Eligible, Interviewed		
Completed - Telephone Interviews	1100	9,920
Completed - In-person Interviews		178

Survey Response, Cooperation and Refusal Rates

Response Rate is calculated as:

$$\frac{\text{Completed Interviews}}{(\text{Eligible, Interviewed} + \text{Eligible, Not-Interviewed} + \text{Unknown Eligible, Not Interviewed})}$$

The overall response rate for SHAPE 2002 was 66.3%, ranging from 53.4% in strata 04B Longfellow to a high of 94.7 in strata 11B University.

Cooperation Rate is calculated as:

<div align="center">Completed Interviews</div>

<div align="center">(Eligible, Interviewed + Refusals)</div>

The cooperation rate for this survey was 75.4%, ranging from a low of 66.8% in strata 06B Nokomis to a high of 95.4% in strata 11B University.

Refusal Rate is calculated as:

<div align="center">Refusals</div>

<div align="center">(Eligible, Interviewed + Eligible, Not-Interviewed + Unknown Eligible, Not Interviewed)</div>

The refusal rate for this survey was 21.6%, ranging from a low of 4.6% in strata 11B University to a high of 28.7% in strata 06B Nokomis.

Conducting the survey in multiple languages and utilizing mixed-mode interviewing methods improved the survey performance. Ten percent of the completed surveys were interviewed in a non-English language. Table 6.3 provides a summary of sample source, mode of interview and language.

Table 6.3

	Interview Language					
Interview source and mode	**English**	**Spanish**	**Somali**	**Hmong**	**Vietnamese**	**Total**
SSI Sample - Telephone	8401	341	179	203	172	9296
Apartment Sample - Telephone	404	2	10	1	2	419
Dormitory Sample - Telephone	65	1	0	0	0	66
Little Earth – In-person	98	0	0	0	0	98
Riverside Plaza – In-person	0	3	76	1	0	80
Total	8968	347	265	205	174	9959

7. Management of SHAPE 2002 data

SRC provided the final data set of 10,098 completed interviews in SPSS format to the project team. The data set included variable labels set to survey questions, and value labels set to response categories.   The data set included the responses for all survey questions, the geography variables from the original sample files, process flag variables that identified sample source/mode and interview language, and computed primary race and ethnicity variables that were used to classify respondents to the cultural groups specified in the project contract with SRC.
A second SPSS data set was provided by SRC that included the actual responses to open-ended questions and the text responses to "Other: specify _____" categories for certain questions.

Data Cleaning
The CATI system used by SRC was programmed to minimize the data cleaning requirements of the project. The CATI system included pre-programmed skip patterns, limited responses to categorical values, minimum and maximum value testing for numeric responses, and some logic testing of specific responses.

The initial data set provided by SRC included 10,098 interviews, consisting of 10,024 completed interviews and 74 partially completed interviews. The first data cleaning step evaluated variables required for weighting the sample including strata, household size, age, gender, and race/ethnicity.

The number of household members by age group was collected in the six-part question 94a-f. Three variables (adults, children and hhsize) were computed by summing the parts of question 94. A final household size for weighting variable (hhsizew) was created after the following adjustments for missing and illogical responses.

Data Problem	Count of Cases	Adjustment Rule
Count is missing for at least 1 age category	129	Set count for missing categories to 0 Compute # of adults and household size
No one in respondent's age category	229	Increase # of adults and household size by 1
Household size = 1 and Married/Living Together	188	Increase # of adults and household size by 1

Evaluation of partial completes, missing values, don't know and refusal responses.
The project data cleaning work group examined all records in the dataset for completeness by evaluating partially completed interviews, and interviews which

contained a high number of missing responses or responses of don't know or refused. As was agreed upon by the project team, SRC eliminated interviews which stopped prior to question 30 from the delivered dataset. Of the 10,098 interviews provided by SRC in the final dataset, 74 interviews were terminated by the respondent after question 30 but before the end of the interview. Questions that were not answered during these 74 interviews and any other missing response on completed interviews were treated like refusals. A summary variable was computed as the total number of refusals for the interview.   After reviewing the distribution of this summary variable it was decided to exclude interviews with 50 or more refusals from the final dataset. In addition, as age of the respondent was identified as a critical variable for analysis and reporting, all interviews with missing age were dropped from the final dataset. Based on the age and missing data criteria, 139 interviews were dropped from the final research dataset. The final research dataset for SHAPE 2002 contained 9,959 interviews.

	Interviews
Total	10098
Drop due to missing age	56
Drop due to 50+ refuses	83
Final File	9959

In the final dataset delivered from SRC, all categorical responses of "don't know " or "refused" were identified as missing values for analysis. The data cleaning work group examined the frequency of don't know responses for each question with results ranging from 0% - 10.7%. Ten questions had don't know response rates exceeding 2%. Questions relating to community and cultural heritage (q13a-d, q14, q15, q16, q17), sexual orientation (q89), and income (q95a-c) had the highest rates for don't know responses. Further analyses were performed to determine the respondent characteristics associated with "don't know" responses for these ten questions.

The project team decided to include don't know as a valid response to the community questions (q13a – d) as we found consistency in the don't know response rate across strata, age, gender, and racial/ethnic groups. Perceptions about a community require time to develop and therefore don't know is an appropriate response to these questions.

Although there were small variations in the characteristics of don't know respondents were found for question 14 – 17 and question 89, no adjustments were made to the dataset.

A three level question (q95a – c) was used to collect household income for SHAPE 2002. Question 95a asked for an income dollar amount which was given by 66% of respondents. Question 95b was asked of the remaining one-third of respondents requesting an income level category from 100% - 600% + of FPL

based on household size. Almost 50% (16% of all respondents) of this group identified an income category.   The remaining group was asked question 95c, which requested an income category of more or less than 200% of poverty level based on household size. Almost 30% (6% of all respondents) of this group provided this level of income information. The remaining 11.5% of the sample did not provide income information.   Household utilization of income eligibility based social service programs identified by the responses to q87d-h and j was used as a proxy for income to categorize 2.5% of the unknown income group to a category of less than 200% of poverty. See income and poverty variables *pov_lev1 through pov_lev4.*

Final Race and Race/Ethnicity Codes.
Included in the final dataset provided by SRC was a computed variable *race* which was developed in order to monitor response goals for specific racial categories as specified in the contract. The project data cleaning work group evaluated this variable in comparison to the responses to Q5. County of birth; Q6. Hispanic/Latino origin; Q7. Race (multiple responses); Q8. Primary race; Q9. Asian heritage; and Interview language. The work group created two new variables based on this evaluation - for final primary race code *f_race* and final racial/ethnic code *f_raceeth* . Fifty-nine (59) records were recoded primarily from Multiracial and Other categories to a single primary race category.

Classification and re-coding of open-end responses
A second SHAPE 2002 dataset was provided by SRC that contained the text responses to one open-end question (Country of birth) and 11 questions which provided an Other specify category.   A project work team reviewed this dataset to categorize the responses.   The category codes for each question were appended or updated to the SHAPE 2002 research dataset.

Section 8
Weighting

When survey data are analyzed, one has the choice of having the answers given by each respondent be "unweighted" or "weighted". If the data are "unweighted", each respondent's answer has the same influence in computing the estimate for the average response for the population. [average = $\sum$response$_i$ / number of respondents; where i goes from 1 to the number of respondents] If the data are "weighted", the responses for each respondent are assigned a specified weight > 0 which varies from one individual to another. That assigned weight is used in estimating the average response for the population [average = $\sum$(response$_i$ * weight$_i$) / number of respondents; where i goes from 1 to the number of respondents].

In a survey such as SHAPE, weighting the data is more appropriate for a number of reasons: A household in one geographic area was more likely to be picked than a household in other geographic areas within the county. A person in a household in which many adults live was less likely to be selected than a person in a household with few adults. The demographics of the SHAPE respondents are different in some ways than the population of the county based on the 2000 Census. It is known from other data sources that health status and behaviors are different in various geographic areas in the county and also differ, in aggregate, by the age, gender, and race or ethnicity of the person. Thus, results using unweighted data are likely to be significantly different than the true value for a particular geographic region or population.

For SHAPE 2002, there were weights computed for the following types of analyses:

1. Geographic areas including the county as a whole [wgeog];
2. Race regardless of Hispanic/Latino ethnicity [wrace];
3. Race or Hispanic/Latino ethnicity when they <u>are</u> mutually exclusive (e.g., Hispanic/Latino, White non-Hispanic, American Indian non-Hispanic, etc.) [wraceeth]; and
4. Households rather than individuals [wgthhold].

These weights were computed using the best available information at the time – primarily SF-1 data from the 2000 U.S. Census. These weights were used to compute the results in the initial reports coming out of the SHAPE project (i.e., SHAPE 2002: Geographic Data Book and SHAPE 2002: Racial and Ethnic Data Book).

Since then, some data have been released by the U.S. Census Bureau which has allowed for better estimating the populations of racial and ethnic sub-groups (e.g., Southeast Asians rather than only the more inclusive Asian/Pacific Islanders) within Hennepin County.   This more recent data has allowed us to compute a new set of weights which gives better estimates for the many sub-populations within Hennepin County.

Thus, the weights listed above should be used only to match the numbers in the published SHAPE 2002 data books. We recommend that all subsequent work use the following set of weights:

1. Geographic areas including the county as a whole [wgt_g];
2. Race regardless of Hispanic/Latino ethnicity [wgt_r];
3. Race or Hispanic/Latino ethnicity when they <u>are</u> mutually exclusive (e.g., Hispanic/Latino, White non-Hispanic, American Indian non-Hispanic, etc.) [wgt_re]; and
4. Households rather than individuals [wgthhold]. Note: this weight did not change because of the change in data sources used to compute the weights.

The philosophy used in developing the weights was to do the least amount of weighting necessary while still taking into account major disproportionate sampling factors and post-stratification disparities. Doing weighting beyond the minimum can sometimes result in unintended consequences, such as trying to fix one misfit of the data and causing a new misfit to appear in a different part of the data.

Some assumptions that were made during the calculation of the weights were the following:

1. The percent of all possible phone numbers that were in our sample from Survey Sampling, Inc. (SSI) and the supplemental list [(# of residential phone numbers from SSI_i + # of residential phone numbers in the supplemental $list_i$)/ # of all actual residential phone $numbers_i$ for all sampling strata i] was relatively constant across all sampling strata.

2. Within each stratum, the households for which we obtained phone numbers – either from SSI or the supplemental list – are similar to the households for which we do not have phone numbers

3. It was not necessary to adjust for differing non-response rates across the strata.

4. The people who self-identify as Hispanic and live in a household whose head of household has a Hispanic surname are similar to people who self-identify as Hispanic but live in a household whose head of household does not have a Hispanic surname

5. The people who self-identify as Southeast Asian and live in a household whose head of household has a Southeast Asian surname are similar to people who self-identify as Southeast Asian but live in a household whose head of household does not have a Southeast Asian surname

6. The probability of being included in the SSI list is similar to the probabiliy of being included in the apartment building supplementary source, so we will treat the apartment building supplementary sample numbers as if they came from SSI; and

7. The mode effect is minor between the responses of those who completed the survey by phone and those at Little Earth and Riverside Plaza who completed the survey face-to-face, so the samples can be combined

8. It was not necessary to make an adjustment for the number of phones within the home. The question as to how many phones there were in the home was not asked of the SHAPE respondents. The assumption was that any adjustment made because of this factor would be very minor, so it was not necessary to ask for this information.

The methods used to compute each of the weights are described in the sections below.

Weight for geographic areas: wgeog and wgt_g

The two steps used to compute wgeog included
1. adjust for disproportionate sampling rates; and
2. do any, but no more, post-stratification adjustment that is needed

The disproportionate sampling rate adjustment is done by including three components:
1. adjust for the percent of households sampled within a given stratum;
2. adjust for the number of adults living within the household; and
3. normalize the resultant weights so that the sum equals the number of respondents to the SHAPE 2002 survey.

Table S2.1 shows the percent of the households and of the adults who were sampled within each of the sampling strata.

The SHAPE 2002 survey asked the respondent for the number of adults living within the household.

The non-normalized weight (NNW) for respondent$_i$ would then be
NNW$_i$ = (number of households in respondent$_i$'s stratum /
   number of respondents in respondent$_i$'s stratum) *
   number of adults living in respondent$_i$'s household)

For example, if respondent$_i$'s stratum had 2000 households in it, 500 SHAPE respondents lived in that stratum, and 4 adults lived in respondent$_i$'s household, then NNW$_i$ would be (2000/500) * 4 or 16.

The normalized disproportionate sampling weight (NDSW) for respondent$_i$ would then be
NDSW$_i$ = [NNW$_i$ / ((∑ NNW$_i$ for all respondents$_i$) / number of respondents)]

Continuing the example from above, if the sum of the NNW$_i$ for all 9,959 SHAPE respondents was 79,672, then NDSW$_i$ would equal 16 / (79,672 / 9,959) or 2.

Now that the equal probability, or disproportionate sampling, factor has been addressed, it is now time to see if any post-stratification adjusting is necessary.

To accomplish that, weight the data with the NDSW and compare the demographics of the sample with the 2000 Census results.

Table S2.2 shows the percent of both the SHAPE 2002 respondents (weighted by NDSW) and the population from the 2000 Census which are in 20 different age/gender/resident of Minnepolis or Suburban Hennepin County groupings. The table shows that there are some groupings which were represented either much more or much less in the SHAPE 2002 survey respondents than the 2000 Census. For example, 10.45% of the Census 2000 population for Hennepin County were women, aged 45-64 who lived in the suburban part of the county. However, 16.23% of the SHAPE 2002 respondents, after weighting to adjust for disproportionate sampling, had those same demographic characteristics.

Since there were a number of these type of differences, it seems appropriate to do post-stratification adjusting.

Assume that j represents the gender of the respondent;
  k represents the age group of the respondent (18-24, 25-44, 45-64, 65-84, and 85+); and
  l represents whether the repondendent lives in Minneapolis (including the Hennepin County portion of St. Anthony) or in suburban Hennepin County.

The post-stratification adjustment$_1$ (PSA1) will equal
(2000 Census count$_{jkl}$ / Total 2000 Census population for Hennepin County) /
(Number of SHAPE 2002 respondents$_{jkl}$ / Total SHAPE 2002 respondents).

In our example, assume that respondent$_i$ is a male who lives in suburban Hennepin County and whose age is between 65 and 84. Table S2.2 shows that there were 31,810 such people living in Hennepin County based on the 2000 Census and that there were 384 such SHAPE respondents. The total 2000 Census adult population was 848,698 and there were 9,959 total SHAPE respondents. So, PSA1 for respondent$_i$ will be
(31,810 / 848,698) / (384 / 9,959) or .972.

The cumulative weight for a given respondent$_i$ will then be NDSW$_i$ * PSA1$_i$. For our example person, that cumulative weight would be 2 * .972 or 1.944.

Because of the special outreach efforts for some populations of interest in SHAPE 2002, a post-stratification adjustment based on race/ethnicity is likely to be also needed.

The four racial/ethnic groups that were chosen to be included for this step included the following:
        a) Hispanic/Latinos;

b) Black-non Hispanics;
c) Asian/Pacific Islander-non Hispanics; and
d) All others

These groups were further broken down by geographic area of residence: Minneapolis and Suburban Hennepin County. These four racial/ethnic groups were selected because their percentage in the sample was influenced greatly because of the over-sampling that was done and their numbers within the adult population in the county were large enough to significantly affect the results.

Table S2.3 shows the percent of both the SHAPE 2002 respondents (weighted by NDSW * PSA1$_i$) and the population from the 2000 Census which are in eight different race/ethnicity and resident of Minnepolis or Suburban Hennepin County groupings. The table shows that there are some groupings which were represented either much more or much less in the SHAPE 2002 survey respondents than the 2000 Census. For example, 4.99% of the Census 2000 population for Hennepin County were Black non-Hispanics who lived in Minneapolis. However, 3.52% of the SHAPE 2002 respondents, after weighting to adjust for disproportionate sampling and doing the post-stratification adjustment PSA1, had those same demographic characteristics.

One final adjustment is needed.

Information gained from a variety of data sources show that there are major differences in behaviors, attitudes, and experiences between Blacks who are U.S.-born and have lived in this country for a long time and Blacks who are recent arrivals from Africa. The SHAPE sampling methodology resulted in a large over-representation of African-born Blacks in the sample. Thus, it is necessary to adjust for this imbalance. The groups included in the adjustment were U.S.-born Blacks and foreign-born Blacks. Once again, we also adjusted for geography, i.e., Minneapolis vs Suburban Hennepin County.

Table S2.4 shows the the percent of both the SHAPE 2002 respondents who self-identified their race as Black [weighted by NDSW * PSA1 for the Black population] and the population from the 2000 Census which are in the four different U.S.-born/Foreign-born and resident of Minnepolis or Suburban Hennepin County groupings.

The table shows that all the groupings were represented either much more or much less in the SHAPE 2002 survey respondents than the 2000 Census. For example, 11.02% of the Census 2000 Black population for Hennepin County were foreign-born who lived in Minneapolis or St. Anthony. However, 30.90% of the SHAPE 2002 respondents, after weighting to adjust for disproportionate sampling and the previous post-stratification adjustments, had those same demographic characteristics.

Assume that m represents the birth place (U.S. or a foreign country) of the respondent and n represents whether the repondendent lives in Minneapolis (including the Hennepin County portion of St. Anthony) or in suburban Hennepin County.

The post-stratification adjustment$_3$ (PSA3) will equal
(2000 Census count of persons who self-identified their race as Black alone$_{mn}$ / Total 2000 Census population of persons who self-identified their race as Black alone for Hennepin County) /
(Number of SHAPE 2002 respondents who self-identified their race as Black$_{mn}$ / Total SHAPE 2002 respondents who self-identified their race as Black).

The cumulative weight wgeog for a given respondent$_i$ who self-identified their race as Black will then be NDSW $_i$ * PSA1$_i$ * PSA3$_i$.

For all others, the cumulative weight wgeog is NDSW $_i$ * PSA1$_i$ * PSA2$_i.$

The computation for wgt_g was similar to wgeog except for the following:

Weight for analyses for racial groups: wrace and wgt_r

The weight, wgeog, is appropriate for any analysis where the unit of analysis is a geographic area, such as suburban Hennepin County. When one wants to look at results for one or more racial groups, however, a different weight is needed. That weight is wrace.

The weight, wrace, is actually a composite of weights for each of the five racial groups –
American Indians;
Asian/Pacific Islanders;
Blacks;
Whites; and
All Others

Five sets of weights are calculated - one each for respondents of each of the five racial groups. Since all respondents are assigned to one of the five racial groups and no respondent is assigned to more than one racial group, all respondents are thus assigned to one racial weight. It is therefore possible to merge the five weights into one weight variable. This step simplifies any future analyses.

The two steps used to compute each of the components of wrace included
1. adjust for disproportionate sampling rates for each racial group; and
2. do any, but no more, post-stratification adjustment that is needed

Tables S2.5 to S2.9 show the percent of the households which were sampled within each of the sampling strata for persons who self-identified of one of the five racial groups.

The SHAPE 2002 survey asked the respondent for the number of adults living within the household.

The non-normalized weight (NNW) for respondent$_i$ would then be
NNW$_i$ = (number of households in respondent$_i$'s stratum /
   number of respondents in respondent$_i$'s stratum) *
   number of adults living in respondent$_i$'s household)

The normalized disproportionate sampling weight (NDSW) for respondent$_i$ would then be
NDSW$_i$ = [NNW$_i$ / (($\sum$ NNW$_i$ for all respondents$_i$) / number of respondents)]

The same post-stratification adjustment will be done for wrace as was done for wgeog.

Tables S2.10 to S2.14 show the percent of both the SHAPE 2002 respondents (weighted by NDSW) and the population from the 2000 Census which are in 20 different age/gender/resident of Minnepolis or Suburban Hennepin County groupings for each of the five racial groups. The tables show that there are some groupings which were

represented either much more or much less in the SHAPE 2002 survey respondents than the 2000 Census.

Since there were a number of these type of differences, it seems appropriate to do post-stratification adjusting.

Assume that j represents the gender of the respondent;
  k represents the age group of the respondent (18-24, 25-44, 45-64, 65-84, and 85+); and
  l represents whether the repondendent lives in Minneapolis (including the Hennepin County portion of St. Anthony) or in suburban Hennepin County.

The post-stratification adjustment$_1$ (PSA1) will equal
(2000 Census count$_{jkl}$ for a given racial group / Total 2000 Census population for Hennepin County for a given racial group) /
(Number of SHAPE 2002 respondents$_{jkl}$ for a given racial group / Total SHAPE 2002 respondents for a given racial group).

The cumulative weight for a given respondent$_i$ will then be NDSW $_i$ * PSA1$_i$.

As was done with wgeog, a second post-stratification adjustment is needed for persons whose self-identified race is Black.

Table S2.4 shows the the percent of both the SHAPE 2002 respondents who self-identified their race as Black (weighted by NDSW * PSA1) and the population from the 2000 Census which are in the four different U.S.-born/Foreign-born and resident of Minnepolis or Suburban Hennepin County groupings.

The table shows that all the groupings were represented either much more or much less in the SHAPE 2002 survey respondents than the 2000 Census. For example, 11.02% of the Census 2000 Black population for Hennepin County were foreign-born who lived in Minneapolis or St. Anthony. However, 30.90% of the SHAPE 2002 respondents, after weighting to adjust for disproportionate sampling and the first post-stratification adjustment, had those same demographic characteristics.

Since there were a number of these type of differences, it seems appropriate to do a second post-stratification adjustment for people who self-identified their race as Black.

Assume that m represents the birth place (U.S. or a foreign country) of the respondent and n represents whether the repondendent lives in Minneapolis (including the Hennepin County portion of St. Anthony) or in suburban Hennepin County.

The post-stratification adjustment$_2$ (PSA2) will equal
(2000 Census count of persons who self-identified their race as Black alone$_{mn}$ /
 Total 2000 Census population of persons who self-identified their race as Black alone for Hennepin County) /

(Number of SHAPE 2002 respondents who self-identified their race as $Black_{mn}$ / Total SHAPE 2002 respondents who self-identified their race as Black).

The cumulative weight wrace for a given $respondent_i$ who self-identified their race as Black will then be $NDSW_i * PSA1_i * PSA2_i$.

For all others, the cumulative weight wrace is $NDSW_i * PSA1_i$.

The computation for wgt_r was similar to wrace except for the following:

Weight for analyses for racial/ethnic groups: wraceeth and wgt_re

The weight, wrace, is appropriate for any analysis where the unit of analysis is a racial group, such as Asian/Pacific Islanders, regardless of that person's Hispanic/Latino ethnicity. When one wants to look at results for Hispanics or for persons who are not Hispanic/Latinos and of a particular race, however, a different weight is needed. That weight is wraceeth.

The weight, wraceeth, is actually a composite of weights for each of the six racial/ethnic groups –
Hispanic/Latinos;
American Indian non-Hispanic/Latinos;
Asian/Pacific Islander non-Hispanic/Latinos;
Black non-Hispanic/Latinos;
White non-Hispanic/Latinos; and
All Other Races non-Hispanic/Latinos

Six sets of weights are calculated - one each for respondents of each of the six racial/ethnic groups. Since all respondents are assigned to one of the six racial/ethnic groups and no respondent is assigned to more than one racial/ethnic group, all respondents are thus assigned to one racial/ethnic weight. It is therefore possible to merge the six weights into one weight variable. This step simplifies any future analyses.

The two steps used to compute each of the components of wraceeth included
1.   adjust for disproportionate sampling rates for each racial/ethnic group; and
2.   do any, but no more, post-stratification adjustment that is needed

Tables S2.15 to S2.20 show the percent of the households which were sampled within each of the sampling strata for persons who self-identified of one of the six racial/ethnic groups.

The SHAPE 2002 survey asked the respondent for the number of adults living within the household.

The non-normalized weight (NNW) for respondent$_i$ would then be
NNW$_i$ = (number of households in respondent$_i$'s stratum /
   number of respondents in respondent$_i$'s stratum) *
   number of adults living in respondent$_i$'s household)

The normalized disproportionate sampling weight (NDSW) for respondent$_i$ would then be
NDSW$_i$ = [NNW$_i$ / (($\sum$ NNW$_i$ for all respondents$_i$) / number of respondents)]

The same post-stratification adjustment will be done for wraceeth as was done for wrace.

Tables S2.21 to S2.26 show the percent of both the SHAPE 2002 respondents (weighted by NDSW) and the population from the 2000 Census which are in 20 different age/gender/resident of Minnepolis or Suburban Hennepin County groupings for each of the six racial/ethnic groups. The tables show that there are some groupings which were represented either much more or much less in the SHAPE 2002 survey respondents than the 2000 Census.

Since there were a number of these type of differences, it seems appropriate to do post-stratification adjusting.

Assume that j represents the gender of the respondent;
  k represents the age group of the respondent (18-24, 25-44, 45-64, 65-84, and 85+); and
  l represents whether the repondendent lives in Minneapolis (including the Hennepin County portion of St. Anthony) or in suburban Hennepin County.

The post-stratification adjustment$_1$ (PSA1) will equal
(2000 Census count$_{jkl}$ for a given racial/ethnic group / Total 2000 Census population for Hennepin County for a given racial/ethnic group) /
(Number of SHAPE 2002 respondents$_{jkl}$ for a given racial/ethnic group /
Total SHAPE 2002 respondents for a given racial/ethnic group).

The cumulative weight for a given respondent$_i$ will then be NDSW $_i$ * PSA1$_i$.

A second post-stratification adjustment is needed for persons whose self-identified racial/ethnic group is Black non-Hispanic/Latino.

Assume that m represents the birth place (U.S. or a foreign country) of the respondent and n represents whether the repondendent lives in Minneapolis (including the Hennepin County portion of St. Anthony) or in suburban Hennepin County.

The post-stratification adjustment$_2$ (PSA2) will equal
(2000 Census count of persons who self-identified their race as Black non-Hispanic alone$_{mn}$ / Total 2000 Census population of persons who self-identified their race as Black non-Hispanic alone for Hennepin County) /
(Number of SHAPE 2002 respondents who self-identified their race as Black non-Hispanic$_{mn}$ / Total SHAPE 2002 respondents who self-identified their race as Black non-Hispanic).

The cumulative weight wraceeth for a given respondent$_i$ who self-identified their race as Black non-Hispanic will then be NDSW $_i$ * PSA1$_i$ * PSA2$_i$.

For all others, the cumulative weight wrace is NDSW $_i$ * PSA1$_i.$

The computation for wgt_re was similar to wraceeth except for the following:

Weight for analyses for households: wgthhold

The weight, wgthhold, is appropriate for any analysis where the unit of analysis is a household rather than an individual.

Because the unit of analysis is a household rather an individual, there needs to be only an equal probability adjustment made for wgthhold rather than any additional post-stratification adjustments.

Table S2.1 shows the percent of the households and of the adults who were sampled within each of the sampling strata.

The non-normalized weight (NNW) for respondent$_i$ would then be
NNW_i = (number of households in respondent$_i$'s stratum /
   number of respondents in respondent$_i$'s stratum)

The normalized disproportionate sampling weight (NDSW) for respondent$_i$ would then be $NDSW_i = [NNW_i / ((\sum NNW_i$ for all respondents$_i$) / number of respondents)]

No post-stratification adjustments are needed for this weight.

Section 9
Public Use Data File

The SHAPE 2002 project team is eager to partner with researchers from academia and community-based organizations to use the SHAPE data in examining the health of Hennepin County adults. As a way to accomplish that, a public use data file from SHAPE 2002 is being made available.

The SHAPE 2002 Public Use Data File (PUDF) contains most of the raw data from the survey. Some of the original variables are not in the PUDF. Those variables were omitted because they could make it possible to identify the respondents, they contained data that was subsequently cleaned and replaced by more accurate variables, or were used only for the administration of the survey.

In addition, the PUDF contains a number of new variables developed during the initial analysis of the SHAPE 2002 data. All variables developed during the generation of the Geographic and Racial and Ethnic Data Books are included so researchers can match the original published results.

To assist researchers in using the PUDF, a SHAPE 2002 Public Use Data File User Guide was developed. The User Guide helps the researcher get started using the data and gives guidance on how best to use and interpret the many variables included. All researchers who partner with the SHAPE team will be given a User Guide to assist them with their analyses.

It needs to be noted that the analysis of SHAPE 2002 data requires the use of statistical software packages that can produce appropriate estimates and standard errors while taking into consideration survey sample features such as clustering, stratification, and unequal probability sampling that were used in the survey design. Two software packages that will produce appropriate standard errors are STATA and SUDAAN.

Table T2 lists the variables that are included in the SHAPE 2002 PUDF.

Individuals and organizations who are interested in obtaining a copy of the SHAPE 2002 PUDF should contact Margaret Hargreaves at 612-348-7416 or at margaret.hargreaves@co.hennepin.mn.us.

Section 10
Lessons Learned

Anytime a group works on a large project such as the SHAPE 2002 study, it is a combination of borrowing from previous similar projects, working with current realities, and trying out new ideas. SHAPE 2002 is no exception. Many decisions and choices along the way were made because they made the most sense at that time. Now that the project is over, and everyone has the benefit of 20/20 hindsight, it is useful to document some of the lessons that the project team has learned throughout the journey. It is hoped that future efforts include many of steps that ended up being valuable, do not repeat mistakes that were made, and are of even higher quality.

Questionnaire Design and Content

1. **Most respondents identified with one racial/ethnic group.** Although the SHAPE respondents were allowed to check all the racial and ethnic groups that applied, most were able to select one group as their "primary" racial/ethnic group. Most adult respondents are able to indicate a primary racial/ethnic group. Currently multiracial categories are chosen primarily for children when the parents are of different races. This situation may change in the future as more multiracial children become adults.

2. **Arthritis for all.** The question "Have you been told that you have arthritis" should be asked of all respondents, not just those age 55 or older. The younger age group is also affected by arthritis and the rate is believed to be increasing due to the rising rate of obesity. Collecting the overall prevalence rate of arthritis among the younger populations will allow us to monitor the arthritis rate over time. We need to pay attention to the recent change of the BRFSS (Behavioral Risk Factor Surveillance System) question/module on arthritis.

3. **Ask questions about "barriers to care" directly.** Ask questions about some specific barriers to accessing care directly rather than leaving it open ended. We know what the common barriers are from previous research. Questions on specific barriers will allow us to measure the frequency of those barriers.

4. **Improve physical activity questions.** Add the occupation code for questions related to physical activity at work so that the type of work, e.g. construction, is clarified for later analysis on the amount of physical activity resulting from the job.

5. **Use core plus module model.** Use a core module for surveillance and then additional modules for subpopulations (e.g., young adults, boomers, seniors, racial/ethnic groups, etc.) and diseases and disabilities of interest. A core plus module model will allow researchers to add questions of

particular interest to each subpopulation without making the survey overly long for all respondents.

6. **Include cancer.** Add cancer and cancer type to chronic condition list to complement the data available from the cancer surveillance system and death records.

7. **Expand ADL and IADL.** Expand the ADL and IADL questions for seniors only. Identify specific ADL/IADL conditions to increase the information available regarding both the extent of these conditions and the characteristics of the persons who have them.

8. **Develop child health module.** Develop a module on child health that could be implemented at the same time as the adult survey. SHAPE 2002 included 2,929 households with children under the age of 18. If the module had been available, information on child health could have been collected at that time.

9. **Modifying questions.** No matter how long it takes, the group should be involved if questions need to added, modified or deleted. The involvement of the Survey Content Task Force when modifying, adding or deleting questions may increase the length of time to design the questionnaire but the final product will be better.

10. **Assess English literacy and primary language spoken.** Such questions may include primary language spoken at home, the questions used in Census 2000, or questions used by other major population surveys. Asking about the respondent's primary language will ensure using the most appropriate survey.

11. **Health Literacy.** Be aware of health literacy issues including English competency and familiarity with medical terminology.

12. **Interpretation of the same question by different racial and ethnic groups.** Some questions may be more susceptible to different interpretation by different racial and ethnic groups. One example is the "fatty foods" question. This was not an appropriate question for Asians/Pacific Islanders since many of them do not each pizza or breakfast sausage but rather eat stir fried food. Modify the question about eating fatty foods so that it is appropriate for persons of various racial/ethnic groups.

13. **Multiple people should review the questions, especially those with expertise in the area the question examines so that the questions will be worded clearly and are most appropriate.**

14. **Revisit questions omitted from SHAPE 2002 that were included in SHAPE 1998 so that the SHAPE 1998 questions are not automatically omitted from consideration for the next survey.** These questions might be good candidates to ask every other survey.

15. **Translation issues. First, allow plenty of time to** translate the survey into other languages. The process of translating, back translating and community review takes longer than one might anticipate. Second, be careful that the reading level for the translated survey is appropriate and not too high. Finally, be alert for medical terms that might not have an equivalent counterpart in other languages. This discrepancy might lead to non-equivalent questions and responses across people using different language versions of the survey.

Request for Proposal

1. **Set up criteria for selecting a vendor before the Request for Proposal is sent to the vendors.** Document everything and have the documentation available to anyone so that all vendors have the same information available to them and no vendor has an "edge" in the application process due to information not available to other vendors. Documentation also keeps all members of the SHAPE team on the "same page" and ensures that the needed information is asked.

2. **Provide as much detail in the RFP (request for proposal)** to save having to answer common questions from many vendors.

3. **A pre-proposal conference is a good idea**. All interested vendors should be invited to a conference to answer all questions about the survey. A conference call can be set up for those interested vendors that can't attend the conference. All questions/answers from the conference should be available in written form to allow all vendors equal access to the information from the conference. This will level the playing field since the same information will be available to all vendors.

4. **Allow enough time for approval of the Request for Proposal by the parties that must approve the RFP.** Each level that must approve the request will increase the amount of time before the RFP can be submitted to interested vendors. The amount of time required by each step of the process should be communicated to members of the committee as well as the vendors.

5. **Check with prospective vendors for availability for the project**. Some prospective vendors may not be available for a project of this scope or have the resources available to do the job.

Sampling Design

1. **Sampling lists may not contain all the information that the company selling the list claims it does.** Be aware when you purchase phone numbers from a vendor that specific housing/dwelling units may not be included in the universe. The SHAPE 2002 project team found 192

properties, each of which had 10 or more housing units (for a total of 12,231 housing units) that were not included in the purchased list of phone numbers. This list came from Survey Sampling International (SSI), a large and well-respected survey and sampling firm based in Fairfield, CT. If this omission had not been caught, and our subsequent development of a methodology to determine many of the missing phone numbers, a large number of potential respondents in certain geographic areas would have been missed.

2. **Goals are not the same as quotas.** Do not define the number of respondents of a particular racial or ethnic group that one expects to get as the goal or target number. The number of U.S-born Blacks that were obtained was substantially lower than was expected and this resulted in problems when it was time to weight the data. This discrepancy between expected and actual was due in large part to the surveyors taking the lower number as a target rather than a bare minimum.

3. **Researchers and Community groups want the SHAPE results for particular subgroups (e.g. Somali, Hmong, cities), which are smaller than the survey was designed for.** It is important to either, inform people up front what the smallest units of analysis will be or to plan for analysis of smaller subpopulations.

Communication

1. **It was beneficial to work closely with community groups and organizations to develop effective communication strategies.** A close working relationship helps the team design the project in culturally appropriate ways, increases awareness of the survey and helps to increase participation if members of the community of interest know that their leaders support the survey. Furthermore, involving the community groups help deepen relationships for future joint projects.

2. **It would be helpful to evaluate the various components of the marketing program to assess their effectiveness at reaching the intended audience and persuading them to participate.** An evaluation of the various components of the marketing effort provides information that may be useful when planning outreach programs for the next survey.

3. **When working with other agencies, the documents setting up the agreement between the agencies need to have roles and responsibilities for all phases of the project fully explained, including the timeline, production and content of jointly-produced reports and press releases.**

4. **Decide at the beginning of the project what reports will be produced.** This planning process at the beginning will help resolve numerous survey

design issues later. For example, wanting to produce a report by country of birth for recent arrivals impacts the sampling design.

Data Collection

1. **Randomize sample lists.** Randomize, or shuffle, the sample lists before providing them to the survey vendor. The vendor may release the sample records in the order that they were provided.

2. **Avoid inserting questions in area with skip patterns.** It is dangerous to insert questions into the CATI system in an area with skip patterns. If all the skips are not caught, then some unexpected situations may occur and the questions will be asked or not asked for respondents that are not planned.

3. **Check wording in category.** Use cognitive interviews that include some of the more complex questions to assess how people will interpret the question. Such procedures prevent including questions that are confusing or are open to multiple interpretations.

4. **Modification of questions by vendors.** Monitor if the persons doing the interviewing are changing the wording of some of the questions. Changing the wording of the question may change the response. This occurs when the interviewer explains or interprets the question to the respondent. When this occurs the questionnaire should be set aside for review by the team. Another recommendation would be to prepare more "canned" explanations for the interviewer to use when asked to interpret a question.

5. **Develop and use survey questionnaire pilot testing protocol.** The pilot version should have instructions to the vendor that indicate how many to call, where to call, whom to call, and what to ask so that phone numbers from a geographic area of interest will not be exhausted and will still be able to be included during the actual study. The pilot test of the 2002 survey was done by the vendor with limited instructions from the SHAPE project team. The vendor drew respondents from a particular SHAPE 2002 geographical area. As a result most of the phone numbers from that area were exhausted and were not available for the final SHAPE sample.

6. **Vendor should deliver all original data when asked.** Be clear with the vendor the expectations of when the various deliverables (e.g. raw data file) are due.

7. **The vendor should record the time it took to complete each question on the survey as well as the entire survey.** SHAPE 2002 did receive "time to complete" data for the pilot study but this information was not delivered for the actual survey.

8. **The vendor should document and deliver the final disposition for in-person surveys, including those who could not be reached, those who refused, and those who were unable to answer the survey due to physical or mental reasons or language issues.**

9. **Starting the conversation in a native language (e.g. Spanish) seemed to help increase the participation of those whose primary language is not English.**

10. **Filtering the list of names and phone numbers to select those of a suspected group of interest (e.g. Hmong. Latino. etc) helped to increase the number of respondents from that group.** Culling out persons from the list of available phone numbers with names that are likely to come from the group of interest increases the probability of finding persons from that group. A culling process reduces the number of phone calls needed compared to a purely random system and avoids implementing a screening process that would complicate the weighting calculation.

11. **Add an administrative question to document the interview language used in the survey.** The language used in the survey was not documented via an administrative code for each of the respondents of SHAPE 2002. The variable listing the language used in the interview was derived from a combination of questions related to the interviewer rather than the respondent. For example, all respondents who were interviewed by the Somali interviewer were classified as being done in Somali. The variable "Hmong language used for interview" was not as clean as was expected since the Hmong interviewer(s) conducted surveys using both English and Hmong, although the vendor reported that most of the interviews completed by the Hmong interviewer(s) were done in Hmong.

Management of SHAPE Data

1. **It's important to take the time to explore the raw data early to look for inconsistent or improbable responses.** Such procedures result in cleaner data and exclude cases whose responses imply that the respondent did not understand the questions or was not giving consistent answers.

2. **Ensure that the budget for the project includes adequate funding for the dissemination of all the reports that will be published.**

Weighting the Data

1. **Add equal probability and post-stratification weights only as much as needed.** Overfitting the data to resolve one problem may cause other problems to appear elsewhere.

2. **Document all the steps used in the weighting process to allow one to go back and confirm the mathematics of each step.**

Things to Definitely do Again

1. **Use of GIS/maps.** The maps from the 2000 Census showing the location of the various sub-populations (e.g., American Indians, etc.) were very helpful in planning how to define the survey strata.

2. **Work with community groups.** Community groups have much to add to a successful project. Actively including researchers and community leaders enhances the quality of all aspects of the project and helps achieve the proper balance between traditional and participatory research.

3. **Breakout work groups were good for dividing up tasks so that people can contribute in those areas where they are most knowledgeable and that multiple tasks can occur concurrently.** For example the Coding work group, the Data Cleaning work group and the Weighting work group all worked in parallel and avoided having everyone involved in every decision.

4. **Plan for alternative modes of administration.** Contacting people by telephone worked for the majority of Hennepin County residents but proved inadequate for American Indians and African-born Blacks. Future surveys are likely to need additional techniques to reach an increasingly mobile and heterogeneous population.

HENNEPIN COUNTY
COMMUNITY HEALTH DEPARTMENT