ICARIS-2 Public Use Data Set

Appendix C – Data Quality Log

Data Quality Log – The Data Quality Log is a summary of data issues that have been investigated by the contractor. These issues were discovered either by the contractor during the data cleaning phase of operations or by NCIPC staff in conducting data analyses. Data issues include questions that were inadvertently skipped by the interviewer, possible response outliers, and inconsistencies in information provided across variables for a given respondent or "R". In some instances, respondents listed in the log have been specifically identified by their unique respondent identifier (R_ID). Should data anomalies be discovered when using these data, the user should refer to this log for a possible explanation.

Telephone Screener
 SC5 was skipped if SC2 + SC3 = 1 (3127 cases)
- for 6 other cases, SC2 or SC3 was 1 and the other variable was DK or RF, and SC5 was skipped
- 1 unique case (R_ID 55742) said 0 men and RF the number of women. SC5 was asked and the answer was no.
- all other cases had legitimate answers to both SC2 and SC3 so SC5 was asked
- if both SC2 and SC3 were answered as Don't Know and/or Refused, the case was considered a refusal to screen since we could not implement the respondent sampling protocol

Module 4: Falls among the Elderly
 FE section is missing for 2 cases where the respondent was ≥ 65 years old
- R_ID 59020 – age was not calculated in real time during the interview so FE was skipped, but respondent was 85 years old
- R_ID 18029 – respondent was 67 years old, unknown why section was blank

Module 5: Enumeration of Children in the Household
 No household had more than 7 children under age 15. Therefore, the K section variables for the 8th, 9th and 10th child in a home are not included in the analytic dataset.

Module 7: Child Counseling
 The entire child counseling section (C3CN C4a-e) was skipped for 1 record (R_ID 10067) where a child older the age of 14 was mistakenly selected. It is not clear how this selection was made. C3CN was later coded as "dot" or inadvertent skip.

 C4d and C4e are missing for 23 randomly selected (R_IDs 02363, 07070, 10088, 12068, 13339, 20238, 23940, 24004, 25125, 29727, 37547, 37743, 40097, 41032, 42782, 44214, 47476, 50158, 55358, 56482, 58229, 60826, and 68142) children for whom age was not known. When age is not known, child age was assumed to be 2-6 years for purposes of asking questions in this section. This incorrect skip was the result of an error in the CATI program.

Module 8: Motor Vehicle
 The entire child motor vehicle section (MV11, MV12a-c, MV13) was skipped for 2 records where a child in the household was age 0-14 years.
- R_ID 10067 – A child older the age of 14 was mistakenly selected (C3CN = "dot" … see note under Child Counseling, Module 7).
- R_ID 26363 – Although a single child under age 15 resided in the household, the value for C3CN was mistakenly coded as "Not Applicable". Although the counseling questions were correctly asked, the motor vehicle question was skipped. C3CN has since been coded to "child 1".

 Two cases (R_IDs 35006 and 40097) with unknown age were not asked questions MV12a and MV12b. In 28 additional instances where age was unknown, age was assumed to be 2-6 years and these questions were asked. It is not known why MV12a and b were skipped for these two records.

Module 8: Motor Vehicle, continued

Unknown why MV3d, MV3e, and MV7 are missing for R_ID 18029 because the respondent age was calculated as being 67. They are not missing for R_ID 59020 because the respondent answered do not drive/no license for MV1 and appropriately skipped out of these questions.

Module 10: Child Bicycle Helmet Safety

BIKE6 is missing for 12 instances (R_IDs 00726, 01318, 01591, 03034, 03623, 07856, 09314, 10066, 12833, 14733, 17843, and 19318) for which is should have been asked (BIKE5 = 1 – 7). This was the result of an inappropriate skip pattern discovered early in the field period and subsequently fixed.

Module 12: Dogs and Dog Bites

The number of cases where DOGNUM = NA/SKIP(97) is inflated by 22. This occurred because the interviewer first entered 0, 98 (DK), or 99 (RF) in DB5 and then went back and changed the response. As soon as DB5 has a value, the random number (DOGNUM) that links to a letter is generated. If the value of DB5 is 0, 98, or 99, DOGNUM is assigned the value of 97. This value is not linked to a letter. In order to prevent DOGNUM and the letter from changing if the case is opened again, DOGNUM is stored and can only be created once. Even though the interviewer corrected the answer to DB5, DOGNUM remained the same so there was no letter associated with it. When the first case like this was brought to our attention, the interviewers were instructed to make up a letter and read that letter in the sentence.

Module 15: Interpersonal Violence

Due to a programming error and some corruption, there are 4 cases in which data are missing for this module. The missing data items are indicated with a check in the following table.

Variable	Case 12294	Case 21347	Case 35349	Case 22525
V6a	✓	✓	✓	
V7a	✓	✓	✓	
V8a	✓	✓	✓	
V2Bb	✓	✓	✓	
V2bCode	✓	✓	✓	
V2bSR	✓	✓	✓	
V3b	✓	✓	✓	✓
V4b	✓	✓	✓	✓
V4bCode	✓	✓	✓	
V5b	✓	✓	✓	✓
V6b	✓	✓	✓	
V7b	✓	✓	✓	✓
V2c	✓	✓	✓	✓
V2cCode	✓	✓	✓	
V2cSR	✓	✓	✓	✓
V3c	✓	✓	✓	✓
4Vc	✓	✓	✓	✓
V4cCode	✓	✓	✓	
V5c	✓	✓	✓	✓
V6c	✓	✓	✓	✓
V7c	✓	✓	✓	✓
Table continued on the next page				

Variable	Case 12294	Case 21347	Case 35349	Case 22525
V2d	✓	✓	✓	✓
V2dCode	✓	✓	✓	
V2dSR	✓	✓	✓	✓
V3d	✓	✓	✓	✓
V4d	✓	✓	✓	✓
4VdCode	✓	✓	✓	
V5d	✓	✓	✓	✓
V6d	✓	✓	✓	✓
V7d	✓	✓	✓	✓

Module 16: Sexual Violence

Due to a programming error, S2, S3A, S3ACODE, S3B, and S3BCODE are missing for 18 cases (R_IDs 01591, 07957, 08966, 15298, 18043, 22560, 23179, 29348, 33568, 40924, 44494, 45826, 45922, 46357, 50186, 53667, 62596, and 68550)

SV7a, b, and c were skipped for 2 respondents (R_IDs 32307 and 32506) who answered SV5=2 "Other people". It is not clear why these questions were inadvertently skipped.

Module 21: End Demographics

This section is missing for R_ID 08001 due to corruption. Items have been shown as missing on the weekly reports with an explanation.

Several instances exist in which a multiple person household with only 1 adult (SC2=1 and SC3=0, or SC2=0 and SC3=1) answered D2 (highest level of household education completed) with a response that differed from and was greater than D1 (highest level of respondent education completed). In 17 instances, the highest level of household education was "Some Post-HS", in 32 instances it was "College Grad" and in 24 instances it was "Post Grad". While a response of "Post HS" could be given for one less than age 18 living in the household, responses of "College Grad" or "Post Grad" seem unlikely. The researcher will need to make his/her own decision as to how to handle these situations.

Some "Possible" Outliers

While these values appear unusual, we can not determine if they are real or typos.

SC2[*] *(Number of Men in the HH)* **range: 0-23**

Response	Frequency	Response	Frequency
11	2	14	2
12	1	23	1

SC3[*] *(Number of Women in the HH)* **range: 0-23**

Response	Frequency	Response	Frequency
11	3	23	1
12	3		

Some "Possible" Outliers, continued

SC4[*] *(Number of Children in the HH)* range: 0-32

Response	Frequency
30	1
32	1

*Note: None of the SC2, SC3 or SC4 outliers are for the same case.

NK0418Y *(Children in HH age 4-18 yrs)* range: 0-32

Response	Frequency
30	1
32	1

DH7 *(Total telephone numbers ringing in HH)* range: 2-90

Response	Frequency	Response	Frequency
6	1	19	1
7	1	21	1
8	1	22	1
10	2	23	1
12	3	90	1

A3 *(Drinks per day on average)* range: 1-63

Response	Frequency	Response	Frequency
20	2	39	1
22	1	45	1
23	1	63	1

M6 *(Miles driven past year, units are thousand)* range: 0-940
  Even after adding the confirmation screen on 04/15/2002, still appear to be many outliers.

DB5 *(Number of dogs in home)* range: 1-30

Response	Frequency	Response	Frequency
8	3	15	1
9	3	16	1
10	3	19	1
11	1	25	1
12	3	29	1
13	1	30	1

FA15 *(Number of adults in HH attending firearms safety workshop)* range: 0-14

Response	Frequency
14	1

FA19 *(Number of crime victimizations, past 12 months)* range: 0-46

Response	Frequency	Response	Frequency
20	5	46	1
30	4		

"Data Issues" Discovered During Analysis

Administrative Variable

Respondent ID 26759

Had a date of interview of May 5, 2003 (after the study ended)

IntDate

5052003

Comment: The contractor reports that the interview date was 7/9/2002. Public use data set edited to reflect the accurate date (IntDate=7092002).

Module 12: Dogs and Dog Bites

Respondent ID 45580

Medical treatment of a dog bite was indicated (DB3=1), yet no evidence of a bite or medical treatment.

SC2	SC3	SC4	K2_01	K2U_01	K3_01	K2_02	K2U_02	K3_02
0 Men	1 Woman	2 Child(ren)	11	1:Years	1:Boy	5	1:Years	2:Girl

DB1	DB2_01	DB2_02	DB2_12	DB3	DB4_01	DB4_02	DB4_12
1:Yes	8:DK	8:DK	8:DK	1:Yes	7:NA	7:NA	7:NA

Comment: There appears to be no evidence that anyone was treated medically for a dog bite, or that anyone was bitten, for that matter. The data for this respondent has been recoded in the public use data set as follows … DB1=8; DB2_01=7; DB2_02=7; DB2_12=7; DB3=8;

Respondent ID 12921

No dog bite was indicated for child 2 (DB2_02), child 3 (DB2_03), the respondent (DB2_12) or other adult (DB2_13) in the household, yet information about treatment of a dog bite for these individuals appears to have been asked (DB4_*, *=02, 03, 12, 13).

SC2	SC3	SC4
1 Man	1 Woman	3 Child(ren)

K2_01	K2U_01	K3_01	K2_02	K2U_02	K3_02	K2_03	K2U_03	K3_03
9	1:Years	2:Girl	6	1:Years	2:Girl	5	1:Years	2:Girl

DB1	DB2_01	DB2_02	DB2_03	DB2_12	DB2_13
1:Yes	1:Yes	2:No	2:No	2:No	2:No

DB3	DB4_01	DB4_02	DB4_03	DB4_12	DB4_13
1:Yes	1:Yes	2:No	2:No	2:No	2:No

Comment: Since no dog bite was indicated for child 2, child 3, the adult respondent, or other adult in the household, the data on medically treated dog bites for these individuals has been recoded in the public use data set as follows … DB4_02=7; DB4_03=7; DB4_12=7; DB4_13=7;