WP. 34
ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Neuchâtel, Switzerland, 5-7 October 2009)

Topic (vii): Indicators for measuring the quality impact of data editing and imputation

STRATEGY FOR MONITORING THE 2011 CANADIAN CENSUS E&I

Invited Paper

Prepared by Michael Bankier and Wesley Benjamin, Statistics Canada

I. INTRODUCTION

1. CANCEIS (Canadian Census Edit and Imputation System) will be used to carry out edit and imputation (E&I) for all census variables in the 2011 Canadian Census of Population. See Bankier (2009) and Bankier and Crowe (2009) for more information on CANCEIS. Much of this paper is based on Benjamin (2009) and Bankier (2008).

2. In the Canadian Census, variables are split into non-overlapping groups called subject matter (SM) topics, and then these groups are processed sequentially in a series of E&I modules for each topic. Variables that were finalized in an earlier E&I module can appear in the edits of the module currently being processed. These variables cannot have their values imputed by the current module and hence are labelled unimputable. If the current module was allowed to change these unimputable variables, there would be no guarantee that the units would still pass the edits of the earlier modules. Occasionally, a variable from a later SM topic is referenced in an earlier SM topic. This reduces the chance of inconsistencies remaining or being generated by imputation in the earlier SM topic that cannot be easily resolved in the later SM topic. The processing of SM topics sequentially allows for these topics to be published sequentially as well.

3. The two types of modules are called derive and donor. Derive modules generate new variables and also perform deterministic imputation. Donor modules do minimum change donor imputation. For a specific module, the records may be partitioned into strata (e.g. two person households might be in one stratum) with each stratum being processed separately by a module.

4. An electronic census data dictionary is used by CANCEIS and census staff to access the following information on the census variables:
- the variables that enter a module and whether they imputable,
- the possible values for these variables, and which ones are considered valid or invalid
- the labels that are associated with qualitative variables since almost all qualitative variables are stored numerically on the census data base (e.g. for the variable SEX, the values 2 and 4 represent FEMALE and MALE respectively).
- the classes of responses that are available for a variable and the definitions of these classes. As an example, for the variable MARST (marital status), the class EVER_MARRIED consists of the following possible values, MARRIED, WIDOWED, SEPARATED, and DIVORCED. A variable can be set equal to a class in an edit which allows that single edit to replace several edits.

5. Census E&I processing is very complex because of the large number of variables involved (particularly for the 20% sample of households that receive a long questionnaire) and the high volume of

households being processed. This processing has to be completed over a relatively short period of time and with a high degree of confidence that it has been carried out correctly. Extensive testing prior to production is done to eliminate most errors. Any errors remaining have to be detected and corrected quickly during processing using monitoring tabulations.

6. This paper will describe how monitoring tabulations can be defined efficiently for all modules (but with a focus on donor modules in the examples in this paper). Certain types of monitoring tabulations should be produced for each module to show the impact of imputation on each imputable variable. Rather than programming these default monitoring tabulations manually, their code will be generated automatically by a SAS macro. The SAS macro will read what are the imputable variables from the census data dictionary plus what are the labels associated with the responses for each of the qualitative variables. Generating the SAS code for these default monitoring tabulations automatically will save significant staff time. Having similar formatting for all default monitoring tabulations will make it easier comparing results across several modules. Additional custom tabulations can also be specified by the user to supplement these default monitoring tabulations if required.

7. Tools to monitor the output of both derive and donor modules are needed soon to support the development of the 2011 E&I modules. These tools will be used at the following stages of the production cycle:

- analysing 2006 Census data to identify problems corrected by imputation or problems remaining after imputation
- updating and testing modules for 2011 including assessing the impact of different sets of parameters or edit rules
- E&I production
- and detailed certification of the data after E&I but prior to publication.

8. Using a clear and consistent monitoring strategy throughout testing and across all SM topics with standard tabulations automatically generated will lower the potential risk of missing inconsistent combinations of responses and will result in a much more efficient production cycle.

9. This paper only gives the high level requirements with regards to monitoring E&I processing and development for 2011. In practice, only a subset of the more important requirements may be implemented for 2011. Sections II to IV and VI to XI I describe these requirements. Section V gives examples of some tabulations that will be produced and how they will be used. Section XIII provides some concluding remarks.

II. GENERAL REQUIREMENTS FOR TABULATIONS

10. i) SAS will be used to create the monitoring tabulations since it is widely used at Statistics Canada.

ii) Monitoring tabulations output from SAS will be in Microsoft Excel workbooks since it is possible to do further analysis within Excel (for example with pivot tables) and then easily insert the results into reports written in Microsoft Word.

iii) The Excel workbooks should be formatted in a similar fashion, such that related tables can be appended into a new worksheet of an existing workbook or results can be appended as new columns in an existing table.

iv) A clear naming convention should be used for the Excel workbooks and the worksheets within the workbooks to easily identify the variables being tabulated.

III. CREATION OF TABULATIONS

11. A default selection of monitoring tabulations should be created automatically after the successful completion of every module. These default tabulations would be determined by information stored in the data dictionary (for example, which variables were imputable in that module). A method should also exist to predefine additional custom tables that are specific to a module that would be created automatically after the default tables have been created. Additionally, the user should be able to request additional custom tables after their initial analysis of the data. The creation of these tables, both default

and custom, should be parameterized so that the user can fine tune the results if desired. However, if the user is content with the default values, then no interaction would be required and a default set of tables would automatically be created.

IV. DEFINE DEFAULT FORMATS FOR TABULATIONS

12. All tabulations should have the response categories formatted such that response labels or class names appear in the tabulations rather than numeric values. These formats should be defined in the dictionary and be automatically applied to all tables when they are created. A default format should be defined for each variable in a module that will be used in the tables that are automatically generated. A method should exist for a user to define an alternate format to be used in additional custom table requests (to group responses into classes for example) as described in Section III.

V. EXAMPLES OF DEFAULT TABULATIONS

13. Examples of default tabulations that will be produced in the 2011 Census automatically are provided in this section. These results were produced from a sample of 2% of the households in the 2006 Canadian Census which were part of the 20% sample that received the long census form.

14. For a particular module, tabulations will be produced to show the impact of imputation on each imputable variable individually and also for pairs and triplets of imputable variables. In this section we will look at tabulations generated for the relationship to person 1 variable. This is an important variable since it is used to form families plus it is involved in many edits which flag inconsistencies between it and the other demographic variables (sex, marital status, common-law status and age).

15. Table 1 gives the ratio of the percentage of persons with a specific relationship after imputation to the percentage before imputation. Because of space limitations, only some of the possible relationships are listed in Table 1 and Table 2 (which defines the meaning of the labels used in Table 1 and the tables which follow). These ratios are given for the persons who passed the edits, the persons who failed the edits plus all persons in the population. For example, grandparents made up 0.04% of the population before imputation compared to 0.02% of the population after imputation with the ratio being .02/.04*100 = 40.82% (this is not 50% since more precise, unrounded values were used for the calculation in the numerator and the denominator). The impact of imputation on a relationship will be studied for ratios not close to 100%.

Table 1: Ratio of Percentage of Persons with a Relationship After Imputation Compared to the Percentage of Persons with a Relationship Before Imputation

	RATIO			BEFORE_PCT			AFTER_PCT		
	Sum			Sum			Sum		
	Status			Status			Status		
	PASSED	FAILED	ALL	PASSED	FAILED	ALL	PASSED	FAILED	ALL
AP_R2P12B*									
GR_PRNT	100.00	12.12	40.82	0.01	0.18	0.04	0.01	0.02	0.02
PERSON1S_SSP	100.00	40.82	82.63	0.11	0.26	0.13	0.11	0.11	0.11
SON_DAUGHT_INLAW	100.00	50.76	75.84	0.25	1.39	0.42	0.25	0.71	0.32
FATHER_MOTHER	100.00	81.74	94.25	0.44	1.17	0.55	0.44	0.95	0.52
PERSON1S_SD	100.00	108.81	101.44	29.62	33.28	30.16	29.62	36.21	30.59
GR_CHILD	100.00	115.24	104.19	0.91	1.99	1.07	0.91	2.29	1.12

Table 2: Description of Labels used with Relation to Person 1 Variable

Label	Description of Relationship to Person 1
BROTH_SIS_INLAW	Brother or Sister-In-Law
BROTHER_SISTER	Brother or Sister
FATH_MOTH_INLAW	Father or Mother-In-Law
FATHER_MOTHER	Father or Mother
GR_CHILD	Grandchild
GR_GR_CHLD	Great Grandchild
GR_PRNT	Grandparent
OSP_P1S_SD	Opposite Sex Common-Law Partner of Son or Daughter
PERSON1S_HW	Husband or Wife (Opposite Sex Married Spouse)
PERSON1S_OSP	Opposite Sex Common-Law Partner
PERSON1S_SD	Son or Daughter
PERSON1S_SMS	Same Sex Married Spouse
PERSON1S_SSD	Stepchild
PERSON1S_SSP	Same Sex Common-Law Partner
SON_DAUGHT_INLAW	Son or Daughter-In-Law

16. Next, the reason why such a high proportion of grandparents were lost to imputation will be studied. The Table 3 Excel worksheet shows the count (FAIL_N) and percentage (FAIL_P) for each combination of unimputed relationship (R2P12B1) and imputed relationship (R2P12B) for the persons who failed the edits. Table 4 shows that this Excel spreadsheet features a filter dropdown box which allows those persons who were grandparents before imputation to be selected and create Table 5. Table 5 shows that most of the grandparents were changed to children of person 1 or grandchildren of person 1.

Table 3: Count And Percentage For Each Combination Of Unimputed Relationship (R2P12B1) And Imputed Relationship (R2P12B) For The Persons Who Failed The Edits

	A	B	C	D
1	R2P12B1	R2P12B	FAIL_N	FAIL_P
2	BLANK	BROTH_SIS_INLAW	41	0.04
3	BLANK	BROTHER_SISTER	111	0.12
4	BLANK	EMPLOYEE	10	0.01
5	BLANK	EMPLOYEES_HW	2	0.00
6	BLANK	FATH_MOTH_INLAW	71	0.08
7	BLANK	FATHER_MOTHER	85	0.09
8	BLANK	FOSTER_CHILD	18	0.02
9	BLANK	GR_CHILD	141	0.15
10	BLANK	GR_PRNT	2	0.00
11	BLANK	GR_GR_CHLD	2	0.00
12	BLANK	LODGER	96	0.10
13	BLANK	LODGERS_HW	1	0.00
14	BLANK	LODGERS_OSP	1	0.00
15	BLANK	LODGERS_SD	2	0.00
16	BLANK	NEPHEW_NIECE	42	0.05
17	BLANK	OSP_GR_CHLD	1	0.00

Table 4: Illustration of Filter Dropdown Box

	A	B	C	D
1	R2P12B1	R2P12B	FAIL_N	FAIL_P
2	(All)	BROTH_SIS_INLAW	41	0.04
3	(Top 10...)	BROTHER_SISTER	111	0.12
4	(Custom...)	EMPLOYEE	10	0.01
5	BLANK	EMPLOYEES_HW	2	0.00
6	BROTH_SIS_INLAW	FATH_MOTH_INLAW	71	0.08
7	BROTHER_SISTER	FATHER_MOTHER	85	0.09
8	EMPLOYEE	FOSTER_CHILD	18	0.02
9	FATH_MOTH_INLAW	GR_CHILD	141	0.15
10	FATHER_MOTHER	GR_PRNT	2	0.00
11	FOSTER_CHILD	GR_GR_CHLD	2	0.00
12	GR_CHILD	LODGER	96	0.10
13	GR_GR_CHLD	LODGERS_HW	1	0.00
14	GR_PRNT	LODGERS_OSP	1	0.00
15	HW_GR_CHLD	LODGERS_SD	2	0.00
16	INVALID	NEPHEW_NIECE	42	0.05
17	LODGER	OSP_GR_CHLD	1	0.00
	LODGERS_HW			
	LODGERS_SD			
	LODGERS_SMS			
	NEPHEW_NIECE			
	BLANK			
	BLANK			
	BLANK			

Table 5: Imputed Relationship for Persons with an Unimputed Relationship of Grandparent

R2P12B1	R2P12B	FAIL_N	FAIL_P
GR_PRNT	BROTH_SIS_INLAW	1	0.01
GR_PRNT	GR_CHILD	10	0.05
GR_PRNT	GR_PRNT	4	0.02
GR_PRNT	GR_GR_CHLD	1	0.01
GR_PRNT	PERSON1S_SD	14	0.07
GR_PRNT	ROOMMATE	3	0.02
GR_PRNT	TOTAL	33	0.18

17. Table 6 allows us to study, for those grandparents changed to children or grandchildren by imputation, what were their ages before imputation (AGEU). It can be seen that these "grandparents" were all well under the age of 45 which is the minimum age allowed for a grandparent. The respondents in these households probably reported their relationship to the other person, i.e. "I am the grandparent of this child" rather than correctly reporting "This is my grandchild".

Table 6: Bivariate Tabulation of Relationship and Age for Grandparents Converted to Either Son/Daughter or Grandchild of Person 1

R2P12B1	R2P12B	AGEU	AGE	FAIL_N	FAIL_P
GR_PRNT	PERSON1S_SD	BLANK/INVALID	TOTAL	1	0.01
GR_PRNT	PERSON1S_SD	[0,14]	TOTAL	7	0.04
GR_PRNT	PERSON1S_SD	[15,19]	TOTAL	3	0.02
GR_PRNT	PERSON1S_SD	[20,29]	TOTAL	2	0.01
GR_PRNT	PERSON1S_SD	[30,39]	TOTAL	1	0.01
GR_PRNT	PERSON1S_SD	TOTAL	TOTAL	14	0.07

R2P12B1	R2P12B	AGEU	AGE	FAIL_N	FAIL_P
GR_PRNT	GR_CHILD	[0,14]	TOTAL	9	0.05
GR_PRNT	GR_CHILD	[15,19]	TOTAL	1	0.01
GR_PRNT	GR_CHILD	TOTAL	TOTAL	10	0.05

18. In another example, Table 7 shows that many of the same sex common-law partners of person 1 are converted to opposite sex common-law partners of person 1 without their sexes being changed. Thus the relationship was changed to be consistent with the reported sex. Table 8 shows that the persons who have their sex changed from male to female or female to male are generally person 1, the opposite sex married spouse of person 1 or the opposite sex common-law partner of person 1.

19. Table 9 in Appendix A shows that a significant proportion of the father/mothers of person 1 are changed to son/daughters of person 1 and that many of those are under the age of 30 which is the minimum age for a father/mother of person 1. Many of these errors are presumably again the result of a respondent reporting that they are the father/mother of another person rather than correctly reporting that this person is their child. Table 10 in Appendix A shows that many son/daughters-in-law of person 1 are converted to son/daughters and the majority of these are under the age of 15. With some of these, the respondent may have been trying to report step-children.

Table 7: Bivariate Tabulation of Relationship and Sex for Same Sex Partner of Person 1

R2P12B1	R2P12B	SEXU	SEX	FAIL_N	FAIL_P
PERSON1S_SSP	BROTHER_SISTER	TOTAL	TOTAL	1	0.01
PERSON1S_SSP	PERSON1S_HW	TOTAL	TOTAL	2	0.01
PERSON1S_SSP	PERSON1S_OSP	TOTAL	TOTAL	28	0.15
PERSON1S_SSP	PERSON1S_SD	TOTAL	TOTAL	2	0.01
PERSON1S_SSP	PERSON1S_SSP	TOTAL	TOTAL	13	0.07
PERSON1S_SSP	ROOMMATE	TOTAL	TOTAL	3	0.02
PERSON1S_SSP	TOTAL	TOTAL	TOTAL	49	0.26

R2P12B1	R2P12B	SEXU	SEX	FAIL_N	FAIL_P
PERSON1S_SSP	PERSON1S_OSP	FEMALE	FEMALE	15	0.08
PERSON1S_SSP	PERSON1S_OSP	FEMALE	TOTAL	15	0.08
PERSON1S_SSP	PERSON1S_OSP	MALE	MALE	13	0.07
PERSON1S_SSP	PERSON1S_OSP	MALE	TOTAL	13	0.07
PERSON1S_SSP	PERSON1S_OSP	TOTAL	FEMALE	15	0.08
PERSON1S_SSP	PERSON1S_OSP	TOTAL	MALE	13	0.07
PERSON1S_SSP	PERSON1S_OSP	TOTAL	TOTAL	28	0.15

Table 8: Bivariate Tabulation of Relationship and Sex for Persons Whose Sex was Changed by Imputation

R2P12B1	R2P12B	SEXU	SEX	FAIL_N	FAIL_P
PERSON1	TOTAL	FEMALE	MALE	15	0.08
PERSON1S_HW	TOTAL	FEMALE	MALE	2	0.01
PERSON1S_OSP	TOTAL	FEMALE	MALE	5	0.03
PERSON1S_SD	TOTAL	FEMALE	MALE	1	0.01
SON_DAUGHT_INLAW	TOTAL	FEMALE	MALE	2	0.01
TOTAL	TOTAL	FEMALE	MALE	25	0.13

R2P12B1	R2P12B	SEXU	SEX	FAIL_N	FAIL_P
BLANK	TOTAL	MALE	FEMALE	2	0.01
FATHER_MOTHER	TOTAL	MALE	FEMALE	2	0.01
PERSON1	TOTAL	MALE	FEMALE	10	0.05
PERSON1S_HW	TOTAL	MALE	FEMALE	20	0.11
PERSON1S_OSP	TOTAL	MALE	FEMALE	18	0.10
PERSON1S_SD	TOTAL	MALE	FEMALE	1	0.01
PERSON1S_SSP	TOTAL	MALE	FEMALE	1	0.01
SON_DAUGHT_INLAW	TOTAL	MALE	FEMALE	3	0.02
TOTAL	TOTAL	MALE	FEMALE	57	0.30

VI. ADDITIONAL DETAILS ON DEFAULT TABULATIONS

20. The default tabulations (as discussed in Section III and V) should include all univariate (e.g. see Table 3) and bivariate (e.g. see Table 6) distributions based on the imputable variables for the selected modules. Trivariate distributions should be available for creation on request as well. Additional auxiliary variables (those not appearing in the edits) should be able to be included in these automatically created tables. Certain auxiliary variables should be included in all of the default tables (e.g. Province), and the user can define extra auxiliary variables in their custom tables. The number of variables included in a table should be carefully considered due to the maximum number of rows and columns constraints available in Excel. If this limit is exceeded, the appropriate Excel worksheet will display only a portion of the table.

21. Tables should be able to display distributions for multiple subpopulations side by side for comparison. This could include things such as comparing failed, passed and all records, or comparing responses from Internet and paper questionnaires. Additionally, tables should be viewable as both counts and percentages. For certain tables with fewer response combinations, the user should be able to request a custom table in a tabular format instead of the compressed summary table format as displayed in Section V. Alternatively, these can be produced from the compressed summary table format by converting it to an Excel pivot table.

VII. DISTRIBUTIONS ACROSS MODULES

22. In the 2011 Canadian Census, whenever a variable has its responses imputed for one or more records, a new version of the variable will be created with the module name as a suffix. This means a variable will never have it values overwritten. Thus, if the variable SEX was initially created in module1 and then updated in modules 2 and 3, the three versions of this variable, SEX_module1, SEX_module2 and SEX_module3, would all exist on the census data base. Doing this creates a complete audit trail of the impact of imputation on each variable.

23. The functionality should exist to easily track the distribution of a variable's responses as it is changed throughout production by imputation. The tabulation system should present the user with a table that lists the distribution for the variable's responses in a series of columns representing the modules in which the variable was updated (e.g. Load, Module1, Module2, etc.). This table would be updated and appended with the latest distribution every time that variable is updated in production. The user should also be able to define bivariate distributions (e.g. Marital Status by Age) that would be updated after either variable has been imputed.

VIII. MICRO LEVEL RECORDS

24. The user should be able to request workbooks that display the micro level (i.e. household or person level) data for records matching certain criteria. For example, the user may want to see the micro level data for a random sample of five records where the imputed age and marital status were different from the initial age and marital status.

25. For derive modules, this workbook would display the before and after responses for all variables for the selected records. For donor modules, this workbook would display responses for the failing record, donor record and final imputed record for the selected records. In addition, a list of the edits that a record matched will be given to help explain why certain variables were imputed.

IX. COMPARE 2006 VS. 2011 DISTRIBUTIONS

26. The user should be able to compare the distributions for the 2011 data to 2006 data in adjacent columns of the same table if the content has not changed between the 2006 Census and the 2011 Census. This could be done prior to all 2011 data being captured and also at the end of processing. If the 2006 and 2011 bases are linked at the dwelling level, this will also allow a comparison to be done of just those dwellings which responded in both 2006 and 2011.

These comparisons will allow for the easy identification of any radical changes in distributions between censuses and will facilitate the correction of the more serious problems as soon as possible.

X. ADDITIONAL CANCEIS OUTPUT FILE TABULATIONS

27. CANCEIS processes data and creates output files at the stratum level. The monitoring program should be able to concatenate output files across strata for a module into a single worksheet, displaying the results for each individual stratum along with the information at the module level (summed across all strata). An example where this would be useful is strata edit failure rates which could be concatenated together into a single table spanning all strata.

XI. TABULATIONS OF EDITS FAILED AND VARIABLES IMPUTED

28. A file will be created by CANCEIS for each donor module listing all edit rules failed by each failed record as well as which variables were imputed to resolve those edit failures and what imputed value they were given. This file will allow one to determine how often an edit rule failed and what percentage of the time each variable entering that edit rule was changed.

XII. STANDARDIZATION OF FAILED RECORDS TO PASSED RECORDS

29. One requirement of imputation is that distributions be preserved in some sense. Often, for a particular variable, both before and after imputation, the distribution of the passed records compared to the distribution of the failed records will be quite different. As an example, the age distribution may be quite different for the failed records compared to the passed records. This can explain some differences in the distributions of other variables when the failed and passed records are compared. For example, if the failed records have a higher proportion of children than the passed records, the failed records will usually have a higher proportion of persons who have never been married compared to the passed records. Because of this, it may be useful to standardize the failed records by, for example, five year age ranges. This would be done by determining weights for the failed records such that the weighted counts for failed records after imputation equalled the counts for the passed records for each five year age range. If the distribution of marital status for the weighted failed records was similar to the passed records, this would suggest that the larger differences in the distributions originally observed were not of concern. The approach used here is similar conceptually to what is done in epidemiology where the death rate for a town, for example, is standardized on age and sex. Further experimentation will be done with standardization to see if it should be widely used in the analysis of the imputation results.

XIII. CONCLUDING REMARKS

30. The purpose of the census monitoring tabulations is to measure the impact of imputation on the census data and also assess the reasonableness of these imputation actions. Careful analysis of monitoring tabulations will prove useful in the development and testing of the E&I modules and their running in production. The automatic generation of tables will result in time savings in the programming of the monitoring tabulations. It will also make it easier for the user to check the results since there will be consistency in the tables across all modules.

References

Bankier, M. (2008), “Methods to Analyse the Impact of E&I on Census Data”, Social Survey Methods Division Report, Statistics Canada, March 5, 2008.

Bankier, M. (2009), "Evolution of Canadian Census E&I Systems – 1976 to 2011", Working Paper ??, Proceedings of the UN/ECE Work Session on Statistical Data Editing, Switzerland (Neuchâtel).

Bankier, M. and Crowe, S. (2009), "Enhancements to the 2011 Canadian Census E&I System", Working Paper ??, Proceedings of the UN/ECE Work Session on Statistical Data Editing, Switzerland (Neuchâtel).

Benjamin, W. (2009), “Methodology’s Requirements for 2011 Edit and Imputation Monitoring Tabulations - Draft”, Social Survey Methods Division Report, Statistics Canada, Dated March 2009.

Appendix A: Other Examples of Monitoring Tables

Table 9: Bivariate Tabulation of Relationship and Age for Father/Mothers Converted to Son/Daughter of Person 1

R2P12B1	R2P12B	AGEU	AGE	FAIL_N	FAIL_P
FATHER_MOTHER	FATHER_MOTHER	TOTAL	TOTAL	154	0.82
FATHER_MOTHER	GR_CHILD	TOTAL	TOTAL	1	0.01
FATHER_MOTHER	LODGER	TOTAL	TOTAL	1	0.01
FATHER_MOTHER	PERSON1S_HW	TOTAL	TOTAL	4	0.02
FATHER_MOTHER	PERSON1S_SD	TOTAL	TOTAL	54	0.29
FATHER_MOTHER	ROOMMATE	TOTAL	TOTAL	3	0.02
FATHER_MOTHER	ROOMMATES_HW	TOTAL	TOTAL	1	0.01
FATHER_MOTHER	SON_DAUGHT_INLAW	TOTAL	TOTAL	1	0.01
FATHER_MOTHER	TOTAL	TOTAL	TOTAL	219	1.17

R2P12B1	R2P12B	AGEU	AGE	FAIL_N	FAIL_P
FATHER_MOTHER	PERSON1S_SD	BLANK/INVALID	TOTAL	1	0.01
FATHER_MOTHER	PERSON1S_SD	[0,14]	TOTAL	23	0.12
FATHER_MOTHER	PERSON1S_SD	[15,19]	TOTAL	8	0.04
FATHER_MOTHER	PERSON1S_SD	[20,29]	TOTAL	11	0.06
FATHER_MOTHER	PERSON1S_SD	[30,39]	TOTAL	5	0.03
FATHER_MOTHER	PERSON1S_SD	[40,44]	TOTAL	2	0.01
FATHER_MOTHER	PERSON1S_SD	[45,64]	TOTAL	4	0.02
FATHER_MOTHER	PERSON1S_SD	TOTAL	TOTAL	54	0.29

Table 10: Bivariate Tabulation of Relationship and Age for Son/Daughters-In-Law Converted to Son/Daughter of Person 1

R2P12B1	R2P12B	AGEU	AGE	FAIL_N	FAIL_P
SON_DAUGHT_INLAW	BROTHER_SISTER	TOTAL	TOTAL	1	0.01
SON_DAUGHT_INLAW	FATH_MOTH_INLAW	TOTAL	TOTAL	6	0.03
SON_DAUGHT_INLAW	FATHER_MOTHER	TOTAL	TOTAL	5	0.03
SON_DAUGHT_INLAW	GR_CHILD	TOTAL	TOTAL	3	0.02
SON_DAUGHT_INLAW	LODGER	TOTAL	TOTAL	1	0.01
SON_DAUGHT_INLAW	NEPHEW_NIECE	TOTAL	TOTAL	1	0.01
SON_DAUGHT_INLAW	OTHER_REL	TOTAL	TOTAL	1	0.01
SON_DAUGHT_INLAW	PERSON1S_HW	TOTAL	TOTAL	2	0.01
SON_DAUGHT_INLAW	PERSON1S_SD	TOTAL	TOTAL	121	0.64
SON_DAUGHT_INLAW	PERSON1S_SSD	TOTAL	TOTAL	2	0.01
SON_DAUGHT_INLAW	ROOMMATE	TOTAL	TOTAL	1	0.01
SON_DAUGHT_INLAW	SON_DAUGHT_INLAW	TOTAL	TOTAL	118	0.63
SON_DAUGHT_INLAW	TOTAL	TOTAL	TOTAL	262	1.39

R2P12B1	R2P12B	AGEU	AGE	FAIL_N	FAIL_P
SON_DAUGHT_INLAW	PERSON1S_SD	BLANK/INVALID	TOTAL	1	0.01
SON_DAUGHT_INLAW	PERSON1S_SD	[0,14]	TOTAL	65	0.35
SON_DAUGHT_INLAW	PERSON1S_SD	[15,19]	TOTAL	29	0.15
SON_DAUGHT_INLAW	PERSON1S_SD	[20,29]	TOTAL	18	0.10
SON_DAUGHT_INLAW	PERSON1S_SD	[30,39]	TOTAL	5	0.03
SON_DAUGHT_INLAW	PERSON1S_SD	[40,44]	TOTAL	1	0.01
SON_DAUGHT_INLAW	PERSON1S_SD	[45,64]	TOTAL	2	0.01
SON_DAUGHT_INLAW	PERSON1S_SD	TOTAL	TOTAL	121	0.64