A Comparison Study of ACS If-Then-Else, NIM, and DISCRETE Edit and Imputation Systems Using ACS Data

Bor-Chung Chen
Yves Thibaudeau
William E. Winkler

U.S. Bureau of the Census
bchen@info.census.gov
October 21, 2003

The Basic Assumption of the Comparison Study

We assume that the inputs (the edit rules) to the three systems are identical.

The Edit and Imputation Systems Compared and Data Set Used

- American Community Survey (ACS) If-Then-Else (ITE) Rules
- Nearest-Neighbor Imputation Method (NIM) (Bankier [1997, 2000])
- DISCRETE Edit and Model-Based Imputation (Winkler [1995, 1997], Chen [1998], Chen, Winkler, and Hemmig [2001], Chen and Winkler [2002], Thibaudeau [2002])
- 1999 ACS Data Set of 26 States

Existing If-Then-Else (ITE) Rules Used by ACS

- The 1999 ACS Edit and Allocation Specifications for Basic Population Variables
- Sex, Age, Household Relationship, Marital Status
- Sections by Variables
- Sequential Edit and Imputation
- SAS Programming Language

Bankier's Nearest-Neighbor Imputation Method (NIM)

- Using Donors---Nearest Neighbors
- All of the imputed records satisfy all of the edits
- The imputed household closely resembles the failed household
- Good imputation actions have equal chance of being selected
- Needs enough donors
- Modifiable Decision Logic Tables (Edit Rules)
- Simultaneous Edit and Imputation

DISCRETE Edit and Model-Based Imputation System

- Edit Generation: generates a complete set of edits
- Error Localization: needs a complete set of edits to determine a minimum number of fields to change if a record fails some of the edits
- Modifiable Edit Tables (Edit Rules)
- Model-Based Item Imputation (Thibaudeau[2002])
- Simultaneous Edit and Imputation

Pre-Edits (Logical Edits)

- Some missing fields in a record can be logically derived from other non-missing fields
- Identify the householder and spouse if present
- Household relationship conversion
- Convert each of the households into at least one 3-person household
- Derive age or date of birth if one of them is missing and check consistency between them

NIM with and without Pre-edits

HH size	Total HHs	W/O Pre-edits (% Failed)	With Pre-edits (% Failed)
3	16954	40.03	27.47
4	14258	45.22	33.56
5	6742	50.86	39.20
6	2129	92.44	55.85
7	719	95.27	60.92
8	319	96.55	64.89
9	150	96.67	64.00
Total	41271	47.90	33.96

Edit Rules: ACS If-Then-Else

In the section of household relationship and marital status:

Universe	Person 2+ and *Relationship* is Husband/wife;
If…	*Marital status* is Widowed, divorced, separated, or never married;
Then…	Make *Marital status* = Married;

Edit Rules: Decision Logic Table of NIM

```
RELANU(01) = PERSON1
;Y;Y;Y;Y;Y;Y;

RELANU(02) = HUSBAND_WIFE
;Y;Y;Y;Y;Y; ;

SEXU(01) = SASMIS          ; ;Y;Y;
; ; ;

SEXU(02) = SASMIS          ; ; ;
;Y;Y; ;

SEXU(01) = MALE            ; ; ;
;Y; ; ;

SEXU(01) = FEMALE          ; ; ; ;
;Y; ;

SEXU(02) = MALE            ; ;Y; ;
; ; ;

SEXU(02) = FEMALE          ; ; ;Y;
; ; ;
```

Edit Rules:
Edit Table of DISCRETE

Explicit edit # 25: 3 entering field(s)

RELANU11	1 response(s): 1
RELANU22	1 response(s): 2
MARSTU22	4 response(s): 2 3 4 5

Passed Households between DISCRETE and NIM

HH size	Total HHs	DISCRETE	NIM
3	16954	12296	12296
4	14258	9473	9473
5	6742	4099	4099
6	2129	940	940
7	719	281	281
8	319	112	112
9	150	54	54
Total	41271	27255	27255

Statistical Comparisons on the Imputed Results

	Edit-passing HH		Imputed HH (NIM)	
4-Person	Freq	Prop (x_i)	Freq	Prop (y_i)
Married	14721	0.390	1250	0.385
Widowed	573	0.015	52	0.016
Divorced	1414	0.038	107	0.033
Separated	507	0.013	37	0.012
N. Married	20569	0.544	1798	0.554
Total	37784	1.000	3244	1.000

$$NIM_4^{ms} = \sum_{i=1}^{n} (x_i - y_i)^2$$

Statistical Comparisons on the Imputed Results (Continued)

	sex	ms	age	hhr	sex-ms	sex-age	sex-hhr	ms-age	ms-hhr	age-hhr
ITE	.017	.072	.047	.095	.047	.031	.059	.045	.105	.045
NIM	.014	.008	.012	.014	.015	.013	.020	.014	.019	.015
DMB	.019	.006	.043	.036	.016	.035	.035	.044	.046	.042

$$S^{v}_{ITE} = \sum_{i=3}^{9} ITE^{v}_{i}, \quad S^{v}_{NIM} = \sum_{i=3}^{9} NIM^{v}_{i}$$

$$v = sex, ms, ..., age - hhr .$$

Imputation Results Agreed and Disagreed(If-Then-Else vs. NIM)

HH size	imputed	agreed	disagreed
3	4658	3564	1094
4	4774	3958	816
5	2643	2007	636
6	1028	720	308
7	438	269	169
8	207	117	90
9	96	54	42
Total	13844	10689	3155

Imputation Results Agreed and Disagreed(If-Then-Else vs. DMB)

HH size	imputed	agreed	disagreed
3	4658	3627	1031
4	4774	3961	813
5	2643	2056	587
6	1028	736	292
7	438	299	139
8	207	144	63
9	96	63	33
Total	13844	10886	2958

Imputation Results Agreed and Disagreed(NIM vs. DMB)

HH size	imputed	agreed	disagreed
3	4658	3632	1026
4	4774	4038	736
5	2643	2091	552
6	1028	728	300
7	438	276	162
8	207	121	86
9	96	56	40
Total	13844	10942	2902

Percentage of Households Failed after Imputations

HH size	ITE	NIM	DMB
3	0.22	0.45	0.61
4	0.34	0.51	0.62
5	0.48	1.22	0.96
6	1.18	1.64	1.90
7	1.50	1.24	2.40
8	1.42	2.30	2.13
9	2.10	2.02	2.80
Total	0.39	0.78	0.79

Why Disagreed?

- Other relative (roomer/boarder, brother/sister) vs. spouse
- Unnecessary change of age by ITE?
- Unnecessary change of sex by ITE?
- Ineffective sequential edit and imputation of ITE?
- Minimum number of fields to change by NIM and DMB?
- Nearest neighbor imputation by NIM?

Why Disagreed (contd.)?

- Unnecessary change of age by NIM
- Imputed households still fail some of the edits by DMB
- Divorced vs. widowed
- Foster child vs. son and other nonrelative
- Unknown marital status

Other Relative vs. Spouse

ID	Sex	Age	HHR	MS	ITE(HHR)	NIM(HHR)	DMB(HHR)
1	M	56	Householder	Married	Householder	Householder	Householder
2	F	16	Daughter	Never Married	Daughter	Daughter	Daughter
3	M	14	Son	Never Married	Son	Son	Son
4	F	53	Daughter	Married	Other Relative	Spouse	Spouse

Roomer/Boarder vs. Spouse

ID	Sex	Age	HHR	MS	ITE(HHR)	NIM(HHR)	DMB(MS)
1	M	41	Householder	Married	Householder	Householder	Widowed
2	F	18	Mother	Never Married	Daughter	Daughter	Never Married
3	M	19	Son	Never Married	Son	Son	Never Married
4	F	16	Other Relative	Never Married	Other Relative	Other Relative	Never Married
5	F	38	Unmarried Partner	Married	Roomer/ Boarder	Spouse	Divorced

Unnecessary Change of Age by ITE ?

ID	Sex	Age	HHR	MS	ITE (Age)	NIM (Age)	DMB (Age)
1	F	36	Householder	Married	36	36	36
2	M	37	Spouse	Unknown (Married)	37	37	37
3	F	12	Daughter	Never Married	24	12	12
4	M	10	Son	Never Married	10	10	10

Unnecessary Change of Sex by ITE ?

ID	Sex	Age	HHR	MS	ITE (Sex)	NIM (HHR)	DMB (HHR)
1	M	46	Householder	Married	M	Householder	Householder
2	F	35	Spouse	Married	F	Spouse	Spouse
3	M	16	Father	Never Married	M	Son	Son
4	F	6	Mother	Never Married	M	Daughter	Daughter
5	M	5	Son	Never Married	M	Son	Son

Ineffective Sequential Edit and Imputation of ITE ?

ID	Sex	Age	HHR	MS	ITE (HHR)	NIM (HHR)	DMB (MS)
1	F	35	Householder	Married	Householder	Householder	Separated
2	F	15	Daughter	Married	Daughter	Daughter	Married
3	M	21	Unmarried Partner	Married	Son	Spouse	Never Married

Minimum Number of Fields to Change by NIM and DMB?

ID	Sex	Age	HHR	MS	ITE (HHR)	NIM (HHR)	DMB (HHR)
1	M	33	Householder	Married	Householder	Householder	Householder
2	F	29	Other Nonrelative	Married	Spouse	Other Nonrelative	Other Nonrelative
3	M	2	Unknown	Never Married	Son	Son	Son

Nearest Neighbor Imputation by NIM?

ID	Sex	Age	HHR	MS	ITE (HHR)	NIM (HHR)	DMB (HHR)
1	M	41	Householder	Married	Householder	Householder	Householder
2	F	43	Spouse	Married	Spouse	Spouse	Spouse
3	F	9	Unknown	Never Married	Daughter	Other Relative	Daughter

Unnecessary Change of Age by NIM

ID	Sex	Age	HHR	MS	ITE (Age)	NIM (Age)	DMB (Age)
1	M	46	Householder	Married	46	47	46
2	F	41	Spouse	Married	41	41	41
3	M	13	Son	Never Married	13	13	13
4	M	11	Brother (Son)	Never Married	11	11	11
5	M	7	Brother (Son)	Never Married	7	7	7

Imputed Households Still Fail Some of the Edits by DMB (1)

ID	Sex	Age	HHR	MS	ITE	NIM	DMB
1	M	52	Householder	Married			
2	F	50	Unknown	Unknown	Spouse Married	6 Daughter Never Married	Daughter Divorced
3	M	12	Son	Never Married			
4	F	41	Spouse	Unknown	Sister Married	Married	Sister Never Married

Imputed Households Still Fail Some of the Edits by DMB (2)

ID	Sex	Age	HHR	MS	ITE	NIM	DMB
1	F	66	Householder	Widowed	Married	36 Married	32 Married
2	M	31	Spouse	Married			
3	F	31	Daughter	Married		10 Never Married	5
4	F	2	Daughter	Never Married	Grandchild		

Imputed Households Still Fail Some of the Edits by DMB (3)

ID	Sex	Age	HHR	MS	ITE	NIM	DMB
1	F	45	Householder	Separated			69
2	M	63	Son	Never Married	5	11	Father
3	F	43	Daughter	Never Married	8	15	
4	F	41	Daughter	Never Married	13	4	Other Nonrelative

Divorced vs. Widowed

ID	Sex	Age	HHR	MS	ITE (HHR)	NIM (HHR)	DMB (HHR)
1	M	47	Householder	Never Married	Never Married	Never Married	Never Married
2	M	33	Brother	Never Married	Never Married	Never Married	Never Married
3	M	79	Father	Married	Married	Married	Married
4	F	72	Mother	Unknown	Divorced	Widowed	Widowed

Foster Child vs. Son and Other Nonrelative

ID	Sex	Age	HHR	MS	ITE	NIM	DMB
1	F	45	Householder	Married			
2	M	53	Spouse	Unknown	Married	Married	Unknown (MS)
3	M	18	Foster Child	Never Married	Other Nonrelative	Son	

Example of Unknown Marital Status

ID	Sex	Age	HHR	MS	ITE	NIM	DMB
1	F	41	Householder	Unknown	Married	Married	Divorced
2	M	39	Unmarried Partner	Unknown	Brother Married	Spouse Married	Never Married
3	M	66	Father	Widowed			

Discussion and Summary

- All three systems could not make some of edit-failing households to pass all edits
- NIM imputation is “closer” to the joint distributions of edit-passing households
- NIM requires enough donor households to do the imputation

Discussion and Summary (Cont.)

- With NIM and DISCRETE, the computer code needs not to be rewritten from a survey to another and/or when the edit rules are changed
- DISCRETE edit generation may require lot of computing resources