

CCSS Data Checking Protocol Manual

This manual contains the checking protocols used when processing data from the baseline and follow-up questionnaires. These were developed in consultation between the Coordinating Center and the Statistical Center.

After the contents of the questionnaire are finalized, several meetings are held to discuss what errors and inconsistencies may arise on each questionnaire, and the best approaches for resolving them. These discussions result in a checking protocol, a SAS program to identify the errors, and an additional program to edit the data (where this can be done in a global manner, as opposed to subject specific edits). After the first few batches of data, the checking protocol is reviewed, at which point existing checks may be amended or removed, and new ones added. For consistency, any additional or more restrictive checks are applied retrospectively to the early data batches.

Data edits are performed at the Coordinating Center, after which the checking program is run again to ensure everything has been corrected. As a final check, the program is run one more time at the Statistical Center, and any outstanding issues are resolved before the new data are incorporated into the existing datasets for analysis.

Updated October 1, 2020

Protocol for resolving SAS dataset errors and inconsistencies – original cohort baseline questionnaire

Pre-scan edit procedure

I. Overview

Look over each page and erase all stray marks and pressure marks (i.e., bubble marks that have been transferred from one page to the next in the process of filling out the questionnaire) especially in the answer grids.

Give applicable CRA any booklets that have machine unreadable answers, for example, booklets in which answers are crossed out or circled, or booklets filled out in pen. These should be pretty obvious from just a casual inspection.

For all questions, correct any obvious incomplete erasures. Make sure that numbers entered into boxes match numbers bubbled in. If not, change the bubbles to match the written numbers.

II. Question by Question

FACE PAGE

Enter the CCSSID on the code sheet. You should find this hand written on the face page. If not, you can find it in the CCSSREG database. (NEED TO MAKE AN RBASE DATA ENTRY/EDIT FORM FOR PERSON EDITING BOOKLETS>)

Enter "Questionnaire date" on the code sheet. This is "Today's date" if provided by the respondent, otherwise it is the date we received the questionnaire (date stamped on face page).

Enter "Maiden name" and "Person completing this questionnaire" into CCSSREG database.

SECTION A

A1. Check that numbers in the boxes match the numbers bubbled in. If missing, enter date of birth from information in the CCSSREG database.

A2. Check for obvious discrepancy between reported sex and patient's name. Compare to information in the registration database if ambiguous.

A3. If a social security number is filled in on the questionnaire, make sure the "Do not have a social security number" bubble is empty. Check that numbers in the boxes match the numbers bubbled in. Compare social security number to information in the registration database. If no SS# exists in reg, enter it from the questionnaire. If database information conflicts with questionnaire, do not change information in the database. We may use both SS#s for patient tracing.

A4. If "other" specifics are given, enter details on codesheet and make sure the "other" bubble is filled in. If details given actually refer to one of the listed options, erase the "other" bubble and fill in the appropriate bubble. A list of equivalents for each option will be provided.

A5. If multiple answers are bubbled or the answer is missing or "no", but A5a is answered (i.e., not missing), change A5 to "yes".

A7. Check that numbers in the boxes match the numbers bubbled in.

A8.If "other" specifics are given, enter details on codesheet and make sure the "other" bubble is filled in. If details given actually refer to one of the listed options, erase the "other" bubble and fill in the appropriate bubble. A list of equivalents for each option will be provided.

A9.If "other" specifics are given, enter details on codesheet and make sure the "other" bubble is filled in. If details given actually refer to one of the listed options, erase the "other" bubble and fill in the appropriate bubble. A list of equivalents for each option will be provided.

A10.Check that numbers in the boxes match the numbers bubbled in.

A11.Check that numbers in the boxes match the numbers bubbled in.

SECTION B

B1.This is a MARK ALL THAT APPLY question. Multiple answers are possible. If "none" is bubbled and other answers are also given, erase "none".

If "other" is bubbled and the care provider specified is actually a physician or other listed health care provider, erase the "other" bubble and fill in the appropriate bubble if it is not already filled in.

If the care provider is genuinely "other", enter the information on the codesheet. A list of equivalents for each option will be provided.

B2.This is a MARK ALL THAT APPLY question. Multiple answers are possible. If "other" specifics are given, enter details on codesheet and make sure the "other" bubble is filled in. If details given actually refer to one of the listed options, erase the "other" bubble and fill in the appropriate bubble. A list of equivalents for each option will be provided.

B6.Check that numbers in the boxes match the numbers bubbled in.

B8.Enter all medications on the codesheet, in the order in which they are listed on the questionnaire. For each medication listed in parts B8.1-B8.15, fill in the first two bubbles of the grid for the drug category (01-15) and the next 4 bubbles for the drug code. A list of drugs will be provided.

For medications listed in part B8.16, use the first two bubbles of the grid to code the reason given for use of the drug and the next 4 bubbles for the drug code. Reasons will start with #17 and can go up to #99. We will create a codelist for this question as we go along and will refer to the codelist to avoid assigning multiple codes to the same reason. Check with Ann to resolve ambiguities. Drugs listed with no reason given should receive a "reason" code of 16.

B9.If "other" specifics are given, fill in details on codesheet and make sure the "yes" bubble is filled in for "other". If details given actually refer to one of the listed options, erase the "other" bubble and fill in the appropriate bubble. A list of equivalents for each option will be provided.

SECTION C-J. MEDICAL CONDITIONS

For each indicated condition, enter years on the codesheet. If an age is given but the question is not answered "yes", fill in the "yes" bubble, unless it's possible that an error of alignment has occurred (bubble filled on one line, age given on the next), in which case check with CRA.

C7.Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

C12.Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

C15. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

C17. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

D5. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

E12. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

E15. If filled in for females, have an MD review questionnaire. Erase if appropriate.

E16-18. If filled in for males, have an MD review questionnaire. Erase if appropriate.

E17. Record an age on the codesheet if one is given. (This question is out of alignment with others on the page, so may be easy to miss.)

F20. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

F21. If filled in for males, have an MD review questionnaire. Erase if appropriate.

G13. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

H2. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

H6. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

H8. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

H11. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

I1. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

I3. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

I5. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

I6. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

I10. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

I14. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

I20. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

I27. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

I31. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled. (There are three grids on the codesheet for this question. If more than three "other surgeries" are reported, use a second codesheet.)

J2. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

J5. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

J15. Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

J16-J38. Check for multiple marks. Correct any obvious incomplete erasures.

J36. If locations given, enter on codesheet. There are 3 grids on the codesheet for answers to this question.

SECTION K. CANCER, LEUKEMIA OR TUMORS

Missing answers in Section K will necessitate telephone recontact of the patient or a recheck of the medical record. Flag booklet with a post-it note marked "Section K" and give to Ann if:

a) there is no information given in Section K, or if K1-K4 are all missing but any of K5-K8 are answered (telephone recontact);

b)K1 is "yes" but any of K2-K4 is missing (medical records check);

c)K5 is "yes" but any of K6-K8 is missing (medical records check).

K1.If multiple, missing or "no" but K2, K3 or K4 is answered, change K1 to "yes".

K2.Record disease code on codesheet.

K4.Record date on codesheet.

K5.If multiple, missing or "no" but K6, K7 or K8 is answered, change K5 to "yes".

K6.Record disease code on codesheet.

K8.Record date on codesheet.

SECTION L. MARITAL STATUS

L1.If multiple, missing or "no" but any of L2-L5 is answered, change L1 to "yes".

L4.Check that numbers in the boxes match the numbers bubbled in.

L6.If multiple, missing or "yes" but L7 or L8 is answered, change L6 to "no".

L7.Check that numbers in the boxes match the numbers bubbled in.

L9.Check that numbers in the boxes match the numbers bubbled in.

L11.If multiple, missing or "yes" but L12 or L13 is answered, change L11 to "no".

L12.Check that numbers in the boxes match the numbers bubbled in.

SECTION M: OFFSPRING/PREGNANCY HISTORY

Missing answers in Section M will necessitate telephone recontact of the patient. If M9, M10 and M11 are all missing flag booklet with a post-it note marked "Section M" and give to Ann.

M1.If multiple, missing or "no" but any other positive answer is given in Section M ("yes" to M2, birth control use reported in M3, "yes" to any of M4-M9, pregnancies reported in M10 or M11) then change M1 to "yes".

M3.Enter "other" details on codesheet. If details given actually refer to one of the listed options, erase the "other" bubble and fill in the appropriate bubble. A list of equivalents for each option will be provided.

M7.Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

M8.Enter specifics on codesheet. If specifics are given, make sure "yes" is bubbled.

M9.If multiple, missing or "no" but a number is entered in M10 or pregnancies are reported in M11, change M9 to "yes".

M10. Enter number on codesheet.

M11. Check for multiple marks. Correct any obvious incomplete erasures and alignment problems. If additional pregnancies are reported either on a separate sheet or in the margins, enter details on codesheet. If more than 15 pregnancies are reported, use a second code sheet.

SECTION N. HEALTH HABITS

N1. All parts. Make sure numbers in boxes match the numbers bubbled in.

If N1 is multiple, missing or "no" but any of N1a, N1b or N1c contains a non-zero answer, change N1 to "yes".

N2. Check for multiple marks. Correct any obvious incomplete erasures.

N3. If multiple, missing or "no" but any of N4-N7 contains a positive answer, change N3 to "yes".

N4. Check that numbers in the boxes match the numbers bubbled in.

N6. Check that numbers in the boxes match the numbers bubbled in.

N18. If filled in for females, have an MD review questionnaire. Erase if appropriate.

N19-22. If filled in for males, have an MD review questionnaire. Erase if appropriate.

N22. Enter age on codesheet. If multiple, missing or "no" but an age is given, change N22 to "yes".

SECTION O. SCHOOL/EMPLOYMENT HISTORY

O3. Check for multiple marks. Correct any obvious incomplete erasures.

O4. This is a MARK ALL THAT APPLY question. Multiple answers are possible.

O5. If multiple, missing or "no" but Q7 has a non-zero answer or Q8 is answered, change O5 to "yes".

O8. Record details on codesheet.

O9. Check that numbers in the boxes match the numbers bubbled in.

SECTION P. FAMILY HISTORY INFORMATION

This section will be key entered rather than scanned. Missing answers in Section P will necessitate telephone recontact of the patient. Flag booklets with any obvious discrepancies or omissions with a post-it note marked "Section P". Give these to Ann.

SECTION Q. INSURANCE/RELIGION/INCOME/OTHER ISSUES

Q2. If multiple, missing or "no" but Q3 is answered, change Q2 to "yes". If "Canadian Resident" is bubbled but any of Q3,

Q3a, or Q3b is filled in, check the registration database to make sure the patient lives in Canada. If the patient is not a Canadian resident, change Q2 to "yes". Do not recode Q3.

Q3.This is a MARK ALL THAT APPLY question. Multiple answers are possible. If "other" specifics are given, enter details on codesheet and make sure the "other" bubble is filled in. If details given actually refer to one of the listed options, erase the "other" bubble and fill in the appropriate bubble. A list of equivalents for each option will be provided.

Q3a.If specifics are given, enter details on codesheet and make sure "yes" is bubbled.

Q5.If multiple, missing or "no" but Q6 is filled out, change Q5 to "yes".

Q6.This is a MARK ALL THAT APPLY question. Multiple answers are possible. If "other" specifics are given, enter details on codesheet and make sure the "other" bubble is filled in. If details given actually refer to one of the listed options, erase the "other" bubble and fill in the appropriate bubble. A list of equivalents for each option will be provided.

Q6a.If specifics are given, enter details on codesheet and make sure "yes" is bubbled.

Q7.If "other" is specified, enter details on codesheet and make sure the "other" bubble is filled in. If details given actually refer to one of the listed options, erase the "other" bubble and fill in the appropriate bubble. A list of equivalents for each option will be provided.

R1-6.Check for multiple marks. Correct any obvious incomplete erasures.

R6.Enter details in CCSSREG database.

R7.Enter details in CCSSREG database.

R8a.Enter details in CCSSREG database.

R9.Enter details in CCSSREG database.

Protocol for resolving SAS dataset errors and inconsistencies – original cohort baseline questionnaire

SAS dataset edit procedure

Section A

1. Date of questionnaire completion (front page) missing
Use date received in Minneapolis. Amend using FSP. Checked at pre-scan edit.
2. Date of birth (A1) missing
Obtain from CCSS institution or other records. Amend using FSP. Checked at pre-scan edit.
3. Twin (A5 & A5a) missing, or answered Y and number of siblings (A7) answered 0.
Examine details of siblings in section P2.
If (only) one sibling birth date matches the patient's birth date then the patient is a twin.
 - a. If the sibling is the same sex, code A5 as "not sure what type, same sex".
 - b. If the sibling is the opposite sex, code A5 as "fraternal, opposite sex".
 - c. If A7 is 0 and P2 is blank, code A5 as "not a twin".
 - d. Otherwise, code A5 as ambiguous.Amend using FSP.
4. Date of questionnaire completion (front page) earlier than 1994 or later than today's date.
Check questionnaire for coding errors. If year incorrect but rest of date is OK, change year to current. Otherwise, use date received in Minneapolis. Amend using FSP.
5. Alive/dead missing or coded 9
Check front page of questionnaire - is the person completing the questionnaire the CCSS subject? Refer to other medical records if necessary. Amend using FSP.
6. .Date of diagnosis missing
Obtain date of diagnosis (from institution?) and enter using FSP.
7. Date of diagnosis earlier than date of birth or later than date of completion or death.
Obtain correct date of diagnosis (from institution?) and/or dates of birth/death. Update using FSP.
8. Age at diagnosis greater than 21, or survival time less than 5 years. Check dates. If correct, case is ineligible - remove from dataset.

Section B

1. Health care providers (B1) answered "none" and other answers (e.g. "physician") given.
Delete "none", i.e. code "none"=N (SAS program). Checked at pre-scan edit.
2. Not seen a physician (B1) but at least one physician visit in B3.
Recode both to ambiguous (SAS program). Also recode cancer physician visits (B4) to ambiguous if non-missing.
3. Seen a physician (B1) but number of visits in B3 is zero.
Recode both to ambiguous unless B4 is non-zero, in which case B3=B4 (FSP).
4. Number of physician visits concerning cancer (B4) exceeds total number of physician visits (B3).
Recode both to ambiguous (SAS program).
5. deleted 11/27
6. More than 40 hospital admissions (B6).
Check questionnaire for scanning/coding errors and consistency. Recode to ambiguous (FSP) if scanning was accurate.

7. deleted 8/9
8. Surgeries in last 2 years (B7) answered N and evidence in part I of surgery in period 7/1/92 - 6/30/94.
Check questionnaire for scanning/coding errors. Recode B7 to ambiguous (FSP) if scanning was accurate. Note that an unambiguous error only occurs if (age) - (surgery age) = 1
(For questionnaire completion dates before 7/1/95. NB Change program if any completion dates later than this).
9. Birth control pills (B8.2) answered Y for males.
Recode to N (SAS program). If any drugs are listed in the box, delete from B8.2 and enter them as female hormones in B8.3 (?) using FSP.
10. Birth control pills (B8.2) answered Y and birth control pills (M3) answered "never used".
Recode both to ambiguous unless a (valid) pill name is given in B8.2, in which case recode M3 only to ambiguous (FSP).
11. Female: Birth control pills (B8.2) answered N and birth control pills (M3) answered "currently use".
Recode both to ambiguous (SAS program). Female hormones (B8.3) answered Y for males, or male hormones (B8.4) answered Y for females.

Check questionnaire for scanning/coding errors. If no errors, review by an MD for decision on editing. Update using FSP if necessary.
12. deleted 8/9
13. deleted 8/9
14. deleted 8/9

Sections C-J

- C-J1. Questions answered N and an age at first occurrence is given.
Recode both automatically to ambiguous unless the question has an "if yes, describe" box in which details are given. In this case, recode N to Y (FSP).
- C-J2. Age of first occurrence is greater than patient's current age.
Check questionnaire to make sure data entry was correct. If so, recode age of first occurrence to ambiguous (FSP).
- E3. Deleted 11/27
- E4. Deleted 12/8
- E5. Deleted 12/8
- E6. Low sperm count (E15) answered by female.
Recode E15 to missing. Checked at pre-scan edit. (Have an MD review questionnaire, then correct using FSP).
- E7. Age at last menstrual period (E17) less than age at first menstrual period (E16).
Check questionnaire to make sure data entry was correct. If so, recode both to ambiguous (FSP).
- E8. Menstrual periods (E16/17/18) answered by male.
Recode E16/17/18 (not sex) to missing. Checked at pre-scan edit. (Have an MD review questionnaire, then correct using FSP).
- E9. No menstrual periods (E16) but answers given for currently having periods (E17) or taking hormones (E18).
Recode both or all three to ambiguous (SAS program). If E17/E18 answered N, or E17 missing and E18 answered, recode E17/E18 to missing (SAS program).

F3. Deleted 12/8

F4. Heart failure during pregnancy (F21) answered by male.

Recode F21 (not sex) to missing. Checked at pre-scan edit. (Have an MD review questionnaire, then correct using FSP).

F5. Heart failure during pregnancy (F21) answered Y but never sexually active (M1) or never been pregnant (M9)

Recode F21 to N (SAS program).

G3. Deleted 12/8

G4. Second part of G9 (oxygen now) answered, and first part (ever had extra oxygen) answered "No" or "not sure".

If second part is "No" or "Not sure", recode to missing. Otherwise, recode both to ambiguous (SAS program).

H3. Heartburn medication (H10) answered Y and age at first occurrence missing.

Check questionnaire to see if an age was coded (this age was missed off the bubble sheet).

H4. Second part of H10 (heartburn medication) answered, and first part (heartburn) answered "No" or "not sure".

If second part is "No" or "Not sure", recode to missing. Otherwise, recode both to ambiguous (SAS program).

I3. Amputation (I1) answered Y but loss of limb or digit (B9) answered N.

If details of amputation are given in I1, recode B9 to Y. Otherwise, recode both to ambiguous. (Have a MD review the questionnaire, and then correct using FSP).

I4. Deleted 12/8

I5. Coronary artery bypass surgery (I7) answered Y but coronary heart disease (F6) answered N.

Recode both to ambiguous. (Have a MD review the questionnaire, and then correct using FSP).

I6. Pericardiectomy (I8) answered Y but pericarditis (F11) and pericardial constriction (F12) both answered N.

Recode all three to ambiguous. (Have a MD review the questionnaire, and then correct using FSP).

I7. Age at colostomy takedown (I13) less than age at colostomy (I12).

Check questionnaire to make sure data entry was correct. If so, recode both to ambiguous. (Have a MD review the questionnaire, and then correct using FSP).

I8. Kidney transplant (I25) answered Y but dialysis (D4) answered N.

Check if any details in D5 which suggest a kidney transplant may have been performed. If not, recode I25 and D4 to ambiguous. (Have a MD review the questionnaire, and then correct using FSP).

I9. Cataract surgery (I28) answered Y but cataracts (C9) answered N.

Recode both to ambiguous. (Have a MD review the questionnaire, and then correct using FSP).

Section K

Note: An MD will review this section if anything is noted, and will verify the answers given in K4 and K8 in relation to the other information given.

1. Second cancer (K1) answered N, but details given in K2, K3, K4.

Recode K1 to Y (SAS program). Checked at pre-scan edit.

2. Third cancer (K5) answered N, but details given in K6, K7, K8.

Recode K5 to Y (SAS program). Checked at pre-scan edit.

3. Missing information in section K.

Recontact the patient if no information is given in Section K, or if K1-K4 are all missing but any of K5-K8 is answered. If K1 is answered Y but K2-K4 are missing, or K5 is answered Y but K6-K8 are missing, attempt to get the missing information from the medical record. Enter new data using FSP.

4. Third cancer (K5) answered N and second cancer (K1) answered N, with K2, K3 and K4 missing.

Recode K5 to missing (SAS program).

Section L

1. Never been married (L1) but details of marriage given in L2-L13.

Recode L1 to Y (SAS program). Checked at pre-scan edit.

2. Currently in first relationship (L6) but year of ending (L7) or details of how ended (L8) are given.

Recode L6 to N (SAS program). Checked at pre-scan edit.

3. Currently in most recent relationship (L11) but year of ending (L12) or details of how ended (L13) are given.

Recode L11 to N (SAS program). Checked at pre-scan edit.

4. Not currently in first or most recent relationship (L6 or L11) but no details given in L7/L8 or L12/L13.

Recode all three to ambiguous (SAS program).

5. Inconsistency between current marital status (L2) and end of most recent relationship (L13) (and end of first relationship if only married once).

Recode both/all three to ambiguous (SAS program).

6. Year of first or most recent marriage (L4/L9) implies that age at the time was less than 12 (?).

Check questionnaire to make sure data entry was correct. If so, recode year to ambiguous (FSP).

7. End date of first/most recent relationship (L4/L9) precedes start date (L7/L12).

Check questionnaire to make sure data entry was correct. If so, recode start date and end date to ambiguous (FSP).

8. Any section L year is later than year of questionnaire completion.

Check questionnaire to make sure data entry was correct. If so, recode year to ambiguous (FSP).

9. Start date of most recent marriage precedes end date of first marriage (if married more than once).

Check questionnaire to make sure data entry was correct. If so, recode dates to ambiguous (FSP).

Section M

1. Never sexually active (M1) but currently sexually active (M2) and/or other evidence of sexual activity in section M.

Recode M1 to Y (SAS program). Checked at pre-scan edit.

2. Did not try for 1+ years to become pregnant (M5), saw doctor (M6) answered N and M7-M8 are blank or N.

Recode M6-M8 to missing (SAS program).

3. Did not try for 1+ years to become pregnant (M5), saw doctor (M6) answered Y and M7-M8 are blank.

Recode M6 and M7 to ambiguous (SAS program).

4. Did not try for 1+ years to become pregnant (M5), saw doctor (M6) answered Y and M7 or M8 answered.

Recode M5 to Y (SAS program).

5. Saw doctor (M6) answered N and reason unable to become pregnant (M7) answered.
6. If M7 is N, SAS program will recode to missing. If Y or not sure, check questionnaire to make sure data entry was correct. If a reason is given in M7, recode M6 to Y. Otherwise recode M6 and M7 to ambiguous (FSP).
7. Reason (M7) answered N and medication (M8) answered.
If M8 is N, SAS program will recode to missing. If Y or not sure, check questionnaire to make sure data entry was correct. If medication details are given in M8, recode M7 to Y. Otherwise recode M7 and M8 to ambiguous (FSP).
8. Ever pregnant (M9) answered N and number of pregnancies (M10) answered, or details given in M11.
Recode M9 to Y, unless M10 is zero, in which case recode M10 to missing (SAS program). Checked at pre-scan edit.
9. Sexually active (M1) answered Y; ever pregnant (M9), number of pregnancies (M10) and pregnancy details (M11) all missing.
Recontact patient.
10. Ever pregnant (M9) answered Y and number of pregnancies (M10) missing but details in M11.
Check questionnaire to make sure data entry was correct. If details given in M11, M10 = number of pregnancies in M11. Amend record using FSP.
11. Never sexually active (M1) but other section M questions answered.
If M2, M4-M9 are missing or answered N, and M10 is missing or 0, recode M2-M10 to missing. Otherwise, recode M1 to Y (SAS program).
12. Age at beginning of pregnancy (M11) exceeds current age.
Recode age in M11 to ambiguous (SAS program).
13. Deleted 12/8

Section N

1. Never smoked (N1) but N1a, N1b or N1c answered.
Check questionnaire to make sure data entry was correct. Recode N1 to Y unless number of cigarettes (N1b) and/or years smoked (N1c) are zero, in which case recode N1-N1c as missing (FSP).
2. Never smoked (N1), N1a-c blank but smoke now (N1d) answered.
Check questionnaire to make sure data entry was correct. If N1d answered N, recode N1d to missing (SAS program). If N1d answered Y, recode N1 and N1d to ambiguous (FSP).
3. Never smoked (N1), N1a-d blank, but whether tried to stop in last 2 years (N1e) and/or age when stopped (N1f) answered.
Check questionnaire to make sure data entry was correct. If N1e answered N and N1f blank, recode N1e to missing (SAS program). Otherwise recode N1, N1e, N1f to ambiguous if N1f non-zero, or recode N1e and N1f to missing if N1f is coded zero (FSP).
4. Age when started smoking (N1a), stopped smoking (N1f), started drinking (N4) or age at first mammogram (N22) exceed current age.
Check questionnaire to make sure data entry was correct. If so, recode section N age to ambiguous (FSP).
5. Age when started smoking (N1a) exceeds age when stopped smoking (N1f).
Check questionnaire to make sure data entry was correct. If so, recode both ages to ambiguous (FSP).
6. deleted 8/9
7. Not smoking now (N1d) but answer to tried to stop in last 2 years (N1e).
Recode N1e to ambiguous (SAS program).
8. Did not try to stop in last 2 years (N1e) but age when stopped is given (N1f).
Check questionnaire to make sure data entry was correct. If N1d is answered N, recode N1e to missing. If N1d is answered Y, recode N1d, N1e and N1f to ambiguous. If N1d is missing, recode N1e and N1f to ambiguous (FSP).

9. Never used other forms of tobacco, but length of use given (N2).
Recode both to ambiguous (SAS program).
10. No more than 2 drinks in life (N3) but answers to any of N4-N8.
Check questionnaire to make sure data entry was correct. If answers in N4-N8 imply that respondent does drink, recode N3 as Y. Otherwise recode N4-N8 as missing (FSP).
11. Under-18 questionnaire only: if drinklyr=N, recode to missing automatically (SAS program).
12. Total number of drinks per month (N6) exceeds 300.
Check questionnaire to make sure data entry was correct. If so, recode to ambiguous (FSP).
13. 1.No drinks in last 2 years (N7) but number of drinks (N6) non-zero.
14. One drink per day (N7) but total in N6 exceeds 60. Two drinks per day (N7) but total in N6 exceeds 90.
15. Three drinks per day (N7) but total in N6 exceeds 120.
16. Four drinks per day (N7) but total in N6 exceeds 150.
17. Five drinks per day (N7) but total in N6 exceeds 180.
For all the above, check questionnaire to make sure data entry was correct. If so, recode N6 and N7 to ambiguous (FSP).
18. General physical examination within last 2 years (N16) but did not see physician within last 2 years (B1).
Recode N16 to ambiguous (SAS program).
19. Testicular examination (N18) answered by female.
Recode N18 to missing. Checked at pre-scan edit. (Have an MD review questionnaire, then correct using FSP).
20. Breast examination, pap smear or mammogram (N19-N22) answered by male.
Recode N19-N22 to missing. Checked at pre-scan edit. (Have an MD review questionnaire, then correct using FSP).
21. Never had a mammogram, but age at first mammogram is given (N22).
Check questionnaire to make sure data entry was correct. If so, recode N22 to Y (FSP). Checked at pre-scan edit.

Section O

1. Did not complete high school (O1) but received standard high school diploma (O2).
Recode both to ambiguous (SAS program).
2. Not in a learning disabled/special education program (O3) but reasons given in O3, or grades given in O4.
Recode all to ambiguous (SAS program).
3. Never had a job (O5) but time since last job (O7) or type of business (O8) answered.
Check questionnaire to make sure data entry was correct. If so, recode O5 to Y (FSP).
4. Worked in last 12 months (O6) but time since last job (O7) answered.
Check questionnaire to make sure data entry was correct. If time in O7 is less than 12 months, or answer suggests respondent has misread the question as "total time worked", recode O7 to missing. Otherwise recode O6 to N (FSP).
5. Did not work in last 12 months (O6) but time since last job (O7) is less than 12 months. Check questionnaire to make sure data entry was correct. If so, recode O6 to Y and O7 to missing (FSP).
6. Time since last job (O7) within 10 years of patient's age.
Check questionnaire to make sure data entry was correct. If so, recode O7 to ambiguous (FSP).

7. Never had a job (O5) or did not work within last 12 months (O6) but sick time (O9) answered.

Recode O9 to missing (SAS program).

8. Turned down by military, police or fire department (O11) but did not apply (O10).

Recode both to ambiguous (SAS program).

Section P

1. Relative's status is alive but a date of death is given.

Recontact patient. Amend record after interview using FSP.

2. Relative's date of death precedes date of birth.

Check questionnaire to make sure data entry was correct. If so, recontact patient to verify the answer, then correct the response (FSP).

3. Relative's date of birth or death is later than date questionnaire completed.

Check questionnaire to make sure data entry was correct. If so, recontact patient to verify the answer, then correct the response (FSP).

4. Patient's date of birth (P1) does not match that given in A1.

Check both dates against other records. Correct whichever is wrong (FSP).

5. Number of full siblings listed in P2 exceeds number of full siblings given in A7.

Recode A7 to match number in P2 (SAS program).

6. Sibling (P2) is more than 20 years older or younger than patient.

Recontact patient. Amend record after interview using FSP.

7. Parent (P3) is less than 13 years older than patient.

Recontact patient. Amend record after interview using

FSP. Age of onset is greater than relative's age.

Recontact patient. Amend record after interview using FSP.

8. Child of patient (P6) less than 13 years younger than patient.

Recontact patient. Amend record after interview using FSP.

9. Patient (P1) or sibling (P2) born after date of death of parent (P3).

Check questionnaire to make sure data entry was correct (and date not within 9 months of father). If appropriate, recontact patient to verify the answer, then correct the response (FSP).

10. Child of patient (P4) born after date of death of other parent (P7) or patient

Check questionnaire to make sure data entry was correct (and date not within 9 months of father). If appropriate, recontact patient to verify the answer, then correct the response (FSP).

11. Female with hypospadias, cryptorchism, or Wiskott-Aldrich or Klinefelter's syndromes, or male with Turner's syndrome.

Check questionnaire to make sure data entry was correct. If so, delete the condition (FSP).

Section Q

1. Health insurance (Q2) answered N but health insurance details (Q3/3a/3b) answered.

Recode Q2 to Y (SAS program).

2. Current health insurance (Q2) but no details given in Q3 for source of health insurance.

Recode the six "source" parts of Q3 as missing (SAS program).

3. No current life insurance (Q5) answered N but life insurance details (Q6/6a/6b/6c) answered.

Recode Q5 to Y (SAS program).

4. Current life insurance (Q5) but no details given in Q6 for source of life insurance.

Recode the four "source" parts of Q6 as missing (SAS program).

5. Personal income (Q9) exceeds household income (Q8) where both are non-missing.

Recode both to ambiguous (SAS program).

Protocol for resolving SAS dataset errors and inconsistencies – follow-up 1 questionnaire

Initial checks (Minneapolis)

1. Date of questionnaire completion (frontpage) missing.
Use date received in Minneapolis.
2. Date of questionnaire completion earlier than 2000 or later than today's date.
Check questionnaire for coding errors. If year incorrect but rest of date is OK, change year to current. Otherwise, use date received in Minneapolis.
3. 1a (education correct) answered No and no details in box.
Check questionnaire to ensure no coding error. If box is empty, change 1a to Not sure.
4. 2a (marital status correct) answered No and no details in box.
Check questionnaire to ensure no coding error. If box is empty, change 2a to Not sure.
5. 3a (employment status correct) answered No and no details in box.
Check questionnaire to ensure no coding error. If box is empty, change 3a to Not sure.
6. 3b (worked since baseline) answered No and details given.
Change 3b to Yes.
7. 4p (other genetic conditions) answered No or Not Sure, and details given in box.
If details appear to be a genetic condition, change 4p to Yes. Otherwise, delete box entry.
8. 5l (other congenital heart defects) answered No or Not Sure, and details given in box.
If details appear to be a congenital heart defect, change 5l to Yes. Otherwise, delete box entry.
9. 5p (other birth defects) answered No or Not Sure, and details given in box.
If details appear to be a birth defect, change 5p to Yes. Otherwise, delete box entry.
10. RT data correct (7) answered No and no details given.
Recontact patient.
11. RT data correct (7) answered Yes, followed by “No, did not receive radiation”.
Check questionnaire for coding/bubbling errors. If no problems, change NORAD (did not receive radiation) to missing.
12. 8 (pregnancy since baseline) answered Yes and no details in 8a.
Check questionnaire to ensure no coding/scanning error.
13. 8 (pregnancy since baseline) answered No and details given in 8a.
Change 8 to Yes.

14. Relative's status (8b, 8c) is alive but a date of death is given.
Recontact patient.
15. Relative's date of death (8b, 8c) precedes date of birth.
Check questionnaire to make sure data entry was correct. If so, recontact patient to verify the answer.
16. Relative's date of birth or death (8b, 8c) is later than date questionnaire completed.
Check questionnaire to make sure data entry was correct. If so, recontact patient to verify the answer.
17. Age of onset is greater than relative's age.
Recontact patient.
18. 9 (cardio/pulmonary conditions correct) answered No and no details in box.
Check questionnaire to ensure no coding error. If box is empty, change 9 to Not sure.
19. 9 (cardio/pulmonary conditions correct) answered Yes and details given.
If entry in box is same as listing printed in 9, delete box entry. Otherwise, change 9 to No.
20. 10, 11 answered N and an age at first occurrence is given.
Recode both to ambiguous unless the question has an "if yes, describe" box in which details are given. In this case, recode No to Yes.
21. Second part of 11i (oxygen now) answered, and first part (ever had extra oxygen) answered "No" or "not sure".
If second part is "No" or "Not sure", recode to missing. Otherwise, recode both to ambiguous.
22. First part of 13 missing or answered No and second part is answered.
Change first part of 13 to Yes.
23. Health insurance (16) answered No or Canadian but health insurance details (16a, b) answered.
If all details are No, recode them to missing. Otherwise, recode 16 to Yes unless address (page 15) is in Canada.
24. Health insurance (16) answered Yes but nothing is bubbled in 16a.
Recode the six "16a" variables as missing.
25. 16b answered No or Don't know and details given in box.
Recode 16b to Yes.
26. New cancer (17) answered N, but details given in 17, 17a.
Recode 17 to Yes.
27. Details given in 17a but not in 17.
Move data from 17a to 17.
28. Missing information in 17, 17a.

Recontact patient.

29. 18q answered No or Not sure and details given in box.

Recode 18q to Yes.

30. 19 answered No and 19a-d answered.

If 19b is answered, or 19c is answered a, recode 19 to Yes. Otherwise, recode 19 to ambiguous.

31. 19b missing, but 19c answered a or b, and 19d not answered.

Use current age in 19b.

32. 19c not answered “Other” and details given in box.

If 19c is blank, or details given indicate that “Other” should be correct, recode as “Other”. Otherwise, delete box details.

33. 19c answered “currently pregnant”.

If no “Currently pregnant” record in pregnancy data (8a) add a record with outcome “Currently pregnant”, age equal to current age, and weeks missing.

34. 19d answered and 19c answered a, b, c or d.

Delete 19d.

35. 19d not answered “Other” and details given in box.

If 19d is blank, or details given indicate that “Other” should be correct, recode as “Other”. Otherwise, delete box details.

36. 21a answered No or missing, and number of hospitalizations or 21b, c answered.

If dates are later than baseline date, or are missing, change 21a to Yes. Otherwise, delete number of hospitalizations, 21b, c.

37. Details given in 22 are a heart or lung problem, paralysis, seizure, or new cancer.

If the information does not duplicate that given in 10n, 11m, 12f, 12h, or 17, add to the appropriate variable. Delete from 22.

38. 22 answered No or Not sure and details given in box.

If details are an appropriate answer for this question (see above), recode 22 to Yes. Otherwise, delete box details.

39. Data in “extra” variables.

Enter additional data from questionnaire.

40. Gaps in sequence – most recent job (3b)

41. Gaps in sequence – other genetic disorders (4p)

42. Gaps in sequence – other heart defects (5l)

43. Gaps in sequence – other birth defects (5p)

44. Gaps in sequence – relative conditions (7)
45. Gaps in sequence – other heart conditions (10)
46. Gaps in sequence – other lung problems (11m)
47. Gaps in sequence – paralysis (12f)
48. Gaps in sequence – seizures, convulsions or blackouts (12h)
49. Gaps in sequence – other health insurance provider (16a)
50. Gaps in sequence – health insurance exclusions (16b)
51. Gaps in sequence – other alternative medicine (18q)
52. Gaps in sequence – hospitalizations (21b/c)
53. Gaps in sequence – other serious medical conditions (22)

For all the above: Check questionnaire. If nothing has been missed, shift entries to the left to eliminate gaps.

54. Female patient with Klinefelter's (4g) or Wiskott-Aldrich (4n) syndromes, or male with Turner's syndrome (4l).
Check questionnaire to make sure data entry was correct. If so, delete the condition.

55. 5o (undescended testes) answered for females.

Change to missing.

56. Birth control pills (6b) answered Yes for males.

Recode to No. If any drugs are listed in the box, delete from 6b and enter them as female hormones in 6c using FSP.

57. Female hormones (6c) answered Yes for males, or male hormones (6d) answered Yes for females.

Check questionnaire for scanning/coding errors. If no errors, review by an MD for decision on editing. Update using FSP if necessary.

58. Age of first occurrence in 10, 11, 12 is greater than patient's current age.

Check questionnaire to make sure data entry was correct. If so, recode age of first occurrence to ambiguous.

59. 19-20 answered by male, or by female under 18

Recode to missing.

60. 19a age greater than 19b age, or either greater than current age.

Check questionnaire for coding errors. If correctly entered, change both to ambiguous.

Checks against other datasets (Seattle)

1. 1a (education correct) answered Yes and details given.

Check baseline data. If entry in box is different, change 1a to No. Otherwise, delete box entry.

2. 1b (more education) answered No and details given. Check baseline data. If details are higher than in baseline, change 1b to Yes. Otherwise, delete details.

3. 2a (marital status correct) answered Yes and details given in box.

Check baseline data. If entry in box is different, change 2a to No. Otherwise, delete box entry.

4. 2b (changed marital status) answered No and details given.

Check baseline data. If details are different in baseline, change 2b to Yes. Otherwise, delete details.

5. 3a (employment status correct) answered Yes and details given in box.

Check baseline data. If entry in box is different, change 3a to No. Otherwise, delete box entry.

6. Female patient with Klinefelter's (4g) or Wiskott-Aldrich (4n) syndromes, or male with Turner's syndrome (4l).

- Check questionnaire to make sure data entry was correct. If so, delete the condition.
7. Hereditary conditions listed in section P baseline for patient, but not listed in 4a-p.
Change appropriate question 4 variable to Yes.
 8. 5d (deafness at birth) answered Yes, and none of 12a-c or baseline C1-3 have an age at first occurrence of 0.
Change earliest age at first occurrence to 0 (if in follow-up data), unless all questions are answered No, in which case recode 5d to ambiguous.
 9. 5e (blindness at birth) answered Yes, and neither 12d, e nor baseline C8-13 have an age at first occurrence of 0.
Change earliest age at first occurrence to 0 (if in follow-up data), unless all questions are answered No, in which case recode 5e to ambiguous.
 10. 5o (undescended testes) answered for females.
Change to missing.
 11. Birth control pills (6b) answered Yes for males.
Recode to No. If any drugs are listed in the box, delete from 6b and enter them as female hormones in 6c using FSP.
 12. Female hormones (6c) answered Yes for males, or male hormones (6d) answered Yes for females.
Check questionnaire for scanning/coding errors. If no errors, review by an MD for decision on editing. Update using FSP if necessary.
 13. RT data correct (7) answered Yes or Not sure and details given, and/or RT treatment dates earlier than year printed in question 7.
Check MRAF data. If data are not present for dates given, contact MD Anderson. If data are present, disregard treatment information in questionnaire.
 14. Follow-up data from 8a start with pregnancy number 1, and baseline pregnancy data exist.
Reassign follow-up pregnancy numbers.
 15. Child's date of birth (8b) precedes baseline or pregnancy questionnaire date.
If child is in baseline family history data, delete follow-up data. Otherwise, move child data to baseline, correcting relative IDs if necessary.
 16. Other parent (8c) or patient is less than 13 years older than offspring (8b).
Recontact patient.
 17. Child of patient (8b) born after date of death of other parent (8c) or patient
Check questionnaire to make sure data entry was correct (and date not within 9 months of father). If appropriate, recontact patient to verify the answer.
 18. Female relative (8b, 8c) with hypospadias, cryptorchism, or Wiskott-Aldrich or Klinefelter's syndromes, or male with Turner's syndrome.
Check questionnaire to make sure data entry was correct. If so, delete the condition.
 19. Age of first occurrence in 10, 11, 12 is greater than patient's current age.
Check questionnaire to make sure data entry was correct. If so, recode age of first occurrence to ambiguous.
 20. Age at first occurrence in 10, 11, 12 is less than age at baseline completion.
If age of occurrence is missing in the baseline data, add the age. If the question (in 10, 11) is answered No, make sure it is included in the "correct information" list (9).
 21. Age of occurrence in 10, 11, 12 differs from equivalent age of occurrence in baseline.
Use youngest age in both datasets. If an "if yes, describe..." question, make sure the age refers to the same condition in both datasets.
 22. New cancer information in baseline section K, and in follow-up questionnaire (17) with date of occurrence missing or earlier than baseline date.
If follow-up data refer to a cancer already reported in the baseline, delete.

23. 19-20 answered by male, or by female under 18

Recode to missing.

24. 19a age greater than 19b age, or either greater than current age.

Check questionnaire for coding errors. If correctly entered, change both to ambiguous.

25. 19 answered Yes and baseline E16 answered No.

If answers to 19a-d are consistent with the answer to 19 being No, change it.

26. 21b dates are earlier than baseline dates.

Delete all of 21b.

Protocol for resolving SAS dataset errors and inconsistencies – follow-up 2 questionnaire

1. Date of questionnaire completion (frontpage) missing.
Use date received in Minneapolis.
2. Date of questionnaire completion earlier than 2001 or later than today's date.
Check questionnaire for coding errors. If year incorrect but rest of date is OK, change year to current. Otherwise, use date received in Minneapolis.
3. 1 (school now completed) Details given on other schooling but did not select OTHER or it is MISSING.
Change OTHER to Yes unless answer entered is clearly covered by the preceding options.
4. 3 (current living arrangement) lives alone but also live with others.
Change to ambiguous.
5. 3 (current living arrangement) Details given on other living arrangements, but did not select OTHER or it is MISSING.
Change OTHER to Yes unless answer entered is clearly covered by the preceding options.
6. 4 (current employment status) answered FULL or PART-TIME but also UNEMPLOYED or UNABLE.
Change to ambiguous.
7. 4 (current employment status) answered not FULL or PART-TIME but did not skip to Question 7.
Delete answers to 5 and 6.
8. A1 (saw doctor) answered YES for nohcare but answered YES to seeing other health care providers also.
Change to ambiguous.
9. A1 (saw other doctor) answered other than YES but gave details to seeing other health care providers.
Change OTHER to Yes unless answer entered is clearly covered by the preceding options.
10. Gaps in sequence - saw other doctor(A1)
Check questionnaire. If nothing has been missed, shift entries to the left to eliminate gaps.
11. A1 (saw doctor) answered NONE but did not skip to Question A9.
If other parts of A1 answered, or A2-8 indicate that patient saw a provider in last 2 years, change first part of A1.
Otherwise, delete answers to A2-8.
12. A2 (Visited other place) answered other than YES but gave details to visiting other place of treatment.
Change OTHER to Yes unless answer entered is clearly covered by the preceding options.
13. Gaps in sequence - visited other treatment(A2)
Check questionnaire. If nothing has been missed, shift entries to the left to eliminate gaps.
14. A3 (Times saw doctor) answered more than zero, but did not answer PHYSICIAN in A1(saw doctor).
Change both to ambiguous.
15. A3 (Times saw doctor) answered zero, but answered YES in A1(see doctor).
Change both to ambiguous.
16. A5 (visits related to cancer) answered more than zero, but answered NONE in A5(# of visits).
Change both to ambiguous.
17. A6 (what discussed in visits) did not answer NONE, but answered NONE in A3(# of times saw doctor).
Change both to ambiguous.
18. A6 (other cancer issues) answered other than YES, but gave details for cancer issues.
Change to YES.

19. Gaps in sequence - discuss other cancer issues(A6)
Check questionnaire. If nothing has been missed, shift entries to the left to eliminate gaps.
20. A7 (Doctor familiar with health problems) answered DID NOT HAVE A CHECK-UP, but did not proceed to Question A9.
If answers in A8 are consistent with having had a check-up, change A7. Otherwise delete A8.
21. A8 (other suggestions) answered other than YES but gave details to given other suggestions by doctor.
Change to YES.
22. Gaps in sequence - other suggestions(A8).
Check questionnaire. If nothing has been missed, shift entries to the left to eliminate gaps.
B4 (Mammogram)and B5(pap smear) Male but answered YES to either questions. Delete questions.
23. B6-B8(blood transfusion, hepatitis C). Gave age but answered NO or NOT SURE.
Change answer to YES.
24. B9(result of hepatitis C test). Answered NO in B8(test for hepatitis C), but did not select NOT SURE in B9.
Change to NOT SURE.
25. C6(as an adult, sunburn). Under 21 but answered C6.
Delete answers.
26. C7(ever sunbathed). Answered NO in C7, but answered C8(days of exposure).
Change C7 to YES.
27. C9(artificial tan). Answered NO in C9, but answered C10(days of exposure).
Change C9 to YES.
28. (deleted)
29. C14(Skin Cancer)Answered other than YES, but provided date of dx.
Change C14 to YES.
30. D2(vigorous activities). Answered NO in D2, but answered D3 (amount of days).
If D3 non-zero, change D2 to YES. Otherwise delete D3.
31. D5(moderate activities). Answered NO in D5, but answered D6(amount of days).
If D6 non-zero, change D5 to YES. Otherwise delete D6.
32. L1(ever smoked at least 100 cigarettes). Answered other than YES, but did not skip to L6.
If age given change L1 to YES. Otherwise, if answered NO without age delete L3, L4, L5, L6.
33. (deleted)
34. L1(age started smoking). Age at time of completing the questionnaire is less that age started smoking.Check data – if correctly entered, delete.
35. M1(have insurance). Answered, Canadian resident or NO, but did not skip to section N.
If answers are consistent with having insurance, change M1 to YES. Otherwise, delete M1a & M1b.
36. M1a (coverage provided) answered NO but gave details to how other coverage is provided.
If answers are consistent with having insurance, change M1 to YES. Otherwise, delete M1a.
37. Gaps in sequence - other coverage provided(M1a).
Check questionnaire. If nothing has been missed, shift entries to the left to eliminate gaps.
38. N1(pregnant since last questionnaire) answered NO, but did not skip to Section O (gave info about pregnancy), or answered YES for N2(currently pregnant) but answered NO for N1.
If responses indicate a recent pregnancy, change N1 to YES.
39. N1(pregnant since last questionnaire). Answered Yes, but no details given.

Check questionnaire. If nothing has been missed, leave as is.

40. N1(pregnant since last questionnaire). Answered NO, but details given.

If responses indicate a recent pregnancy, change N1 to YES.

41. Relative's status (N4) is alive but a date of death is given

Recontact patient.

42. Relative's date of death (N4) precedes date of birth

Check questionnaire to make sure data entry was correct. If so, recontact patient to verify the answer.

43. Relative's date of birth or death (N4) is later than date questionnaire completed.

Check questionnaire to make sure data entry was correct. If so, recontact patient to verify the answer.

44. Age of onset is greater than relative's age

Recontact patient.

45. Gaps in sequence - relative conditions (N5)

Check questionnaire. If nothing has been missed, shift entries to the left to eliminate gaps.

46. O1-14(General dental health): Age of first occurrence given but did not answer YES to question.

Check questionnaire to make sure data entry was correct. If so, change answer to YES.

47. O1-14(General dental health): Received care in last 2 years but didn't answer YES to initial question.

Check questionnaire to make sure data entry was correct. If so, change answer to YES.

48. (deleted)

49. Age of first occurrence in section O is greater than patient's current age.

Check questionnaire to make sure data entry was correct. If so, recode age of first occurrence to ambiguous.

50. (deleted)

51. O13(Other dental treatment): Answered NO, but gave age, answered about care or gave details.

Check questionnaire to make sure data entry was correct. If so, change answer to YES.

52. Gaps in sequence - other dental treatment(O13)

Check questionnaire. If nothing has been missed, shift entries to the left to eliminate gaps.

53. O14(Other dental problems): Answered No, but gave age, answered about care or gave details.

Check questionnaire to make sure data entry was correct. If so, change answer to YES.

54. Gaps in sequence - other dental problems(O14).

Check questionnaire. If nothing has been missed, shift entries to the left to eliminate gaps.

55. P1(have osteoporosis/osteopenia): Answered other than YES but gave age or about current information.

If age given, change P1 to YES.

56. P2 (taking meds for osteoporosis). Answered other than YES but did not skip to P4.

Change P2 and P3 to ambiguous.

57. P3g (Other medication for osteoporosis) answered No, but gave age, answered about care or gave details.

Change to YES.

58. Gaps in sequence - other medication for osteoporosis(P3g).

Check questionnaire. If nothing has been missed, shift entries to the left to eliminate gaps.

59. P3a-g(ever taken pills) answered NO, but YES to (currently taking).

Change to YES.

60. P3a-g: Answered NO to currently taking but indicated taking similar medicine in Section Q.

- Change to YES.
61. P4(have avascular necrosis) answered other than YES, but gave age.
Change to YES.
62. P4(have avascular necrosis) answered other than YES, but did not skip to Section Q.
Change to YES, unless all answers are consistent with No to P4, in which case delete them.
63. P5(received treatment for avascular necrosis) answered other than YES, but did not skip to Section Q.
If answered NO for P5, delete entries P6a-g; if MISSING for P5, but YES for P6a-g, change P5 to YES.
64. P6e(Other treatments for AVN) answered NO, but gave age, answered about care or gave details.
Change P6e to YES.
65. Gaps in sequence - other treatments for AVN (P6e)
Check questionnaire. If nothing has been missed, shift entries to the left to eliminate gaps.
66. P6a-e(receive treatment) answered NO but YES to currently receiving
Change to YES.
67. Q1-9(medication) answered other than YES but gave age at first use.
Change to YES.
68. (deleted)
69. DRUGTYPE=99 Recode as 3
digits.DRUGTYPE or DRUGNAME
missing Change to unknown.
70. New cancer (R1) answered N, but details given in R1,R2.
Recode R1 to Yes.
71. Details given in R2 but not in R1.
Move data from R2 to R1.
72. Missing information in R1, R2
Recontact patient.
73. (deleted)
74. T1(read LTFU study newsletter) answered NO but did not skip to the next page.
Change to YES.
75. T3(seek more information about health problems) answered NO but did not skip to T5.
Change to YES.
76. S4 (Other resources for information) answered NO, but gave details.
Change S4 to YES.
77. Gaps in sequence - Other resources for information (S4).
Check questionnaire. If nothing has been missed, shift entries to the left to eliminate gaps.

Protocol for resolving SAS dataset errors and inconsistencies – follow-up 4 questionnaire

1. Unfeasible height.
Check responses. Change to ambiguous if unable to correct.
2. Unfeasible weight.
Check responses. Change to ambiguous if unable to correct.
3. Lower level of education at FU2007 (levelschool) than at baseline (educatn), FU1 (ednow) or FU@ (f2educatn).
Check response - if data entered matches response on questionnaire form, change to ambiguous.
4. Not currently employed (employmentstatusfulltime or ...parttime) and job title given.
Delete job title, unless all employmentstatus questions are "no" or missing, in which case change everything to missing.
5. Personal income exceeds household income.
Check responses. If correct, change both to ambiguous.
6. Saw no health care provider (medcaretalknone) and other health care questions answered.
If responses indicate that patient did see a provider, change medcaretalknone to "no". Otherwise, change medcaretalkphysician - medcaretalkother to "no" and subsequent responses to missing.
7. Did not see a physician according to B1 (medcaretalkphysician), but did see one according to B3 (seephysician), or the reverse.
Check responses. If B4 (seephysician2) also indicates a physician was seen, change medcaretalkphysician to "yes". Otherwise, change medcaretalkphysician and seephysician to ambiguous.
8. More cancer physician visits than total physician visits.
Change both to ambiguous.
9. Did not see a physician according to B1 (medcaretalkphysician) or B3 (seephysician), but had a check-up by a physician (checkup).
If both medcaretalkphysician and seephysician agree, change checkup to missing. Otherwise change medcaretalkphysician to "yes", and seephysician to missing if "none".
10. Never had a check-up (checkup) and outcomes of check-up listed.
If all outcomes are "no" or missing, change to missing. Otherwise change checkup to ambiguous.
11. Breast exam or pap-smear questions answered by male.
Check that the questionnaire is for the right CCSSID. If so, change to missing.
12. Did not take drug (bcpill_yn, etc.), but age is given, and/or the conditional question "If yes, are you *currently* taking the drug" was answered.
If age is missing and currently taking is "no", change currently taking to missing. If both age and currently taking are answered, or a drug name is given, change (drug)_yn to "yes". Otherwise, change age or currently taking to missing.
13. Sections D, E, F, G, H, I, J and K: Medical condition not answered "yes" and age at first occurrence given.
Change both to ambiguous.
14. Sections D, E, F, G, H, I, J and K: Age at first occurrence greater than current age.
Change to ambiguous.
15. Not blind in (only) one eye (blind1) and "if yes, do you have any sight in this eye" question (blind2) answered.
If blind2 is "no", change blind1 to "yes, condition still present". Otherwise, change blind2 to missing.
16. Not blind in both eyes (blind3) and "if yes, do you have any sight" question (blind4) answered.
If blind4 is "no", change blind3 to "yes, condition still present". Otherwise, change blind4 to missing.
17. Menstrual history questions answered by males.
Check that CCSSID is correct. If so, change to missing.
18. No menstrual periods (menstrualnat) but age given (menstrualnatage, menstrualageyr, menstrualagemo).

If menstrualnat missing, change it to "yes". Otherwise change all four to ambiguous.

19. Menstrual age months greater than 11.

Change both years and months to ambiguous.

20. Male survey question answered by females.

Check that CCSSID is correct. If so, change to missing.

21. No hypertension (hypertension2) but hypertension medication (hypertension3) answered.

If hypertension3 is "no", change to missing. Otherwise change hypertension2 to ambiguous.

22. No high cholesterol (hicholmed1) but cholesterol medication (hicholmed2) answered.

If hicholmed2 is "no", change to missing. Otherwise change hicholmed1 to ambiguous.

23. No hepatitis (hepatitis) but hepatitis details answered.

Change hepatitis to ambiguous.

24. Mastectomy not answered "yes" but number of breasts removed answered.

If mastectomy is missing or "not sure", change to "yes". Otherwise change both to ambiguous.

25. No current memory problems (memory1) but problem severity (memory2) answered.

If memory1 missing, change to "yes and still present". Otherwise change both to ambiguous.

26. No epilepsy or seizures (epilepsyseizures1) but medication (epilepsyseizures2) answered.

If epilepsyseizures2 is "no", change to missing. Otherwise change epilepsyseizures1 to ambiguous.

27. No current equilibrium problems (equilibrium1) but problem severity (equilibrium2) answered.

If equilibrium1 missing, change to "yes and still present". Otherwise change both to ambiguous.

28. Stroke1 not answered "Yes" and subsequent stroke questions answered.

If subsequent questions suggest patient had a stroke, change stroke1 to "yes", otherwise change stroke2-9 to missing.

29. Bodily pain answered "none" and subsequent pain questions answered.

If subsequent questions suggest patient had pain, change bodilypain to missing, otherwise change subsequent questions to missing.

30. Currently living alone (Currentlivingalone) and also living with other people.

Change all to ambiguous.

31. Currently "never married" and number of times married is answered.

Check responses, including those on earlier questionnaires. If never married, change nummarried to missing. Otherwise change both to ambiguous.

32. Never drink (everdrink), but drinking details given.

If drinking details are consistent with never having drunk, change them to missing. If they are consistent with having drunk, change everdrink to "yes". If unsure, change everything to ambiguous.

33. Never smoked (eversmoke), but smoking details given.

If smoking details are consistent with never having smoked, change them to missing. If they are consistent with having smoked, change eversmoke to "yes". If unsure, change everything to ambiguous.

34. No drinks in last year (drinksfreq1) and drink details given.

If drinking details are consistent with no drinks in last year, change them to missing. If they are consistent with having drinks in last year, change drinksfreq1 to missing. If unsure, change everything to ambiguous.

35. Drink heavily (drinksfreq2) more often than ever drink (drinksfreq1).

Check responses. If correct, change both to ambiguous.

36. Not a current smoker (smokenow) and quit attempts answered.

If smokenow is missing, change to "yes". If quitsmok12mo is zero, change it to missing. Otherwise, change both to ambiguous.

37. Smoking start age or (smoking start age + years smoked) greater than current age.

- Check responses. If correct, change to ambiguous.
38. Never used a tobacco product, and years of use given.
- If “<1”, change to missing. Otherwise change both to ambiguous.
39. No vigorous activity (vigactivity) and frequency (vigactivitydays) or duration (vigactivitytime) given.
- Check responses. If days and time are zero, change to missing. Otherwise, change all to ambiguous.
40. No moderate activity (modactivity) and frequency (modactivitydays) or duration (modactivitytime) given.
- Check responses. If days and time are zero, change to missing. Otherwise, change all to ambiguous.
41. Second cancer (diagnosissince) not answered "yes" and details given.
- Check full questionnaire data and existing second neoplasm data. If we already know about this condition, delete details. Otherwise, change diagnosissince to “yes”.
42. Year of second cancer same as or before year of last CCSS questionnaire (d_lcontact).
- Check full questionnaire data and existing second neoplasm data. If we already know about this condition, delete details. If we don't already know about it, keep the data. Not pregnant since last questionnaire (pregnancysince) and pregnancy details listed.
- Check dates and data from other questionnaires. If we already know about the pregnancies, change to missing. Otherwise, change pregnancysince to “yes”.
43. Not pregnant since last questionnaire (pregnancysince) and offspring details listed.
- Check dates and data from other questionnaires. If we already know about the offspring, change to missing. Otherwise, change pregnancysince to “yes”.
44. No additional radiation (addrad) but month or year given.
- Change to “yes”.
45. No additional chemo (addrad) but month or year given.
- Change to “yes”.
46. Not a CCSS participant.
- If outcome is missing, the CCSSID is not an eligible case and is probably incorrect. If outcome is not missing and not “participant”, check to see if there has been a status change since November 2008. If not, the patient is not a participant: either the ID is incorrect or the data need to be deleted.
47. Partial new-cancer, radiation or treatment date (missing month or year).
- Check questionnaire. Set computed date to missing (if year unavailable) or 7/1/YYYY (if month unavailable).
48. Partially-competed questionnaire date.
- Edit by hand If day missing, use MM/15/YYYY. If month missing, use 07/01/YYYY.
49. Drug name given and drug question not answered “yes”.
- Change “yn” variable to “not sure”.

Protocol for resolving SAS dataset errors and inconsistencies – expansion cohort baseline

Section A

1. Date of questionnaire completion (front page) missing.
Check St Jude data. Use mailing date if necessary.
2. Completion date precedes date of birth or later than today's date.
3. Check St Jude data for both dates. Use mailing date as date of completion if necessary.
4. Height feet or inches missing.
5. Amend if an obvious misread of the question (e.g. 5 feet 64 inches). Otherwise change to missing.
6. Unfeasible height (≥ 8 feet or ≤ 2 feet).
7. Amend if an obvious misread of the question. Otherwise change to missing.
8. Unfeasible weight (≥ 600 lb or ≤ 20 lb).
Change to missing.
9. Unfeasible height/weight combination (>3 IQR away from $\log(\text{weight})$ vs height regression line).
Change one or both to missing.
10. Twin (A6) but not multiple birth.
11. Twin (A6) but no siblings (A8).
12. Not twin (A6) but multiple birth question answered
13. Check number of siblings (A8) and sibling dates of birth in section P. If no sibs, delete twin status.
14. Live with parents in A9 but answered currently lives alone in M1.
Change both to missing.

Section B

1. Answered "no health care" but gave other answers in B1-B6
2. Change "no health care" to "no" if b1-b6 includes a "yes" answer, otherwise delete other answers.
3. Other provider not answered "yes" and details given, or gaps in details variables.
Change "other provider" to "yes" unless the provider is one of the earlier categories in B1.
4. Mismatch between saw a physician and number of physician visits.
Change "saw a physician" to "yes" unless number is 0.
5. Cancer physician visits greater than total physician visits.
Change both to missing.
6. Other scarring/disfigurement not answered "yes" and details given, or gaps in details variables.
Change to "yes" if appropriate. Need to look at codes.
7. Drug yes/no not answered "yes" and details given.
Change to "yes".
8. Drug age greater than age at baseline. Change to missing if age at baseline is correct.
9. Birth control Pill, female hormone answered by males and male hormone answered by females.
Delete drug data.
10. Answered yes to B8.4 (Pills or Insulin for diabetes) but no to E6 and E7 ever been told that have diabetes that can be controlled by pills or insulin.
11. Delete drug data.

Section C

1. Condition not answered Yes and age and/or details given.
2. Change to "yes, not sure if condition is present" unless the age is 0 (DATSTAT default).
3. Age of condition greater than age at baseline.
Change to missing if age at baseline is correct.
4. Gaps in sequence: other hearing problems.
5. Gaps in sequence: retina problems.
6. Gaps in sequence: other eye problems.

7. Gaps in sequence: other speech problems.
Move data to the left to close gaps.

Section D

1. Condition not answered Yes and age and/or details given.
2. Change to “yes, not sure if condition is present” unless the age is 0 (DATSTAT default).
3. Age of condition greater than age at baseline
Change to missing if age at baseline is OK.
4. Gaps in sequence: other kidney/urinary problems.
Move data to the left to close gaps.

Section E

1. Condition not answered Yes and age and/or details given.
2. Change to “yes, not sure if condition is present” unless the age is 0 (DATSTAT default).
3. Age of condition greater than age at baseline
Change to missing if age at baseline is OK..
4. Gaps in sequence: Broken Bones.
5. Gaps in sequence: Other Hormonal Problems.
Move data to the left to close gaps.
6. Menstrual Periods (E13, E14, E15, E16) answered by males.
Delete unless any answers aren’t “no”, in which case check that we have the right sex.
7. Answered No to E13 and didn't skip to E15.
8. Look at the whole menstrual history record – change answers to give best fit. If unsure, change everything to missing.
9. Answered a, b, c, or d to E15 but didn't skip to F1.
Change E16 to missing.
10. Age of first natural period greater than age at baseline.
Delete if age at baseline is OK.
11. Answered currently pregnant (E15.d) and no on N1 (ever had intercourse) or no on N4 (currently pregnant).
Check section N for pregnancy data. Change E15 or N1-N4 depending on which gives consistent data. Change all to missing if unsure.
12. Menstrual age months greater than 11.
13. Change age to missing unless it looks like they have given us the same age in years and months (e.g. 25 years 304 months)

Section F

1. Condition not answered Yes and age and/or details given.
2. Change to “yes, not sure if condition is present” unless the age is 0 (DATSTAT default).
3. Age of condition greater than age at baseline
Change to missing if age at baseline is OK..
4. Gaps in sequence: other Coronary Problems.
5. Gaps in sequence: other Heart/circulatory Problems.
Move data to the left to close gaps.

Section G

1. Condition not answered Yes and age and/or details given.
2. Change to “yes, not sure if condition is present” unless the age is 0 (DATSTAT default).
3. Age of condition greater than age at baseline.
Change to missing if age at baseline is OK.
4. Gaps in sequence: other Breathing or Lung Problems.
Move data to the left to close gaps.

Section H

1. Condition not answered Yes and age and/or details given.
2. Change to “yes, not sure if condition is present” unless the age is 0 (DATSTAT default).
3. Age of condition greater than age at baseline.
Change to missing if age at baseline is OK.

Gaps in sequence: other Liver Problems. Move data to the left to close gaps.

Section I

1. Condition not answered Yes and age and/or details given.
2. Change to “yes, not sure if condition is present” unless the age is 0 (DATSTAT default).
3. Age of condition greater than age at baseline.
Change to missing if age at baseline is OK.
4. Gaps in sequence: Amputation.
5. Gaps in sequence: Spine Surgery.
6. Gaps in sequence: Joint Replacement.
7. Gaps in sequence: Bone Surgery.
8. Gaps in sequence: Heart Surgery.
9. Gaps in sequence: Lung Surgery.
10. Gaps in sequence: Organ Transplant.
11. Gaps in sequence: Any Other Surgery.
Move data to the left to close gaps.
12. Males answered I32, I33 or I34.
13. Females answered I35 or I36.
14. Delete unless any answers aren’t “no”, in which case check that we have the right sex.

Section J

1. Condition not answered Yes and age and/or details given.
2. Change to “yes, not sure if condition is present” unless the age is 0 (DATSTAT default).
3. Age of condition greater than age at baseline.
Change to missing if age at baseline is OK.
4. Gaps in sequence: Epilepsy/seizures.
5. Gaps in sequence: Paralysis.
6. Gaps in sequence: Other Brain/Nervous system Problems.
Move data to the left to close gaps.

Section K

1. Answered 'none' to K21 and didn't skip to section L.
Change to missing, unless subsequent answers are consistent with “none”, in which case delete them.

Section L

1. Answered no to L1 and didn't skip to section M.
2. Change L1 to “yes” if details given, unless details are clearly the first cancer diagnosis. If unsure leave in – these will be reviewed later when we get path reports.
3. Answered no to L6 and didn't skip to section M.
4. Change L6 to “yes” if details given, unless details are clearly the first cancer diagnosis or the same as L1. If unsure leave in. These will be reviewed later when we get path reports.
5. Answered no to L3 or L8 (no treatments) but gave type of treatment.

Change L3 and/or L8 to “yes”.

6. Inconsistencies in dates. Change to missing as required.

Section M

1. Ever married according to marital status, but not according to "ever married".
Change “ever married” to “yes”. May need to look at all of section M if unsure.
2. Answered no to M2 and did not skip to section N.
3. Change M2 to yes or delete rest of M if consistent with M2 being “no”.
4. Answered Yes to M7 and didn't skip to section N.
5. Change M7 to no or delete rest of M if consistent with M7 being “yes”.
6. Answered Yes to M12 and didn't skip to section N.
7. Change M12 to no or delete rest of M if consistent with M12 being “yes”.
- a. Answered no to M12 or M7 (currently in relationship) but answered currently divorced, widowed or separated for M3, and vice-versa.
8. May need to change all to missing – look at whole of section M.
9. Inconsistencies between dates of start and end of relationships and between 1st and current relationship.
Change to missing.
10. Inconsistencies in dates.
Change to missing if dates of birth and/or completion are OK.

Section N

1. Answered no to N1 (ever had sexual intercourse) and didn't skip to O1.
Recode N1, or delete rest of N if consistent with N1 being “no”.
2. Answered no to N6 (ever pregnant) and did not skip to section O.
Recode N6, or delete rest of N if consistent with N6 being “no”.
3. Answered yes to N6 and N4 (currently/ever pregnant) but zero to how many times (N7).
Check pregnancy dataset (N8). If no pregnancy details, change N6/N4. Otherwise, edit N7.
4. Number of pregnancies (N7) does not match number of pregnancies listed in N8.
Change N7 if number in N8 is larger.
5. Pregnancy details given in N8, never sexually active or pregnant.
6. Change ever pregnant/ever sexually active if rest of N is consistent with pregnancies having occurred.
7. Missing pregnancy outcome.
8. Code most recent pregnancy to “currently pregnant”.
9. Pregnancy age greater than current age. Change to missing if no problems with current age.
10. Pregnancy age less than 12.
11. Review related dates and ages. Change to missing if necessary.
12. Live birth or stillbirth, and duration less than 20 weeks.
13. If live birth, change duration to missing. If stillbirth, change outcome to miscarriage.
14. Pregnancy duration more than 42 weeks.
Review related dates and ages. Change to missing if necessary.

Section O

1. Answered no to O1 but didn't skip to O7.
2. Change O1 to “yes” if remaining questions indicate smoking, otherwise delete O2-6.
3. Number of years smoked and age started smoking, drinking, first mammogram > age at baseline.
Change to missing if age at baseline is OK.
4. Answered no to O3 (do you smoke now?) but answered O6.
Change both to missing.
5. Never chew, snuff, smoked pipe or cigars but length of use answered.
6. Change both to missing unless length of use is <1, in which case change length of use only to missing.
7. Answered no to O9 (never had at least 2 alcoholic drinks) but did not skip to O15.
Change O9 to “yes” if remaining questions indicate drinking, otherwise delete O10-14.

8. Answered zero to O12 but did not skip to O15. Change O12 to missing.
9. Males answered questions O27 - O30.
10. Delete unless any answers aren't "no", in which case check that we have the right sex.
11. Never had a mammogram but age given (O30).
12. Change to "yes" if age is >10, otherwise change both to missing.
13. Age of first mammogram less than 10 or greater than current age.
Change to missing age at baseline is OK.

Section P

1. Number of siblings in section A (n_sibls) does not match number of siblings listed in section P (numsiblings).
2. If there are more siblings listed in section P, change n_sibls.
3. Difference of more than 30 years between case DOB and sibling DOB. Edit dates if obvious typos. Otherwise change to missing.
4. Sibling age less than 0.
Delete age and date of birth.

Section Q

1. Female patient with Klinefelter's syndrome (Q1a-f) or Wiskott-Aldrich syndrome (Q1a-n), or male with Turner's syndrome (Q1a-l).
2. Change Q1a to "no" and delete condition.
3. Mother, sister, daughter or wife with Klinefelter's syndrome or Wiskott-Aldrich syndrome; father, brother son or husband with Turner's syndrome (Q1b).
4. Delete condition.
5. Deafness or impaired hearing at birth (Q3a-d) but none of C1-C7 answered "yes".
Change Q3a-d to missing.
6. Blindness or difficulty seeing at birth (Q3a-e) but none of C8-C18 answered "yes".
Change Q3a-e to missing.
7. Hole in the heart or other heart defect (Q3a-l) but none of F1-F13 answered "yes".
Change Q3a-l to missing.
8. Any congenital abnormality of the pancreas, liver or digestive tract (Q3a-m) but none of H3-H9 answered "yes".
Change Q3a-m to missing.
9. Female patient with undescended testes (Q3a-o).
Change Q3a-o to missing.
10. Mother, sister, daughter or wife with undescended testes (Q3b).
Delete condition.
11. Genetic conditions (Q1b), birth conditions (Q3b) or cancer (Q4) in sibling, but number of siblings (A8) is 0 and no siblings listed in P1.
Delete conditions.

Section R

1. Answered no to R3 learning disabled/special program but completed grades (R4) and reasons (R3).
Change to "yes".
2. Didn't graduate from high school but got a high school diploma. Change to "completed high school".

Section S

1. Answered no to S1 (never had a job) and did not skip to S4.
2. If S2 and S3 are consistent with having had a job, change S1 to yes. Otherwise delete S2 and S3.
3. Answered yes to S2 - Not currently working and did not skip to S4.
Change both to missing.
4. Rejected from Military, Police and/or Fire dept, but did not apply (S4 and S5).
Change both to missing.

Section T

1. Personal income larger than household income. Change both to missing.

Section U

1. Answered no or Canadian Resident to U2 and did not skip to U4.
2. If U3 indicates respondent has health insurance, change U2 to Yes. Might need to check U3a and U3b answers.
3. Answered yes to U2 but answered no to all of U3.
Change U3 to missing.

Protocol for resolving SAS dataset errors and inconsistencies – follow-up 5 questionnaire

Section A

1. Date of questionnaire completion (front page) missing.
Check St Jude data. Use mailing date if necessary.
2. Completion date precedes 2014 or later than today's date.
Check St Jude data for both dates. Use mailing date as date of completion if necessary.
3. Height feet or inches missing.
Amend if an obvious misread of the question (e.g. 5 feet 64 inches). Otherwise change to missing.
4. Unfeasible height (≥ 8 feet or ≤ 2 feet).
Amend if an obvious misread of the question. Otherwise change to missing.
5. Unfeasible weight (≥ 600 lb or ≤ 20 lb).
Change to missing.
6. Unfeasible height/weight combination (>3 IQR away from $\log(\text{weight})$ vs height regression line).
Change one or both to missing.

Section B

1. Answered "no health care" but gave other answers in B1-B6
Change "no health care" to "no" if b1-b6 includes a "yes" answer, otherwise delete other answers.
2. Other provider not answered "yes" and details given, or gaps in details variables.
Change "other provider" to "yes" unless the provider is one of the earlier categories in B1.
3. Mismatch between saw a physician and number of physician visits.
Change "saw a physician" to "yes" unless number is 0.

Section C

1. Drug yes/no not answered "yes" and details given.
Change to "yes".
2. Drug age greater than age at baseline.
Change to missing if age at baseline is correct.
3. Birth control Pill, female hormone answered by males and male hormone answered by females.
Delete drug data.
4. Answered yes to B8.4 (Pills or Insulin for diabetes) but no to E6 and E7 ever been told that have diabetes that can be controlled by pills or insulin.
Delete drug data.

Section D

1. Condition not answered Yes and age and/or details given.
Change to "yes, not sure if condition is present" unless the age is 0 (DATSTAT default).
2. Age of condition greater than age at baseline.
Change to missing if age at baseline is correct.

3. Gaps in sequence: other hearing problems.
 4. Gaps in sequence: retina problems.
 5. Gaps in sequence: other eye problems.
 6. Gaps in sequence: other speech problems.
- Move data to the left to close gaps.

Section E

1. Condition not answered Yes and age and/or details given.
Change to “yes, not sure if condition is present” unless the age is 0 (DATSTAT default).
2. Age of condition greater than age at baseline
Change to missing if age at baseline is OK.
3. Gaps in sequence: other kidney/urinary problems.
Move data to the left to close gaps.

Section F

1. Condition not answered Yes and age and/or details given.
Change to “yes, not sure if condition is present” unless the age is 0 (DATSTAT default).
2. Age of condition greater than age at baseline
Change to missing if age at baseline is OK..
3. Gaps in sequence: other Coronary Problems.
4. Gaps in sequence: other Heart/circulatory Problems.
Move data to the left to close gaps.

Section G

1. Condition not answered Yes and age and/or details given.
Change to “yes, not sure if condition is present” unless the age is 0 (DATSTAT default).
2. Age of condition greater than age at baseline
Change to missing if age at baseline is OK..
3. Gaps in sequence: Broken Bones.
4. Gaps in sequence: Other Hormonal Problems.
Move data to the left to close gaps.
5. Menstrual Periods (G13, G14, G15, G16) answered by males.
Delete unless any answers aren’t “no”, in which case check that we have the right sex.
6. Answered No to G13 and didn't skip to G15.
Look at the whole menstrual history record – change answers to give best fit. If unsure, change everything to missing.
7. Answered a, b, c, or d to G15 but didn't skip to H1.
Change G16 to missing.
8. Age of first natural period greater than age at baseline.
Delete if age at baseline is OK.
9. Menstrual age months greater than 11.
Change age to missing unless it looks like they have given us the same age in years and months (e.g. 25 years 304 months)

Section H

1. Condition not answered Yes and age and/or details given.
Change to “yes, not sure if condition is present” unless the age is 0 (DATSTAT default).
2. Age of condition greater than age at baseline.
Change to missing if age at baseline is OK.
3. Gaps in sequence: other Breathing or Lung Problems.
Move data to the left to close gaps.

Section I

1. Condition not answered Yes and age and/or details given.
Change to “yes, not sure if condition is present” unless the age is 0 (DATSTAT default).
2. Age of condition greater than age at baseline.
Change to missing if age at baseline is OK.
3. Gaps in sequence: other Liver Problems.
4. Gaps in sequence: other digestive problems.
Move data to the left to close gaps.

Section J

1. Condition not answered Yes and age and/or details given.
Change to “yes, not sure if condition is present” unless the age is 0 (DATSTAT default).
2. Age of condition greater than age at baseline.
Change to missing if age at baseline is OK.
3. Gaps in sequence: Amputation.
4. Gaps in sequence: Spine Surgery.
5. Gaps in sequence: Joint Replacement.
6. Gaps in sequence: Bone Surgery.
7. Gaps in sequence: Heart Surgery.
8. Gaps in sequence: Lung Surgery.
9. Gaps in sequence: Organ Transplant.
10. Gaps in sequence: Any Other Surgery.
Move data to the left to close gaps.
11. Males answered J32, J33 or J34.
12. Females answered J35 or J36.
Delete unless any answers aren’t “no”, in which case check that we have the right sex.

Section K

1. Condition not answered Yes and age and/or details given.
Change to “yes, not sure if condition is present” unless the age is 0 (DATSTAT default).
2. Age of condition greater than age at baseline.
Change to missing if age at baseline is OK.
3. Gaps in sequence: Epilepsy/seizures.
4. Gaps in sequence: Severe headaches
5. Gaps in sequence: Paralysis.

6. Gaps in sequence: Stroke/paralysis
7. Gaps in sequence: Other Brain/Nervous system Problems.
Move data to the left to close gaps.

Section M

1. Never married but gave number of times married.
Change “ever married” to “yes”. May need to look at earlier marriage data if unsure.

Section N

1. Answered no to N1 but didn't skip to N7.
Change N1 to “yes” if remaining questions indicate smoking, otherwise delete N2-6.
2. Number of years smoked and age started smoking, drinking, first mammogram > age at baseline.
Change to missing if age at baseline is OK.
3. Answered no to N3 (do you smoke now?) but answered N6.
Change both to missing.
4. Never chew, snuff, smoked pipe or cigars but length of use answered.
Change both to missing unless length of use is <1, in which case change length of use only to missing.
5. Answered no to N9 (never had at least 2 alcoholic drinks) but did not skip to N15.
Change N9 to “yes” if remaining questions indicate drinking, otherwise delete N10-14.
6. Answered zero to N12 but did not skip to N15.
Change N12 to missing.
7. Did not answer yes to physical activity, but gave details.
Change to yes, unless details suggest no physical activity, in which case delete the details.
8. More than 7 days per week.
Change to missing.

Section T

1. No RT or chemo, but details given.
Change to “yes”.
2. RT/chemo before last contact or after FU5 date.
3. Month not in range 1-12.
Change date to missing.

Section U

1. No hospitalizations, but details given.
Change to “yes”.
2. Hospitalization not in 12 months before questionnaire completion.
Delete hospitalization data.
3. Month not in range 1-12.
Change date to missing.

Section V

1. V1-4 skip patterns.

Make consistent with pregnancy details in V5, change to missing if unsure.

2. Reported pregnancies, but not pregnant per V1 or V3.

If ages are consistent with this being a recent pregnancy, or if unsure, change V1/V3 to “yes”. Otherwise, delete pregnancy data.

3. Pregnancy age greater than current age.

Change to missing if no problems with current age.

4. Pregnancy age less than age at last contact.

Check to see if we already have details of this pregnancy from earlier questionnaires. If so, delete the new data.

5. Missing pregnancy outcome.

Code most recent pregnancy to “currently pregnant” if ages are consistent with this.

6. Pregnancy age less than 12.

Review related dates and ages. Change to missing if necessary.

7. Live birth or stillbirth, and duration less than 20 weeks.

If live birth, change duration to missing. If stillbirth, change outcome to miscarriage.

8. Pregnancy duration more than 43 weeks.

Review related dates and ages. Change to missing if necessary.

Protocol for resolving SAS dataset errors and inconsistencies – follow-up 6 questionnaire

Short version

Section A

1. Date of questionnaire completion (front page) missing.
Check St Jude data. Use mailing date if necessary.
2. Completion date precedes 2016 or later than today's date.
Check St Jude data for both dates. Use mailing date as date of completion if necessary.

Medium version

Section A

1. Date of questionnaire completion (front page) missing.
Check St Jude data. Use mailing date if necessary.
2. Completion date precedes 2016 or later than today's date.
Check St Jude data for both dates. Use mailing date as date of completion if necessary.

Section C

1. No health insurance in C1, but C2-C4 answered.
If answers in C2-C4 are consistent with having insurance, change C1. Otherwise, delete C2-C4.
2. No problems paying bills (C5) but unable to pay bills in C6.
Change to missing.
3. Not working in C13, but work details in C14.
If answers in C14 are consistent with having insurance, change C13. Otherwise, delete C14.
4. No home ownership in C22, but details given in C23-C26.
If answers in C23-C26 are consistent with home ownership, change C22. Otherwise, delete C23-C26.
5. No debts in C27, but details given in C28.
If answer in C28 is consistent with debt, home ownership, change C27. Otherwise, delete C28.
6. No bankruptcy in C30, but a year given in C31.
Change both to missing.
7. Numbers entered in section C out of range.
Change to missing.

Long version

Section A

1. Date of questionnaire completion (front page) missing.
Check St Jude data. Use mailing date if necessary.
2. Completion date precedes 2016 or later than today's date.
Check St Jude data for both dates. Use mailing date as date of completion if necessary.

3. Height feet or inches missing.
Amend if an obvious misread of the question (e.g. 5 feet 64 inches). Otherwise change to missing.
4. Unfeasible height (≥ 8 feet or ≤ 2 feet).
Amend if an obvious misread of the question. Otherwise change to missing.
5. Unfeasible weight (≥ 600 lb or ≤ 20 lb).
Change to missing.
6. Unfeasible height/weight combination (>3 IQR away from $\log(\text{weight})$ vs height regression line).
Change one or both to missing.

Section B

1. Answered "no health care" but gave other answers in B1-B6
Change "no health care" to "no" if b1-b6 includes a "yes" answer, otherwise delete other answers.
2. Other provider not answered "yes" and details given, or gaps in details variables.
Change "other provider" to "yes" unless the provider is one of the earlier categories in B1.
3. Mismatch between saw a physician and number of physician visits.
Change "saw a physician" to "yes" unless number is 0.

Section C

1. Answered "test" questions for wrong sex.
Change to missing.

Section D

1. Did not answer "yes" to physical activity questions, but gave details.
If answers are consistent with physical activity, change to "yes". Otherwise, delete details.
2. More than 7 days per week.
Change to missing.