PILOT OF THE NATIONAL COST OF CANCER CARE SURVEY

M. Sirken and P. Royston, National Center for Health Statistics
R. Warnecke, E. Eastman, and R. Czaja, University of Illinois
D. Monsees, National Cancer Institute

INTRODUCTION

The rapidly increasing costs of health care have created a need for public health programs to deal with the devastating financial effect of serious illness on families. To develop sound, cost-effective programs, national estimates of medical expenditures must be obtained. However, existing survey methods have not been able to provide these estimates because of difficulties in identifying a national sample of patients with a serious illness and obtaining verifiable and accurate reports of the various direct and indirect costs of care for these patients.

Despite design difficulties, the National Cancer Institute has found that accurate estimates of the cost to the nation for cancer care are required for program planning and evaluation. Accurate baseline data are needed against which to assess the cost benefits of early detection at various sites and various modes of therapeutic and rehabilitative interventions. The data are also needed to evaluate the current use and effectiveness of proven interventions as well as to assess the need for new interventions or for programs to disseminate the current "state-of-the-art." Such data will be instrumental in developing program priorities both at the National Cancer Institute and in other programs concerned with the financing of catastrophic illness.

The National Cancer Institute is presently funding a series of survey experiments to test and develop design protocols for a national survey of cancer care costs. The survey experiments were designed by statisticians at the National Center for Health Statistics and are being carried out by the Survey Research Laboratory at the University of Illinois. This paper describes these survey experiments.

Conducting national surveys to estimate disease costs involves identifying a national probability sample of diagnosed cases of the disease, and collecting the cost information for the national sample of diagnosed cases. The first phase, identifying a national probability sample of cases, presents difficult design problems which vary somewhat depending on the survey strategy that is adopted. Essentially, two types of survey strategies are available for identifying a national sample of diagnosed cases. They differ with respect to the sampling frame that is used in the survey. One strategy is based on a sampling frame of medical providers. The other strategy is based on a household sampling frame.

The first stage in a medical provider survey includes soliciting the cooperation of the providers, such as hospitals, and screening the provider's medical records to identify diagnosed cases of the illness. If hospital records are used as the sampling frame, it is usually necessary to obtain the attending physician's permission to contact the patient. Once permission is granted, the patient is asked to take part in the panel cost survey.

The household survey strategy is quite different. A sample of households is interviewed to identify household members with the illness. All patients identified in this way are asked to participate in a panel cost survey. During the panel survey, the patient is asked to identify the physician who made the diagnosis, and this physician is contacted to confirm the diagnosis.

It might appear as if the survey of medical providers would surely be a more efficient design for identifying patients than a household survey since the patients could be more easily located, and their diagnoses more easily determined from their medical records than from reports of household respondents. However, several efforts to estimate costs by conducting surveys of patients identified through hospital files were essentially unsuccessful [1,2,3]. Although more than 80 percent of the hospitals cooperated by permitting their files to be screened for patients, nearly all cooperating hospitals required permission from the attending physicians and the patients before releasing the patient identifiers that would be needed to conduct the cost of care survey. Consequently, interviews were completed with fewer than 30 percent of the eligible patients.

The technical problems encountered with hospitals has generated renewed interest in the possibility of screening patients by means of household surveys. However, household surveys of disease incidence designed along traditional lines are also known to have serious technical problems. Estimates based on traditional household surveys are subject to large sampling errors because most serious diseases have relatively low prevalence rates and the statistics are subject to large biases due to coverage errors and reporting errors. The reasons for the biases are not well understood but it is known that some patients are missed because household surveys exclude the institutional population, and it is believed that some patients are sensitive about disclosing their serious illnesses and do not report them.

In view of the technical problems associated with screening patients by household surveys based on traditional sampling methods, this project proposes to test the design of household surveys that are based on network sampling. The essential difference between traditional sampling designs and the network sampling designs is the counting rule [4]. This rule specifies the households at which patients can be reported in the survey. Network sampling allows patients to be reported at the residences of close relatives, such as spouse, children, parents, and siblings, as well as at their own residences. Thus, in a household survey to identify persons with a serious illness, households would be asked to identify household members with the illness and also close relatives with the illness living elsewhere. These relatives are then interviewed to confirm the report of illness, and to collect information, such as medical care costs, which is available only from the patient.

There are several possible advantages of the proposed network sampling design but some potential problems as well.

The possible advantages of the network sample design are illustrated by the findings of a national network survey of diabetes prevalence [5]. In the 1976 Health Interview Survey (HIS), diabetics were eligible to be reported by their siblings, children and parents as well as by themselves. Prevalence estimates of diabetes were compared based on traditional sampling and network sampling. The estimates based on network sampling were significantly larger and had substantially smaller relative sampling errors. The sampling errors were smaller because more households were eligible to report diabetics. It is believed that the estimates based on network sampling were larger and apparently more accurate for two reasons: (1) institutionalized patients were enumerated; and (2) relatives not living with the patients were more willing to report them than the patients were willing to report themselves.

The potential disadvantages of the network sample design include: (1) it requires collecting supplementary weighting information from the sample households; and (2) if information is needed which is only available from the patient, an additional interview must be conducted with the patient's household. In surveys to estimate disease incidence and prevalence, such as the Diabetes Survey, the patient interview is not necessary, but in a survey of costs of illness, the patient must report the illness and must agree to participate in the cost survey.

The proposed strategy for a national network survey of cancer costs involves three data collection stages:

STAGE 1: As a part of the ongoing National Health Interview Survey (NHIS), a national sample of households will be interviewed using the NHIS questionnaire and a cancer supplement which identifies cancer patients in the household and close relatives living elsewhere who have cancer. Relatives with cancer identified in the HIS will then be interviewed to confirm the report of cancer.

STAGE 2: All persons reported as having cancer by their own households will be asked to participate in a four-round panel survey to collect data on health care and costs for a one year period. During the panel survey, names and addresses of medical providers will be obtained.

STAGE 3: A survey of medical providers will be carried out to validate and supplement the diagnostic, utilization, and cost-of-care data reported in the panel survey.

Four experiments are being conducted in the pilot study to develop the protocols for the three stage national survey design. Flow charts of the National Survey design and the Pilot Study design are shown in Figures 1 and 2, respectively. The four Pilot Study experiments are described in the subsequent sections.

EXPERIMENTS 1 and 2

Experiments 1 and 2 both relate to the household survey. They will address complementary aspects of the same issue. Experiment 1 is designed to evaluate the degree to which cancer patients will be reported in their own households using a general household health interview format. A second purpose is to obtain names and addresses of close relatives of cancer patients to interview in the second experiment.

Experiment 1 will be carried out with a sample of approximately 600 households. Three hundred and twenty-four households will be drawn from two regional cancer registries and a sample of approximately 270 "decoy" households will be selected from the same geographic areas.

To acquire the necessary cases from the registries, the Survey Research Laboratory (SRL) will collaborate in this study with the Illinois Cancer Council (ICC), which is one of the 21 designated Comprehensive Cancer Centers in the United States. Because of the sensitive nature of the identifying information about patients, the data center at ICC will select and control both the patient and the decoy sample. As a result, the patients and decoys will be indistinguishable to anyone connected with the experiment.

The registry sample will be grouped into three strata based upon geographic location, nine cancer site groups, and three diagnostic periods. The nine cancer groups are: colon/rectum, breast, cervix/corpus/ovary, prostate, kidney/bladder, leukemia/lymphoma, oral/larynx, melanoma, and miscellaneous (all other sites). The three one-year diagnostic periods will be August 1977 through July 1978, August 1978 through July 1979, and August 1979 through July 1980. To the extent possible, the cases will be equally distributed between the following geographic strata: the Chicago SMSA, other Illinois SMSA's, and non-SMSA counties. The registry sample is a 3 x 9 x 3 design and four cases will be selected for each cell. The only exception will be the cells for the miscellaneous sites. Due to the short median survival time of these cases, the miscellaneous site will contain 12 cases from each geographic stratum diagnosed in the period February 1980 through July 1980. Thus, the total number of cells will be 75.

In selecting the decoy cases, emphasis will be placed on the location of decoy cases rather than on meeting strict probability sampling criteria normally applied to surveys. In selecting the decoy sample, the main criterion will be to select cases located within reasonable proximity to the selected cancer patient households. Decoy addresses will be selected randomly from one of the following sources; reverse telephone directories, city directories, or recent listings of households from area probability studies.

Personal interviews will be conducted with both registry cases and decoys using the most recent version of the National Health Interview Survey core questionnaire and a supplement designed specifically for this survey. The HIS core questionnaire obtains information for each household member about personal and demographic characteristics, illness, injuries, impairments, and chronic conditions, and the kinds of health services received. The supplement, which has been developed for the National Cost of Cancer Survey and is being tested in the pilot, is designed to identify and obtain descriptive data about cancer patients

living in the household and about relatives with cancer living in other households. The first section, labeled "Screening Page," asks about household members with cancer, and parents and siblings living elsewhere with cancer. Three probe questions are used; the first asks about malignant tumors or growths or any type of cancer, the second asks specifically about Hodgkin's Disease, lymphoma and leukemia, and the third asks about a borderline or mild cancer or one that was cured. If there is a recent widow or widower in the household, a similar series of questions is asked about the deceased spouse. Appendix 1 shows the sections of the "Screening Page" which identify household members and parents with cancer.

Following the screening section is a section called the "Patient Page" which collects a few items of information about each cancer patient reported in the Screening section. Here, questions are asked on the name and vital status of the patient and on the site and date of diagnosis of the cancer. To weight the reported cancer cases appropriately and to locate the relatives with cancer for interview in the panel cost survey, it is necessary to collect two additional pieces of information; the size of the patient's network, i.e., the number of relatives living elsewhere who are eligible to report the patient, and the names and addresses of relatives living elsewhere with cancer. Appendix 1 shows portions of the "Patient Page" section.

In Experiment 1, a third section, called the Extended Family Page, is included which will not appear in the proposed National survey. This section serves to develop the sampling frame for Experiment 2 by obtaining for each cancer patient living in the household, the names and addresses of one randomly selected sibling and child living in Illinois.

Experiment 2 is designed to test the reporting level and accuracy of various relatives of cancer patients to determine whether network sampling will be an improvement over traditional sampling. A second aim is to determine which combination of relatives will provide the most efficient network sampling frame.

In Experiment 2, HIS interviews will be conducted with relatives of cancer patients interviewed in Experiment 1, without revealing the foreknowledge that they are relatives of known cancer patients. A sample of approximately 200 relative households will be selected for this experiment. Insofar as possible this sample will include an equal number of households of siblings and children of cancer patients. One hundred and seventy decoy households will be selected for Experiment 2. The rules for selecting decoy households will be the same as those used to select the decoys for Experiment 1.

Data collection procedures and instruments used for Experiment 2 will be virtually identical to those used in Experiment 1, with one exception; the Extended Family Section of the supplement will be omitted.

EXPERIMENT 3

Experiment 3 is a pilot for the panel cost survey. Procedures will be evaluated for obtaining cooperation of cancer patients for the panel survey. Also, the effects of bounded and unbounded recall and varying lengths of recall periods will be studied to determine the optimum design for minimizing recall errors and costs. The panel study will involve two rounds of data collection at three month intervals.

Approximately 200 cancer patients identified in their own households in the Experiments 1 and 2 HIS interview will comprise the sample for Experiment 3. These 200 patients will be mailed a letter and calendar shortly after the HIS interview which notifies them of their selection for the panel study, and asks that they record all medical care on the calendar and retain all bills until the interviewer visits them.

The same core questionnaire will be used in both rounds of Experiment 3. The instrument is modeled after that used in the National Medical Care Utilization and Expenditure Survey (NMCUES), which collects medical care utilization and cost data for the general population. The pilot study instrument is designed to collect certain information for cancer patients which can be compared to the NMCUES data for the general population, such as hospital, physician, and clinic charges and costs for prescription and nonprescription medicine. To estimate a total cost for cancer care, however, cost items are needed in the Cost of Cancer Survey which are not in NMCUES, but which may be a substantial expense for cancer patients. These include certain nonmedical costs, such as transportation for medical care, counseling services, hired help, home modifications,

The first cost interview will have a variable recall period, which will depend on the length of time since diagnosis. For those cases which were diagnosed less than one year before the HIS interview, costs since diagnosis, and probably some prediagnosis costs, will be collected in the first cost interview. For all other patients, the recall period will cover the time between the HIS interview and the first cost interview, which will be three to four months. The recall period for the second panel cost interview will be the same for all patients. Interviews will be conducted approximately three months after Round 1 and will cover all utilization and costs in that three month period.

Each panel respondent will be asked to sign an Authorization form permitting the medical providers listed in the survey to be contacted for further information.

EXPERIMENT 4

This experiment is a pilot for the Medical Provider Survey. Its primary purpose will be to validate diagnostic, utilization, and direct cost-of-care data reported by cancer patients participating in the Experiment 3 panel survey by conducting a record check survey with their medical providers. A secondary purpose will be to obtain supplementary diagnostic data from the medical providers with which to determine the certainty of the diagnosis.

The survey of medical care providers will include a maximum of 600 health care providers from whom patients reported receiving medical care in Experiment 3. A Medical care provider can be an individual, facility or institution.

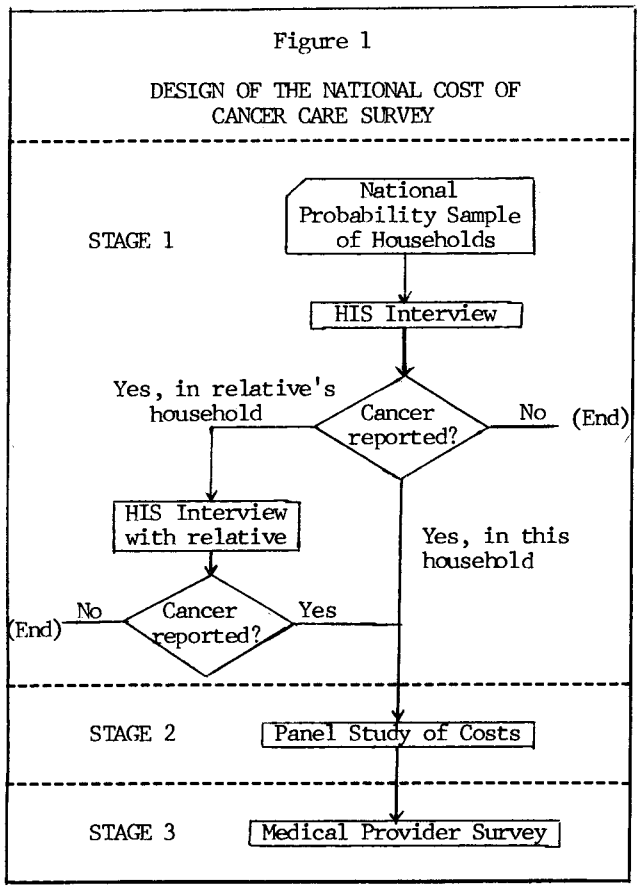
The survey instruments are patterned after those developed for the Medical Provider component of the NMCUES survey. The initial series of questions will collect descriptive data about each medical provider. Then, in order to evaluate the patient data, medical care providers will be asked the same questions regarding diagnosis, type of care provided and cost-of-care as those in the panel survey instrument. Finally, the Medical Provider Survey instruments will gather detailed data related to diagnosis and care provided which panel respondents would not necessarily know or be able to recall accurately, particularly data which will determine the certainty of the diagnosis.

SUMMARY

Four survey experiments have been planned as a pilot for a National Cost of Cancer Care Study. Experiments 1 and 2 address complementary issues concerning the effectiveness of a network procedure in identifying persons with cancer. In Experiment 1, households of known cancer patients will be interviewed to see whether the cancer will be reported. In Experiment 2, close relatives of these known cancer patients will be interviewed to see whether they will report the relative with cancer. Two hundred of the cancer patients who report the condition in Experiment 1 will be asked to take part in the Experiment 3 panel cost survey. Two rounds of cost interviews will be conducted at three month intervals to collect medical care utilization and cost data. Finally, medical providers for these patients will be contacted for information to verify the medical care and costs reported by the patients, and to assign a degree of certainty to the diagnosis.

Figure 1

DESIGN OF THE NATIONAL COST OF CANCER CARE SURVEY



References

[1] Kalsbeek, W.D., et.al. "Final Report: Survey of the Incidence, Prevalence, and Costs of Head and Spinal Cord Injury", July, 1977.*

[2] Eldred, C.A., et.al. "A Report of the Survey of Intracranial Neoplasms", June, 1977.*

[3] Robins, M., et.al. "Nationwide Study of Stroke: Final Report", December, 1978.*

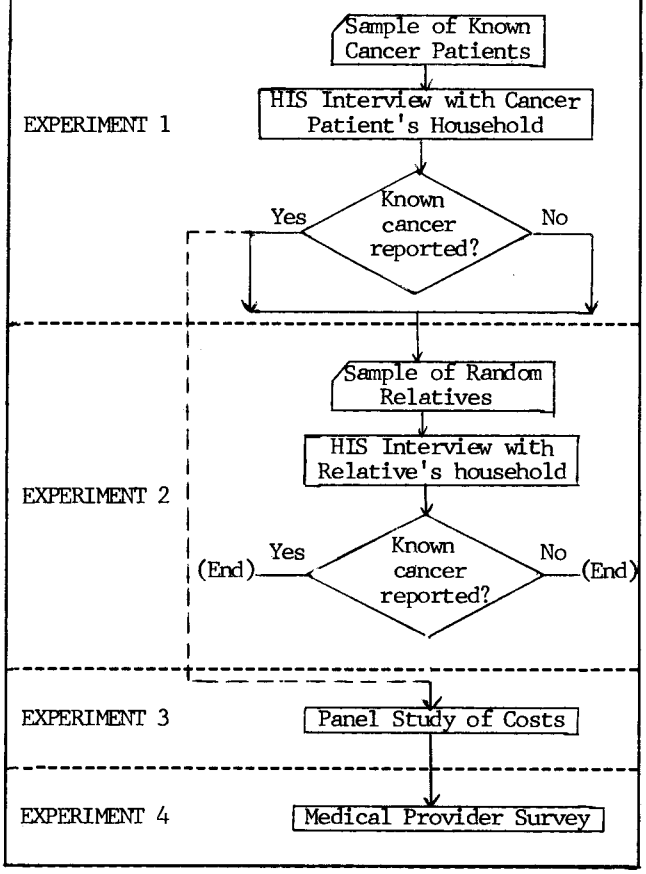
[4] Sirken, Monroe G. "The Counting Rule Strategy in Sample Surveys", Proceedings of the Social Statistics Section, American Statistical Association, 1974, pp. 119-123.

[5] Sirken, Monroe G., Graubard, B.I., McDaniel, M.J. "National Network Surveys of Diabetes", Proceedings of the Section on Survey Research Methods, American Statistical Association, 1978, pp. 631-635.

* Final Contract Reports prepared for the National Institute of Neurological and Communicative Disorders and Stroke.

Figure 2

DESIGN OF THE COST OF CANCER CARE PILOT STUDY



SCREENING PAGE

1

Record time: _____ ☐ AM ☐ PM

If "cancer" or "malignant" tumor or growth reported in C2, mark box in person's column.

1a. (Besides — —) Has anyone in the family *(read names)* **ever had a malignant tumor, a malignant growth, or any type of cancer?**

☐ Yes ☐ No *(2a)*

b. Who is this? *Mark "Tumor/cancer" box in person's column.* — 1b. ☐ Tumor/cancer

c. Has anyone else in the family ever had a malignant tumor, a malignant growth, or any type of cancer?

☐ Yes *(Reask 1b and 1c)* ☐ No

For each person with box marked in 1b, complete Item 1 on PATIENT PAGE, then return to 2a on THIS page.

2a. (Besides *(persons in 1b)*) **Has anyone in the family ever had Hodgkin's Disease, lymphoma, or leukemia?**

☐ Yes ☐ No *(3a)*

b. Who is this? Which condition did — — have? *Mark box in person's column.* — 2b. ☐ Hodgkin's Disease ☐ Lymphoma ☐ Leukemia

c. Has anyone else in the family ever had Hodgkin's Disease, lymphoma, or leukemia?

☐ Yes *(Reask 2b and 2c)* ☐ No

For each person with box marked in 2b, complete Item 1 on PATIENT PAGE, then return to 3a on THIS page.

3a. Sometimes tumors or growths can be detected before they are a serious threat to health. (Besides *(persons in 1b and 2b)*) **Has anyone in the family ever had borderline cancer, a mild form of cancer, or a cancer that was cured?**

☐ Yes ☐ No *(Go to next page)*

b. Who is this? *Mark "Borderline/mild/cured cancer" box in person's column.* — 3b. ☐ Borderline/mild/cured cancer

c. Has anyone else in the family ever had borderline cancer, a mild form of cancer, or a cancer that was cured?

☐ Yes *(Reask 3b and 3c)* ☐ No *(Go to next page)*

For each person with box marked in 3b, complete Item 1 on PATIENT PAGE, then return to S-1 on SCREENING PAGE, p. 4.

SCREENING PAGE — CONTINUED

1

S-1 *Mark first appropriate box* — S-1. ☐ Under 17 *(NP)* ☐ Mother in household *(S-2)* ☐ Mother NOT in household

4a. Is — —'s mother still living? — 4a. ☐ Yes *(4c)* ☐ No

b. Did — —'s mother die before *(12 month date)*? — b. ☐ Yes *(S-2)* ☐ No

c. Did — —'s mother ever have a malignant tumor, a malignant growth, or any type of cancer? — c. ☐ Yes *(4g)* ☐ No

d. Did — —'s mother ever have Hodgkin's Disease, lymphoma, or leukemia? — d. ☐ Yes *(4f)* ☐ No

e. Did — —'s mother ever have borderline cancer, a mild form of cancer, or a cancer that was cured? — e. ☐ Yes *(4g)* ☐ No *(S-2)*

f. Which condition did she have? — f. ☐ H.D. ☐ Lymph. ☐ Leuk.

g. *Complete Item 1 on PATIENT PAGE, then return to S-2 on THIS page.*

S-2 *Mark appropriate box* — S-2. ☐ Father in household *(NP)* ☐ Father NOT in household

5a. Is — —'s father still living? — 5a. ☐ Yes *(5c)* ☐ No

b. Did — —'s father die before *(12 month date)*? — b. ☐ Yes *(NP)* ☐ No

c. Did — —'s father ever have a malignant tumor, a malignant growth, or any type of cancer? — c. ☐ Yes *(5g)* ☐ No

d. Did — —'s father ever have Hodgkin's Disease, lymphoma, or leukemia? — d. ☐ Yes *(5f)* ☐ No

e. Did — —'s father ever have borderline cancer, a mild form of cancer, or a cancer that was cured? — e. ☐ Yes *(5g)* ☐ No *(NP)*

f. Which condition did he have? — f. ☐ H.D. ☐ Lymph. ☐ Leuk.

g. *Complete Item 1 on PATIENT PAGE, then return to S-1 on THIS page for next person.*

PATIENT PAGE

1a. Patient lives in this household:

☐ Yes *Enter name* ↙

_____ *(Go to 1b)*

☐ No *Enter col. no. and name of person in this household to whom patient is related.* ↙

_____ (Col. no.) _____ (Name)

Patient is this person's:

☐ Brother ☐ Mother ☐ Husband
☐ Sister ☐ Father ☐ Wife

Patient is: ☐ Living ☐ Deceased

b. Cancer reported as:

☐ malignant tumor, growth or cancer
☐ Hodgkin's disease
☐ Lymphoma
☐ Leukemia
☐ Borderline, mild or cured cancer

→ _____ (Volunteered information on site or type of cancer)

↙

Does (*patient*)'s condition have a specific name?

☐ Yes _____ (Name of condition) ☐ No ☐ DK

Read to respondent
Now I'd like to ask just a few questions about (*patient*).

If patient lives in household, go to 3a. Otherwise ask:

2. What is (*patient*)'s name?

_____

If "leukemia" in 1b, go to 4.

3a. What part of the body [is/was] affected by (*patient*)'s (*condition in 1b*)?

☐ Skin
☐ Other *(Specify)* _____ *(Go to 4)*

Hand Card A

b. There are several different kinds of skin tumors or growths. Please look at this card and tell me which kind of skin condition (*patient*) [has/had]. [Is/Was] it —

1 ☐ Squamous cell carcinoma
2 ☐ Basal cell carcinoma
3 ☐ Bowen's disease
4 ☐ Paget's disease
5 ☐ Malignant melanoma
6 ☐ Some other kind of skin condition *Specify*

_____

4. In what month and year was (*patient*) first told by a doctor that (*patient*) had (*condition in 1b*)?

_____ (Month) _____ (Year) ☐ DK *(5a)*

6a. How many living brothers and sisters 17 years old or older does (*patient*) have?

____ Living 17+ brothers and sisters ☐ None *(7a)* ☐ DK *(6d)*

Hand Card B

b. How many of these (*number in 6a*) brothers and sisters live in any of the places shown on that card?

____ In institution ☐ None ☐ DK

c. How many of (*patient*)'s (*number in 6a*) brothers and sisters 17 years old or older are now on full time active duty with the Armed Forces?

____ In Armed Forces ☐ None ☐ DK

d. (Not counting the brothers and sisters [living in any of the places on that card/(and) in the Armed Forces]) How many of (*patient*)'s brothers and sisters 17 years old or older live in Illinois?

____ Live in IL ☐ None *(7a)* ☐ DK

e. How many of (*patient*)'s brothers and sisters 17 years old or older live in this household?

____ Live in household ☐ None

Footnotes

7a. How many living children 17 years old or older does (*patient*) have?

____ Living 17+ children ☐ None *(P-2)* ☐ DK *(7d)*

Hand Card B

b. How many of these (*number in 7a*) children live in any of the places shown on that card?

____ In institution ☐ None ☐ DK

c. How many of (*patient*)'s (*number in 7a*) children 17 years old or older are now on full time active duty with the Armed Forces?

____ In Armed Forces ☐ None ☐ DK

d. (Not counting (*patient*)'s children [living in any of the places on that card/(and) in the Armed Forces]) How many of (*patient*)'s children 17 years old or older live in Illinois?

____ Live in IL ☐ None *(P-2)* ☐ DK

e. How many of (*patient*)'s children 17 years old or older live in this household?

____ Live in household ☐ None

P-2

☐ Cancer patient is household member *(Go to Next Patient)*
☐ Cancer patient is deceased *(P-3)*
☐ Other *(9a)*

P-3 Is the spouse of this deceased patient living in this household, i.e. is "widowed" box marked?

☐ Yes *(Go to Next Patient)* ☐ No *(Ask 8)*

8a. At the time (*patient*) died, was (*patient*) married?

☐ Yes ☐ No *(8g)* ☐ DK *(8g)*

b. Is (*patient*)'s [husband/wife] still living?

☐ Yes ☐ No *(8g)* ☐ DK *(8g)*

c. Has (*patient*)'s [husband/wife] remarried?

☐ Yes *(8g)* ☐ No ☐ DK

Hand Card B

d. Does (*patient*)'s [husband/wife] now live in any of the places shown on that card?

☐ Yes *(8g)* ☐ No ☐ DK

e. Is (*patient*)'s [husband/wife] now on full time active duty with the Armed Forces?

☐ Yes *(8g)* ☐ No ☐ DK

f. Does (*patient*)'s [husband/wife] now live in Illinois?

☐ Yes *(8g)* ☐ No *(8g)* ☐ DK *(8g)*

g. *Go to Next Patient.*

9a. We will be conducting interviews in a small sample of households with people who have had serious illnesses. Could you please give me (*patient*)'s name, address and telephone number so that — —'s household could be contacted, if it is selected?

(Under no circumstances will the interviewer or anyone connected with this survey tell anyone in (*patient*)'s household about (*patient*)'s illness. The interviewer will not know anything about the health of the people in that household before she conducts the interview.)

Name _____

Number and Street _____

City _____ State _____

( ) _____

Area code Telephone Number