

Table of Contents

Introduction.....	1
Information about the first wave of the project	1
Versions of the Questionnaire.....	2
General Considerations.....	4
Use of a proxy	4
Denomination of the Spouse Variables.....	4
Analysis of Couples	5
Notes on the Minimental.....	5
Age Variables.....	5
Weighting the Data	5
Missing Values.....	6
Data Files and Related Documents	7
General Notes about the Structure of the Rosters	8
The Household Roster: Available Variables.....	8
The Children Roster: Available Variables.....	9
The Siblings Roster: Available Variables.....	10
The Parents Roster: Available Variables	10
The Other Persons Roster: Available Variables	10
Variables of Position and Variables of Duplication	11
Utilization of Rosters' Data	12
Calculated Variables	13
Errors and Limitations Identified in Data	14
Codebook	15

Introduction

The research project Puerto Rican Elderly: Health Conditions (*Condiciones de salud de los adultos mayores en Puerto Rico*), known as PREHCO, is a study funded with federal funds from the National Institute on Aging / NIA (Grant number RO1 AG1622090-01) through a contract between the University of Wisconsin-Madison and the University of Puerto Rico. This project also benefits from the technical support of the Pan American Health Organization (Washington D.C.) and has received economic support from the Legislature of Puerto Rico and the Office of the Women's Advocate in Puerto Rico.

In its first wave, between May 2002 and May 2003, we carried out a cross sectional study of the non-institutionalized elderly population 60 years or older in Puerto Rico, which serves as the database for the study of this age group. Its codebook is explained in this document.

We used a multistage probabilistic sample by clusters, using as a sample frame from the 2000 Population and Household Census of the Census Bureau of The United States.

A questionnaire that included the following sections was designed for the project: health conditions, physical and mental disability, use of medication, health and social services needs, access to and use of health services, abuse, migration, living conditions, help and assistance patterns of the family, community and public and government agencies, as well as others.

Information about the first phase of the Project

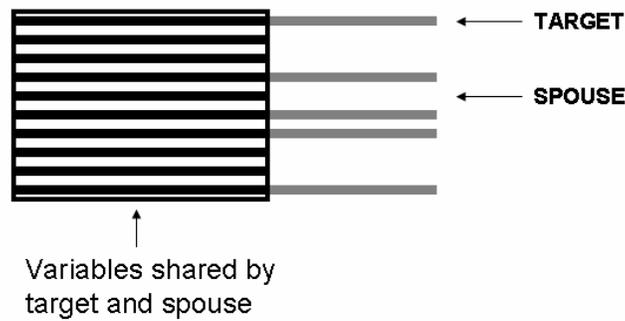
In the first phase of the PREHCO project, 4,291 elderly persons 60 years or older and 1,442 spouses were interviewed, the latter independent of their age. With the data obtained in these interviews, we created a database that has created in the same record the information of each one of the 4,291 participants and that of their spouses when interviewed.

Figure 1. Structure of the Database of PREHCO1 (n=4,291)



Even though this codebook presents the structure of that database, since the interviewee and his/her spouse answered the same questions in many sections of the questionnaire, it is also possible to construct, in order to increase the sample size, a database that includes the targets and their spouses (60 or older) as separate records. This new database would have a total of 5,333 records (Figure 2), and, as previously mentioned, would include only those variables common to both types of interviews and those spouses who are 60 years or older. There is a SPSS syntax file that permits us to automatically create this database from the 4,291 cases (see section titled "Data Files and Related Documents".)

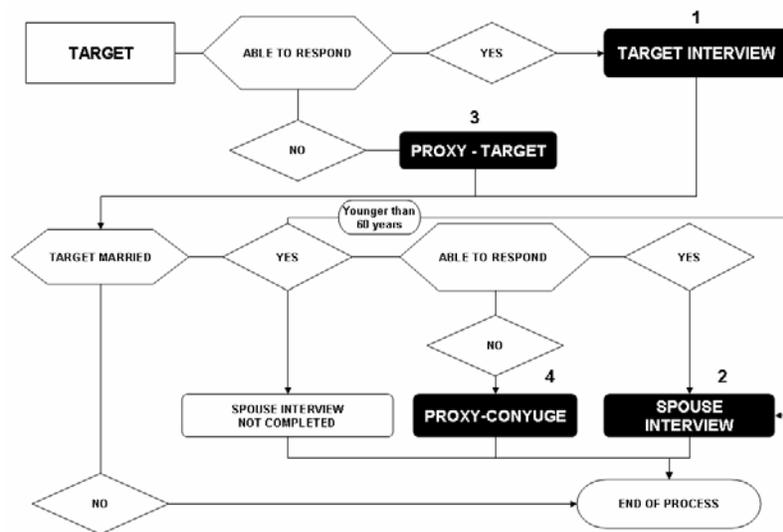
Figure 2. Structure of the PREHCO database (n=5,333)



Versions of the questionnaire

The main questionnaire changes according to various factors. These include principally the capacity of the interviewee to respond, the existence of a spouse, and the capacity of the spouse to respond (see Figure 3).

Figure 3. Flowchart used to determine the type of interview



To summarize, there are four main versions of the questionnaire (see Table 1); these are: (1) target questionnaire, (2) spouse questionnaire, (3) questionnaire of the proxy¹ for the target, and (4) questionnaire of the proxy for the spouse.

1. **Target questionnaire:** the principal questionnaire is the one administered to the older adult who is capable of answering the questionnaire on his/her own. This version includes all the sections of the questionnaire.

¹ Criteria for qualifying as proxy: he/she must have known the interviewee for at least one year and must be at least 18 years of age.

Table 1. Versions of the Questionnaire

1 Target	2 Spouse	3 Proxy-Target	4 Proxy-Spouse
Section A: Personal Data	Complete	Complete	Complete
Section B: Minimental	Complete (if spouse is 60 years or older)	Complete (if spouse is the proxy and older than 60 years) Reduced (if proxy is another person and is older than 60 years) or questions	No (if the proxy is the target) Reduced (if the proxy is another person and is older than 60 years and has not previously acted as proxy of the target)
Section C: Household roster		Complete	
Section D: Personal Data	D1, D15 to D20, D31 to D33	D1 to D25, D27, D28	D1, D15 to D20
Section E: SES Steps	Complete		
Section G: Health Status	Complete	Only G4-G67, G109, G114ck, G114-G131, G133-G139, G141, G143-G144, G147-G156, G161-G170	Only G4-G67, G109, G114ck, G114-G131, G133-G139, G141, G143-G144, G147-G156, G161-G170
Section H: Childhood conditions	Complete		
Section I: Functional Status	Complete	Section Z	Section Z
Section J: Medications	Complete	Complete	Complete
Section K: Access and Utilization	Complete	Complete	Complete
Section L: Family Members data		Complete	
Section M: Transfers			
Section N: Labor History			
Section O: Migration	Complete		
Section P: Household		Complete	
Section Q: Properties		Complete	
Section R: Race	Complete		
Section S: Abuse	Complete		
Section T: Sexuality			
Section U: Measurements	Complete	Complete (if possible)	Complete (if possible)
Section V: Contacts			
Section X: Perception of the Interviewer	Complete	Complete	Complete

2. Spouse questionnaire: the questionnaire of the spouse is the one administered to the spouse of the older adult capable of answering the questionnaire on his/her own. This questionnaire does not include some of the sections of the principal questionnaire and also excludes some questions in the included sections.
3. Proxy for the target questionnaire: this questionnaire is answered by a proxy, in the cases in which the older adult is not capable of responding on his/her own. The determination of whether it is necessary to have a proxy is carried out using the following criteria: (1) the score obtained by the interviewee in the minimal and (2) the opinion of the interviewer in some cases². The proxy provides information relative to the older adult, excluding some sections or individual questions that imply opinion or state of mind. In some cases the spouse acts as the proxy for the interviewee and in others, in which there is no spouse or the spouse is not capable of responding, the proxy is another person.
4. Proxy for the spouse questionnaire: this questionnaire is completed by a proxy, in the cases in which the spouse of the older adult is not capable of responding on his/her own (only for spouses 60 or older). The determination of the necessity for a proxy was realized with the criteria previously outlined. The proxy provides information relative to the spouse of the older adult, excluding sections or individual questions that imply opinion or state of mind. In some cases the target acts as a proxy for the spouse and in others the proxy is another person.

General Considerations

Below are some general considerations that should be taken into account before using the database.

Use of a Proxy

13.5% of the targets were interviewed through a proxy. There are two variables that can be used to filter, if necessary, those interviews: CODE or PROXY. The CODE variable has the 25 (for regular interviews) and 26 (for proxy interviews) values. The PROXY variable has the values 0 (when there is no proxy), 2 (when the proxy is the spouse), and 3 (when the proxy is another person). Some characteristics of the proxy are also reflected in the variables: PRXYAGE (age), PRXYSEX (gender), PRXYRL_1 (general relationship with the interviewee), and PRCYRL_2 (specific relationship)³.

Denomination of the Spouse Variables

The variables corresponding to the spouse of the interviewee maintain the same nomenclature of those of the interviewee, but are preceded by the letter **Y**. So, for example, the variable for general health status is G1 for the target interviewee and YG1 for his/her spouse.

² It was established as criteria that a total score on the minimal of 11 or more indicated that the interviewee was capable of responding to the questionnaire on his/her own. A score of less than 9 indicated that it was necessary to use a substitute informant. In those cases in which the score was 9 or 10 the interviewer was the one who determined whether the interviewee was capable of responding.

³ For more information on these variables consult page C-1 of the questionnaire.

Analysis of Couples

In order to carry out the analysis of couples, it is possible to use two variables as reference. The variable SPOUSE=1 indicates that there was a couple at the moment in which the interview was realized. However, the variable CONYUGE=1 is more useful, since it indicates that the interview of that spouse was completed.

Notes on the Minimental

As indicated in Table 1, Section B of the questionnaire contains a brief mental examination or minimental. The minimental also works as a filter used to determine the need for a proxy, following the criteria previously explained. In the cases in which the target's impediment or incapacity to answer the questionnaire on his/her own was obvious, the minimental was not carried out and a proxy was chosen directly. These cases can be identified using the variable MINRVAL (MINRVAL=1). The participants who demonstrated an obvious impediment do not have points (FINALSCR) in the minimental. In some cases the interviewee had a physical impediment (BRCAP=1) that inhibited him/her from completing some of the tests on the minimental that require skills such as drawing or folding. In those cases, the total points on the minimental were obtained by extrapolating the points obtained on the questions that were answered to the entire set of questions and tests included in the minimental.

Age Variables

In some parts of the interview the age at which a specific event took place is recorded. Initially, the interviewee is directly asked the age. If the age is unknown, the interviewee is asked for the month and the year. Then a final age is calculated for that event that indicates the age supplied by the interviewee or, in the event that it's faulty, the age calculated based on the year and the month. So that, for each age attributed to each event, there is a series of four variables. The variables for all these series have the following format: *_E (age supplied by the informant), *_A (year), *_M (month), and *EDAD (age summary: supplied age or calculated age). This means that in order to analyze the age in which a specific event took place, the last variable is the one that should be used, since it reflects the most complete information. In the following example, please find the variables that record the age at which the first union of the interviewee ended:

d10_e (supplied age)
d10_a (year)
d10_m (month)
d10age (age summary: supplied or calculated)

Weighting the data

The weighting factor for each record is located in the variable **FAC_T** in the database (n=4,291). Its calculation is based on a poststratification by gender and age group, considering six age groups: 60-64, 65-69, 70-74, 75-79, 80-84, and 85 years or more. If the database (n=5,333) is constructed with interviewees and spouses as separate records (see "Information about the first phase of the PREHCO Project" on page 1) a different weight factor should be used. This weighting factor is available in the file FACTORTC.SAV (see section "Data Files and Related Documents" below).

Missing Values

The entire database uses common codes for missing values. These values are presented in the following table:

Table 2. Codes and Labels for Missing Values

-1	Does not know	Interviewee indicated that he/she does not know the answer to the question.
-2	Does not answer	Interviewee indicated that he/she would not answer the question.
-3	Undetermined, data unavailable	Data not compiled due to some type of error.
-5	Does not apply to the proxy-target	The question does not apply to the proxy for the target questionnaire.
-6	Does not apply to the proxy-spouse	The question does not apply to the proxy for the spouse questionnaire.
-7	There is no spouse	The interviewee does not have a spouse or partner; therefore, there is no interview of the spouse.
-8	Does not apply to the spouse younger than 60 years of age	The question does not correspond to spouses younger than 60 years of age.
-9	Minimental does not apply	When the interviewee answers the minimental and passes the test, the variables of the minimental of the spouse and the minimental of the possible substitute informant or proxy do not have to be answered.
-10	Does not apply, obvious impediment	The interviewee has an obvious impediment (determined by the interviewer before starting the minimental test) and does not answer the questions of the minimental.
-11	Does not apply, physical impediment	The interviewee has a physical impediment and does not answer the questions of the minimental that require hand movements as with writing and drawing.
-12	Does not apply, there is no spouse interview	The target interviewee has a spouse or partner but the spouse/partner was not interviewed. The variable that identifies the final result of the interview of the spouse/partner (lycodigo) has a code that indicates the reason for which the questionnaire was not completed.
-13	Does not apply to the target or to the spouse/partner	This code only applies to section Z, which is answered by the target-proxy or the spouse-proxy.

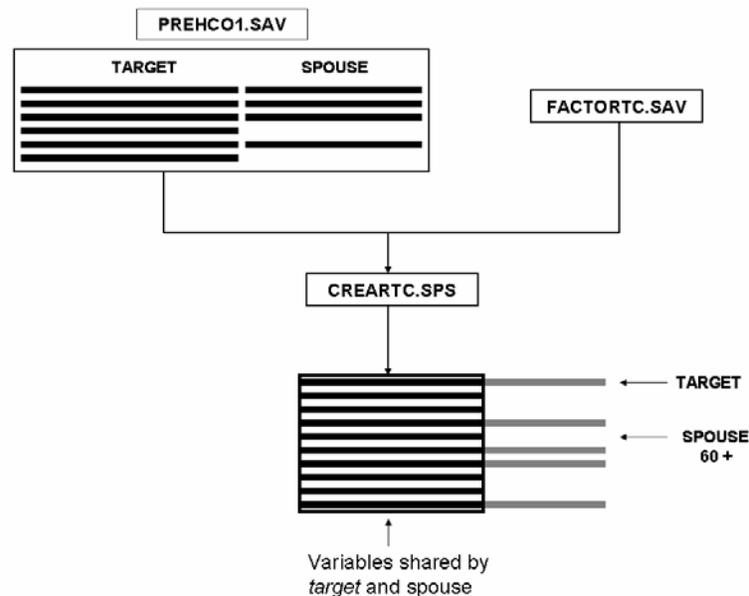
Data Files and Related Documents

The following data files are available:

1. **PREHCO1.SAV** (approximately 10Mb)
Contains all of the data of the first wave of PREHCO in SPSS format. The data is stored in 2,167 variables.
2. **creartc.sps** (19 Kb)
This SPSS syntax file automatically creates a file of 5,333 records that has the interviewees and their spouses (60 years or older) in separate records. Since both interviews share at least 700 variables, this procedure allows for working with these common variables (see above for the explanation on versions of the questionnaire) with a larger sample size. To weight the resulting file (PREHCOTC.SAV) the factor FACTC should be used.
3. **factortc.sav** (72 Kb)
This file contains the weight factor necessary to work with the database of targets and spouses (n=5,333), which can be created using the procedure CREARTC.SPS. It is also in the SPSS format.

Important: For the creation of a file with separate records for the target and the spouse using the procedure CREARTC.SPS. All three files above should be copied to a folder called C:\PREHCO1, then the file PREHCO1.SAV should be opened and the procedure CREARTC.SPS should be executed. The result will be a new file called PREHCOTC.SAV.

Figure 4. Creation of the data file of targets and spouses (n=5,333)



4. **intro_s.pdf** and **intro_e.pdf**

This introductory document, available in English and Spanish.

5. **codebook.pdf** (approximately 15Mb)

See 'Codebo9ok' on page 15. If you would like to electronically view the codebook, it is possible to obtain a copy of the program **ViewSav**⁴ which allows the viewer to see a codebook of a database in SPSS format in real time.

General Notes about the Structure of the Rosters⁵

There are two sections in the questionnaire that record information about the residents of the household and about other family members and people close to the interviewee. Section C contains the data of the household members and Section L contains four rosters: the children's roster, the siblings' roster (the three with whom the interviewee maintains the most contact), the parents' roster, and other people who are close to the interviewee's roster.

The people in each roster are assigned certain positions in the database. The household roster occupies positions between 1 and 10. In the household roster the target always occupies the position 1. The spouse, when he/she exists, does not always necessarily occupy the same position, that is to say, he/she may theoretically be in any one of the household records between 2 and 10, since the first record always corresponds to the target. This implies that, in order to identify the record occupied by the spouse, there should be a search carried out that inquires as to the relationship of the person in each record with the target: PARE_R(2-10)=2 or 4.

The Household Roster: Available Variables

Variable	Description
pos(1-10)	Position in the household roster
c4r(1-10)_1	General relationship with the interviewee
pare_r(1-10)	Specific relationship with the interviewee
c5r(1-10)_1	General relationship with the head of household
c5r(1-10)_2	Specific relationship with the head of household
sexo_r(1-10)	Gender
edad_r(1-10)	Age
c8r(1-10)	Ability to read
c9r(1-10)	Ability to write
c10r(1-10)	Schooling/Education
mari_r(1-10)	Marital status
c12r(1-10) (a-i)	Work status (series of 9 questions)
dd(1-10)	Duplication in another roster

⁴ ViewSav is a free Real-Time Codebook for SPSS files. It has been created by Karel Asselberghs, from the Department of Sociology of the University of Amsterdam and can be obtained at: <http://www.asselberghs.dds.nl/downloads/ViewSav.zip>

⁵ A detailed presentation about the rosters (ROSTERS.PDF) is available in our web site <http://prehco.rcm.upr.edu>.

The variables of the first person in the household roster will appear in the database as follows:

	caseid	c1num	c3	pos1	c4r1_1	pare_r1	c5r1_1	c5r1_2	sexo_r1	edad_r1	c8r1	c9r1	c10r1	mar_r1	ddt1
1	11050110	2	2	1	0	1	1	2	2	64	1	1	13	1	0
2	11050210	1	1	1	0	1	0	0	1	87	1	1	6	3	0
3	11050310	5	2	1	0	1	2	7	2	66	1	1	13	3	0
4	11050410	1	1	1	0	1	0	0	1	62	2	2	4	6	0
5	11050610	1	1	1	0	1	0	0	2	70	1	1	10	4	0
6	11050810	2	1	1	0	1	0	0	1	75	1	1	13	1	0
7	11050910	1	1	1	0	1	0	0	1	82	1	1	15	6	0
8	11060610	1	1	1	0	1	0	0	2	78	2	2	2	4	0

The roster of children occupies positions between 21 and 40; that of siblings occupies positions between 61 and 63; that of parents between 71 and 72; and that of other persons positions between 81 and 83.

The Children Roster: Available Variables

Variable	Description
pos(21-40)	Position in the children roster
pare_r(21-40)	Specific relationship with the interviewee
sexo_r(21-40)	Gender
l4r(21-40)	Indicates whether child is alive
edad_r(21-40)	Age at the time of the interview or age at the time of death
dond_r(21-40)	Place of residence
l7r(21-40)	Schooling/Education
mari_r(21-40)	Marital Status
l9r(21-40)	Number of Children
l10r(21-40)	Whether he/she lived or is living in the United States
l11r(21-40)	Received financial assistance from him/her to move to the United States
dd(21-40)	Duplication in the household roster

The variables of the first person in the Children Roster will appear in the data base in as follows:

	caseid	l1num	pos21	pare_r21	sexo_r21	l4r21	edad_r21	dond_r21	l7r21	mar_r21	l9r21	l10r21	l11r21	ddt1
39	11100820	2	21	5	2	1	28	1	15	6	0	3	.	3
40	11101310	2	21	5	2	1	47	5	15	4	3	1	2	0
41	11110110	1	21	5	2	1	47	4	16	4	3	3	.	0
42	11110120	2	21	5	2	1	47	4	16	1	1	3	.	0
43	11110210	5	21	5	1	1	52	6	16	6	0	3	.	0
44	11110310	1	21	5	1	2	0	2	1	0
45	11110410	2	21	5	2	1	41	1	13	6	0	3	.	4
46	11110510	3	21	5	2	1	47	4	13	2	3	3	.	0
47	11110610	2	21	5	2	1	44	1	15	6	0	3	.	3
48	11110710	3	21	5	1	1	40	4	12	1	4	3	.	0

The Siblings Roster: Available Variables

Variable	Description
pos(61-63)	Position in the Siblings Roster
pare_r(61-63)	Specific relationship with the interviewee
sexo_r(61-63)	Gender
edad_r(61-63)	Age
dond_r(61-63)	Place of Residence
mari_r(61-63)	Marital Status
dd(61-63)	Duplication in the Household Roster

The Parents Roster: Available Variables

Variable	Description
pos(71-72)	Position in the Parents Roster
pare_r(71-72)	Specific relationship with the interviewee
sexo_r(71-72)	Gender
edad_r(71-72)	Age
dond_r(71-72)	Place of residence
dd(71-72)	Duplication in the Household Roster

The Other Persons Roster: Available Variables

Variable	Description
pos(81-83)	Position in the Other Persons Roster
pare_r(81-83)	Specific Relationship with the Interviewee
l28r(81-83)	Whose child or sibling he/she is son/daughter
sexo_r(81-83)	Gender
dond_r(81-83)	Place of residence
dd(81-83)	Duplication in the Household Roster

As can be seen in the previous tables, the variables common to all rosters (see Table 3 above) have a common nomenclature. This means that the marital status of the child that occupies position 22 will be stored in the variable MARI_R22 and the marital status of the brother that occupies position 61 will be stored in the variable MARI_R61. This occurs the same way with gender (SEXO_R##), age (EDAD_R##), relationship (PARE_R##), and place of residence (DOND_R##).

Though each roster's variables maintain their name, which corresponds to the question on the questionnaire, they also include the position of that person and, consequently, a reference to his/her primary roster. Therefore, the number of children of the child who occupies position 35 will be stored in L9R35, and the number of children of the child who occupies position 22 will be stored in L9R22.

The structure of these rosters is not homogeneous; that is, not all rosters compile the same information. In Table 3 there is a breakdown of the presence and absence of a series of basic variables in each of the rosters.

Table 3. Presence of some basic variables in each one of the five rosters.

Variable	Roster				
	Household	Children	Siblings	Parents	Others
Relationship with the Interviewee	X	X	X	X	X
Gender	X	X	X	X	X
Age	X	X	X	X	no
Marital Status	X	X	X	no	no
Place of Residence	⁶	X	X	X	X

Variables of Position and Variables of Duplication

The sections of the rosters include the so-called variables of position, whose names follow the sequence POS1 to POS10 for the household, POS21 to POS40 for the children, POS61 to POS63 for the siblings, POS71 to POS72 for the parents, and POS81 to POS83 for other persons. These variables have a value corresponding to their position (for example: POS23=23) if a person exists in that position, or they are empty if no one in that position exists.

Since the household roster is not exclusive of other rosters, a person could belong to the household roster and to any other of the rosters. This duplication can only occur between the household roster and any other of the rosters. A child who resides in the home could occupy position 5 in the household roster (POS5=5) and the position 22 in the children roster (POS22=22). This is an example of a duplicated record.

Each roster, even if it has variables in common with other rosters, as was presented in Table 3, contains some information not included in the others (see sections C and L in the questionnaire.) For this reason, it is convenient to understand whether a person is duplicated and, as such, has additional information in another roster. The variables of duplication are used to identify if a record is duplicated.

The duplication variables follow the same nomenclature as those of position (DD1 to DD10, DD21 to DD40, DD61 to DD63, DD71 to DD72, and DD81 to DD83.) These variables can have a value of zero (0) when the person is not duplicated, a value corresponding to the position in which the duplicated person is found, or a missing value if there is no one in this record.

To follow the previous example of the son who is found in position 5 of the household and is duplicated in position 22 of the children roster, the variable DD5 would have a value of 22 (DD5=22) and the variable DD22 would have a value of 5 (DD22=5).

⁶ Even though there is not a variable that indicates the place in which the household members live, it is understood that, following the logic of the variables of place of residence in the other rosters, all the persons in the household will have the value 1 (Lives in the home) in order to be compared with other people.

Utilization of Rosters' Data

The information in these rosters can be used for various purposes. One of them is the analysis of the family arrangements and family structure of the interviewee, and his/her network of support. Also, it can be used for the purpose of identifying characteristics of certain people with whom the interviewee has contact. This task implies a certain complexity and requires explanation.

There are three sections in the questionnaire: section I (Functional Status), section M (Transfers), and section N (Labor History and Income), that use the information recorded in the rosters. Basically, in certain parts of these sections, the interviewee is asked about the assistance he/she receives from or provides to others. The interviewee who declared that he/she received or provided assistance for a given activity, indicates who is the person who assists him/her or whom among those in the rosters, is the recipient of his/her help. This information is codified using the number of position of that person in the rosters. If the selected person is duplicated, for example in the household roster and in the children roster, in other words, a child who lives in the home, the number indicating his/her position in his/her primary roster, the household roster, will appear. The following example illustrates this situation:

If an interviewee indicates having difficulty preparing meals for him/herself (I4=1 or I4=3), he/she will be asked about the person who helps in the preparation of meals (I4A). The value of the variable I4A indicates the position of that person in the rosters. In this way, if I4A=4, the person who helps the interviewee would also be the one in the fourth position of the household roster, his/her age would be EDAD_R4, his/her gender would be SEXO_R4, his/her marital status would be MARI_R4, his relation to the interviewee PARE_R4, and his/her place of residence, in this example would be the home. But, in addition to these common variables, there are other particular variables of the household roster in which there is additional information about the person. In this case all of these variable names will have the ending R4.

If we observe the variable DD4, the variable of duplication for that member of the household, and it has a value of 22, this would mean that the person is duplicated in the children roster in position 22. In this record of the children roster there is also additional information about that person. In this case all these variable names will end in R22.

Even though the norm is that the interviewee would choose from among the members in the rosters, in the previously mentioned questions in sections I, M, and N, the interviewee was also given the opportunity to identify other people who might not be included in any of the rosters, not even those of other people, when he/she was asked about assistance given or received. As a result, in some of the variables, the value 77 arises on occasion, indicating that the person mentioned is not in any of the rosters. The net result is that it is impossible to gather information about the characteristics of those persons, except that they are neither household members, nor children, nor siblings, nor parents. In the same way, in the questions regarding the person who provides assistance in section N, there can be values such as 99 (the interviewee does not have that expense) or 98 (all household members pay for that expense).

Calculated Variables

In the data base there is a group of calculated variables⁷ that present important data from some of the sections of the questionnaire. These variables offer the user more clear or summarized information on some of the specific data regarding the interviewees. Some of the variables are the product of the combination of two or more variables, others simply the re-codification of one variable of the questionnaire. In Tables 4 and 5 are the lists of constructed variables for the target and the spouse, indicating the section of the questionnaire to which they refer, its position in the database, its name, and its description. As with the original variables of the questionnaire of the spouse, his/her calculated variables begin with the letter “y”.

Table 4. List of calculated variables of the target in the database

SECTION	POSITION	VARIABLE	DESCRIPTION
Initial	8	fac_t	Weighting Factor
	9	targvi	Targets in the home
	10	totarget	Total targets in the home
B	64	finalscr	Minimental Final Score
	65	pasomin	Did the interviewee pass the minimental?
C	73	qjefe	Who is the head of the household?
G	642	depre_y	Score on the Depression Scale
	476	g77_code (R)*	What did the doctor tell him/her that pain or discomfort in his/her chest was?
	615	g168fisi (R)	What kind of pastime (physical activities) did he/she engage in regularly?
	616	g168soci (R)	What kind of pastime (social activities) did he/she engage in regularly?
	617	g168ocio (R)	What kind of pastime (leisure activities) did he/she engage in regularly?
H	648	code_h4 (R)	Father's occupation
	649	rcode_h4 (R)	Father's occupation recoded
	656	code_h6d (R)	Mother's occupation
	657	rcodeh6d (R)	Mother's occupation recoded
K	743	k60a_cod (R)	Health service that he/she was not able to obtain or receive
N	1188	code_n7 (R)	Current occupation (last week)
	1189	code_n9 (R)	Type of establishment or business where he/she works
	1215	code_n18 (R)	Last job
	1216	code_n20 (R)	Type of establishment or business where he/she used to work
	1228	code_n25 (R)	Occupation during most of his/her life
	1229	cod_n25b (R)	Type of establishment where he/she worked during most of his/her life
	1241	n50_code (R)	Type of health problem he/she is experiencing
O	1339	o10_code (R)	Main reason for returning from the United States to Puerto Rico
U	1460	bmi	Body Mass Index
	1461	rbmi	Body Mass Index Recoded
	1469	legstand	Result of exercise: standing on one foot
	1473	get_up	Result of exercise: stand up and walk

* (R)=Recoded from an original alphanumeric variable.

⁷ All the calculated variables include in their label the word CALCULATED.

Table 5. List of calculated variables of the spouse in the database

SECTION	POSITION	VARIABLE	DESCRIPTION
Initial	1483	conyuge	Identifies interview with spouse is complete
	1488	y marital	Marital status
	1489	yleer	Knows how to read
	1490	y escribe	Knows how to write
	1491	y escolar	Schooling/Education
B	1543	y finascr	Score at the end of the minimal
	1544	y pasomin	Passed the minimal
G	1690	y g77code (R)*	What did the doctor tell him/her that pain or discomfort in his/her chest was?
	1829	y g168fis (R)	What kind of physical activity did he/she engage in regularly?
	1830	y g168soc (R)	What kind of social activity did he/she engage in regularly?
	1831	y g168oci (R)	What kind of leisure activity did he/she engage in regularly?
	1856	y depre_y	Score on the Depression Scale
H	1862	y code_h4 (R)	Father's occupation
	1863	y rcodeh4 (R)	Father's occupation recoded
	1870	y codeh6d (R)	Mother's occupation
	1871	y rcodeh6d (R)	Mother's occupation recoded
K	1957	y k60a_co (R)	Health service that he/she was not able to obtain or receive
N	1997	y code_n7 (R)	Current occupation (last week)
	1998	y code_n9 (R)	Type of establishment or business where he/she works
	2024	y coden18 (R)	Last job
	2025	y coden20 (R)	Type of establishment or business where he/she used to work
	2037	y coden25 (R)	Occupation during most of his/her life
	2038	y codn25b (R)	Type of establishment where he/she worked during most of his/her life
	2050	y n50code (R)	Type of health problem he/she is experiencing
O	2106	y o10code (R)	Main reason for returning from the United States to Puerto Rico
U	2146	y bmi	Body Mass Index
	2147	y rbmi	Body Mass Index Recoded
	2155	y legstand	Result of exercise: stand on one foot
	2159	y get_up	Result of exercise: stand up and walk

* (R)=Recoded from an original alphanumeric variable.

Errors and Limitations Identified in Data

There are some errors or known limitations in the database that the users should consider:

1. The variables G55 (has fallen after 60 years of age) and G56 (number of falls in the last year) present some cases without information (22 and 53, respectively) due to a programming error, corrected during the uploading of the data. Even though in some cases the information was obtained via telephone, for other cases the data is unknown. In the database, these are properly identified with the value -3.
2. See the last paragraph of the section "Utilization of the information contained in the rosters" on the limitation of the identification of characteristics of some people who lend or receive assistance (value -77).

3. Section S on abuse was answered by a reduced number of participants. This is due to the instructions given to the interviewers. Due to the delicate nature of the questions, interviewers were told not to go forward with this section if someone else was present at the time of the interview. Those users of the database interested in studying the abuse of elderly adults should take into account this limitation.
4. In section G on Health, there is an experimental sub-section of symptoms taken from the clinical environment with the double purpose of validating some of the illnesses identified by the interviewee and being able to analyzing the predictive power of those symptoms. Due to an initial programming error later corrected, some of the interviewees did not answer the questions related to the respiratory symptoms (G92-G92I3).
5. In section O on Migration, questions O10a to O10d were introduced once fieldwork had begun, which means that some of the cases do not have valid information. This problem only affects some of the interviewees who might have resided in the United States and are duly identified with the value -3.
6. Due to a programming error, the marital status of some of the siblings in section L does not have value.
7. The question regarding the diagnosis of bilharziasis (G107) contains a large amount of missing data because of an error in the programming of the questionnaire.
8. The variables M16 (son or daughter who most helps him/her), M17 (brother or sister who most helps him/her), and M18 (person who most helps him/her), have no information for some cases.
9. Some asset and income variables were collected as categorical variables, instead of continuous. For obvious reasons, the variables in this form allow limited analysis of effects of assets and income. We conducted a follow-up telephone survey of randomly selected PREHCO respondents from the original PREHCO survey to obtain information on assets and household income variables. In the near future we will provide a new version of the full data set with continuous measures for those data.

Codebook

The codebook book is available in the document CODEBOOK.PDF. For each one of the variables in the database there is the following information: name of the variable, label of the variable, position in the database, text of the corresponding question in the questionnaire, type of variable, size of variable, missing values, codes and labels used, and distribution of frequencies and commentaries, the latter for some complex variables. The distribution of frequencies is not presented for those variables that have more than 20 categories, except in the case of important variables such as the age of the interviewee.