

**CODEBOOK FOR CSSCD PATIENT ROSTER**  
 CSSCD INFANT COHORT PATIENTS

CONTENTS OF SAS DATASET: ROST2N3P.SD2  
 DATA FROM CSSCD PHASE 2/3 ROSTER  
 VARIABLES ARE LISTED IN ALPHABETICAL ORDER AND IN ORDER OF THEIR POSITION  
 IN THE SAS DATASET 09:55 Friday, June 1, 2001  
 DATE VARIABLES HAVE BEEN REMOVED & CSSCD ID #S REPLACED W/ ANONYMIZED ID #

CONTENTS PROCEDURE

|   |                       |     |
|---|-----------------------|-----|
| Data Set Name: OUT.ROST2N3P               | Observations:         | 675 |
| Member Type: DATA                         | Variables:            | 22  |
| Engine: V612                              | Indexes:              | 0   |
| Created: 12:35 Friday, June 1, 2001       | Observation Length:   | 176 |
| Last Modified: 12:35 Friday, June 1, 2001 | Deleted Observations: | 0   |
| Protection:                               | Compressed:           | NO  |
| Data Set Type:                            | Sorted:               | YES |
| Label:                                    |                       |     |

-----Engine/Host Dependent Information-----

|                           |      |
|---------------------------|------|
| Data Set Page Size:       | 8192 |
| Number of Data Set Pages: | 16   |
| File Format:              | 607  |
| First Data Page:          | 1    |
| Max Obs per Page:         | 46   |
| Obs in First Data Page:   | 29   |

-----Alphabetic List of Variables and Attributes-----

| #  | Variable | Type | Len | Pos | Format | Informat | Label                               |
|----|----------|------|-----|-----|--------|----------|-------------------------------------|
| 2  | AGE_D0E  | Num  | 8   | 8   |        |          | AGE AT DATE OF PHASE 1 ENTRY (DAYS) |
| 13 | AGE_D0E2 | Num  | 8   | 152 |        |          | AGE AT DATE OF PHASE 2 ENTRY (YRS)  |
| 14 | AGE_D0E3 | Num  | 8   | 160 |        |          | AGE AT DATE OF PHASE 3 ENTRY (YRS)  |
| 15 | AGE_E0F  | Num  | 8   | 168 |        |          | AGE AT END OF FOLLOW-UP (YRS)       |
| 1  | ANONID   | Char | 8   | 0   |        |          | ANONYMIZED ID #                     |
| 10 | DIED     | Num  | 8   | 80  | 2.     | 2.       | DIED?                               |
| 5  | DXSOURCE | Num  | 8   | 32  |        |          | SOURCE OF HEMOGLOBIN DIAGNOSIS      |
| 6  | HEMO     | Num  | 8   | 40  | 2.     | 2.       | CDC HEMOGLOBIN DIAGNOSIS            |
| 11 | HEMO2    | Num  | 8   | 88  |        |          | LOCAL HEMOGLOBIN DIAGNOSIS          |
| 3  | PHASE2   | Num  | 8   | 16  |        |          | ENROLLED IN PHASE 2                 |
| 4  | PHASE3   | Num  | 8   | 24  |        |          | ENROLLED IN PHASE 3                 |
| 9  | PTYPE    | Num  | 8   | 64  | 2.     | 2.       | PROTOCOL TYPE DURING PHASE 1        |
| 8  | RACE     | Num  | 8   | 56  | 2.     | 2.       | RACE                                |
| 12 | SETNO    | Num  | 8   | 144 |        |          | # OF SERA SAMPLE DATES              |
| 7  | SEX      | Num  | 8   | 48  | 2.     | 2.       | GENDER                              |

**CODEBOOK FOR CSSCD PATIENT ROSTER**  
CSSCD INFANT COHORT PATIENTS

---

---

CONTENTS OF SAS DATASET: ROST2N3P.SD2  
DATA FROM CSSCD PHASE 2/3 ROSTER  
VARIABLES ARE LISTED IN ALPHABETICAL ORDER AND IN ORDER OF THEIR POSITION  
IN THE SAS DATASET 09:55 Friday, June 1, 2001  
DATE VARIABLES HAVE BEEN REMOVED & CSSCD ID #S REPLACED W/ ANONYMIZED ID #

CONTENTS PROCEDURE

-----Variables Ordered by Position-----

| #  | Variable | Type | Len | Pos | Format | Informat | Label                               |
|----|----------|------|-----|-----|--------|----------|-------------------------------------|
| 1  | ANONID   | Char | 8   | 0   |        |          | ANONYMIZED ID #                     |
| 2  | AGE_DOE  | Num  | 8   | 8   |        |          | AGE AT DATE OF PHASE 1 ENTRY (DAYS) |
| 3  | PHASE2   | Num  | 8   | 16  |        |          | ENROLLED IN PHASE 2                 |
| 4  | PHASE3   | Num  | 8   | 24  |        |          | ENROLLED IN PHASE 3                 |
| 5  | DXSOURCE | Num  | 8   | 32  |        |          | SOURCE OF HEMOGLOBIN DIAGNOSIS      |
| 6  | HEMO     | Num  | 8   | 40  | 2.     | 2.       | CDC HEMOGLOBIN DIAGNOSIS            |
| 7  | SEX      | Num  | 8   | 48  | 2.     | 2.       | GENDER                              |
| 8  | RACE     | Num  | 8   | 56  | 2.     | 2.       | RACE                                |
| 9  | PTYPE    | Num  | 8   | 64  | 2.     | 2.       | PROTOCOL TYPE DURING PHASE 1        |
| 10 | DIED     | Num  | 8   | 80  | 2.     | 2.       | DIED?                               |
| 11 | HEMO2    | Num  | 8   | 88  |        |          | LOCAL HEMOGLOBIN DIAGNOSIS          |
| 12 | SETNO    | Num  | 8   | 144 |        |          | # OF SERA SAMPLE DATES              |
| 13 | AGE_DOE2 | Num  | 8   | 152 |        |          | AGE AT DATE OF PHASE 2 ENTRY (YRS)  |
| 14 | AGE_DOE3 | Num  | 8   | 160 |        |          | AGE AT DATE OF PHASE 3 ENTRY (YRS)  |
| 15 | AGE_EOF  | Num  | 8   | 168 |        |          | AGE AT END OF FOLLOW-UP (YRS)       |

-----Sort Information-----

Sortedby: ANONID  
Validated: YES  
Character Set: ANSI

**CODEBOOK FOR CSSCD PATIENT ROSTER**  
CSSCD INFANT COHORT PATIENTS

---

---

\*\*\*\*\*  
\* RST.FMT contains value labels for numerical codes assigned to categorical \*  
\* variables in the SAS dataset ROST2N3P.SD2 \*  
\*\*\*\*\*;

PROC FORMAT;

VALUE HEMO 1='SS'  
          2='SC'  
          3='SB+ THAL'  
          4='SBO THAL'  
          5='SS ALPHA THAL'  
          6='SBO + ALPHA THAL'  
          7='SBO + DELTA THAL'  
          8='OTHER VARIANT'  
          9='TRANSFUSED';

VALUE RACE 1='BLACK'  
          2='OTHER';

VALUE SEX 1='FEMALE'  
          2='MALE';

VALUE PTYPE 1='FULL PROTOCOL'  
          2='SEMI-PROTOCOL';

VALUE DXSOURCE 1='CDC' 2='FORM 14' 3='CLINIC EDIT';

\* Format NO\_YES used for the following variables: PHASE2 PHASE3 HASTUBE  
DIED;

VALUE NO\_YES 0='NO'  
          1='YES';

\* NOTE: VARIABLE DIED HAS A VALUE ONLY IF PATIENT DIED - I.E., MISSING  
VALUES =NO;

\* format HEMO HEMO2 HEMO.  
      SEX SEX.  
      RACE RACE.  
      PHASE2 PHASE3 HASTUBE DIED NO\_YES.  
      DXSOURCE DXSOURCE.  
      PTYPE PTYPE.;

run;  
quit;

CODEBOOK FOR CSSCD PATIENT ROSTER  
CSSCD INFANT COHORT PATIENTS

---

---

CLINICL ----- CLINIC LETTER CODE **DELETED**

type: string (str1)

unique values: 15

coded missing: 0 / 467

| tabulation: | Freq. | Value |
|-------------|-------|-------|
|             | 28    | "A"   |
|             | 15    | "D"   |
|             | 20    | "E"   |
|             | 53    | "F"   |
|             | 17    | "G"   |
|             | 27    | "H"   |
|             | 40    | "J"   |
|             | 47    | "M"   |
|             | 49    | "O"   |
|             | 10    | "P"   |
|             | 28    | "Q"   |
|             | 6     | "R"   |
|             | 7     | "U"   |
|             | 42    | "W"   |
|             | 78    | "Z"   |

PHASE2 ----- ENROLLED IN PHASE 2

type: numeric (float)  
label: PHASE2

range: [1,1]  
unique values: 1

units: 1  
coded missing: 0 / 467

| tabulation: | Freq. | Numeric | Label |
|-------------|-------|---------|-------|
|             | 467   | 1       | YES   |

PHASE3 ----- ENROLLED IN PHASE 3

type: numeric (float)  
label: PHASE3

range: [0,1]  
unique values: 2

units: 1  
coded missing: 0 / 467

| tabulation: | Freq. | Numeric | Label |
|-------------|-------|---------|-------|
|             | 89    | 0       | NO    |
|             | 378   | 1       | YES   |

**CODEBOOK FOR CSSCD PATIENT ROSTER**  
CSSCD INFANT COHORT PATIENTS

---

---

HEMO ----- CDC HEMOGLOBIN DIAGNOSIS

          type: numeric (float)  
          label: HEMO

          range: [-9,8]                                  units: 1  
unique values: 8                                  coded missing: 0 / 467

          tabulation: Freq.  Numeric  Label

|     |    |                  |
|-----|----|------------------|
| 11  | -9 |                  |
| 207 | 1  | SS               |
| 135 | 2  | SC               |
| 22  | 3  | SB+ THAL         |
| 3   | 4  | SBO THAL         |
| 84  | 5  | SS ALPHA THAL    |
| 4   | 6  | SBO + ALPHA THAL |
| 1   | 8  | OTHER VARIANT    |

HEMO:

1. CDC hb diagnosis - also see HEMO2

HEMO2 ----- LOCAL HEMOGLOBIN DIAGNOSIS

          type: numeric (float)  
          label: HEMO2

          range: [1,4]                                  units: 1  
unique values: 3                                  coded missing: 454 / 467

          tabulation: Freq.  Numeric  Label

|   |   |          |
|---|---|----------|
| 8 | 1 | SS       |
| 3 | 3 | SB+ THAL |
| 2 | 4 | SBO THAL |

HEMO2:

1. Local hb diagnosis
2. Used to update HEMO to include more patients in analyses

DXSOURCE ----- SOURCE OF HEMOGLOBIN DIAGNOSIS

          type: numeric (float)

          range: [1,3]                                  units: 1  
unique values: 3                                  coded missing: 447 / 467

          tabulation: Freq.  Value

|   |   |             |
|---|---|-------------|
| 7 | 1 | CDC         |
| 4 | 2 | FORM 14     |
| 9 | 3 | CLINIC EDIT |

DXSOURCE:

1. Applicable only if HEMO2>0

**CODEBOOK FOR CSSCD PATIENT ROSTER**  
CSSCD INFANT COHORT PATIENTS

---

---

F04ALPHA ----- # OF ALPHA GENES **DELETED**

type: numeric (float)  
range: [-9,5] units: 1  
unique values: 6 coded missing: 0 / 467

| tabulation: | Freq. | Value |
|-------------|-------|-------|
|             | 32    | -9    |
|             | 147   | -7    |
|             | 9     | 2     |
|             | 80    | 3     |
|             | 196   | 4     |
|             | 3     | 5     |

F04ALPHA:

1. Not required if HEMO=2 or 3

HAP5COMP ----- HAS RESULTS BASED ON 5 RFLP SITES **DELETED**

type: string (str3)  
unique values: 2 coded missing: 159 / 467

| tabulation: | Freq. | Value |
|-------------|-------|-------|
|             | 5     | "NO"  |
|             | 303   | "YES" |

HAP5COMP:

1. Not required if hemo=2 or 3

C\_HAP5 ----- HAPLOTYPE BASED ON 5 RFLP SITES **DELETED**

type: string (str12)  
unique values: 17 coded missing: 159 / 467

| tabulation: | Freq. | Value        |
|-------------|-------|--------------|
|             | 38    | "BEN"        |
|             | 93    | "BEN/BEN"    |
|             | 8     | "BEN/CAM"    |
|             | 64    | "BEN/CAR"    |
|             | 29    | "BEN/SEN"    |
|             | 1     | "CAM"        |
|             | 1     | "CAM/CAM"    |
|             | 10    | "CAR"        |
|             | 6     | "CAR/CAM"    |
|             | 17    | "CAR/CAR"    |
|             | 6     | "CAR/SEN"    |
|             | 5     | "INCOMPLETE" |
|             | 21    | "NO MATCH"   |
|             | 4     | "SEN"        |
|             | 1     | "SEN/CAM"    |
|             | 1     | "SEN/SAU"    |
|             | 3     | "SEN/SEN"    |

C\_HAP5:

1. Not required if hemo=2 or 3

CODEBOOK FOR CSSCD PATIENT ROSTER  
CSSCD INFANT COHORT PATIENTS

HAPCOMPL ----- HAPLOTYPING STATUS BASED ON 7 RFLP SITES **DELETED**

type: string (str12)

unique values: 2                          coded missing: 159 / 467

| tabulation: | Freq. | Value          |
|-------------|-------|----------------|
|             | 302   | "ALL RESULTS"  |
|             | 6     | "SOME RESULTS" |

HAPCOMPL:

1. Not required if hemo=2 or 3

C\_HAP7 ----- HAPLOTYPE BASED ON 7 RFLP SITES **DELETED**

type: string (str12)

unique values: 8                          coded missing: 159 / 467

| tabulation: | Freq. | Value        |
|-------------|-------|--------------|
|             | 52    | "BEN"        |
|             | 8     | "BEN/CAM"    |
|             | 9     | "CAR"        |
|             | 6     | "CAR/CAM"    |
|             | 6     | "INCOMPLETE" |
|             | 218   | "NO MATCH"   |
|             | 7     | "SEN"        |
|             | 2     | "SEN/SEN"    |

C\_HAP7:

1. Not required if hemo=2 or 3

HASTUBE ----- DNA SAMPLE AT REPOSITORY? **DELETED**

type: numeric (float)  
label: HASTUBE

range: [0,1]                                  units: 1  
unique values: 2                          coded missing: 127 / 467

| tabulation: | Freq. | Numeric | Label |
|-------------|-------|---------|-------|
|             | 37    | 0       | NO    |
|             | 303   | 1       | YES   |

HASTUBE:

1. DNA sample NOT REQUIRED if hemo=2 or 3

**CODEBOOK FOR CSSCD PATIENT ROSTER**  
 CSSCD INFANT COHORT PATIENTS

SETNO ----- # OF SERA SAMPLE DATES

type: numeric (float)

range: [1,6] units: 1  
 unique values: 6 coded missing: 41 / 467

tabulation: Freq. Value

|     |   |
|-----|---|
| 129 | 1 |
| 57  | 2 |
| 111 | 3 |
| 104 | 4 |
| 20  | 5 |
| 5   | 6 |

SEX ----- GENDER

type: numeric (float)  
 label: SEX

range: [1,2] units: 1  
 unique values: 2 coded missing: 0 / 467

tabulation: Freq. Numeric Label

|     |   |        |
|-----|---|--------|
| 222 | 1 | FEMALE |
| 245 | 2 | MALE   |

RACE ----- RACE

**MODIFIED (1=Black, 2=Other)**

type: numeric (float)  
 label: RACE

range: [-9,3] units: 1  
 unique values: 4 coded missing: 0 / 467

tabulation: Freq. Numeric Label

|     |    |           |
|-----|----|-----------|
| 1   | -9 |           |
| 449 | 1  | BLACK     |
| 1   | 2  | CAUCASIAN |
| 16  | 3  | OTHER     |

PTYPE ----- PROTOCOL TYPE DURING PHASE 1

type: numeric (float)  
 label: PTYPE

range: [1,2] units: 1  
 unique values: 2 coded missing: 0 / 467

tabulation: Freq. Numeric Label

|     |   |               |
|-----|---|---------------|
| 465 | 1 | FULL PROTOCOL |
| 2   | 2 | SEMI-PROTOCOL |



**CODEBOOK FOR CSSCD PATIENT ROSTER**  
CSSCD INFANT COHORT PATIENTS

---

---

AGE\_D0E ----- AGE AT DATE OF PHASE 1 ENTRY (DAYS)

          type: numeric (float)

          range: [3,282]                                  units: 1

unique values: 172                                  coded missing: 0 / 467

          mean: 96.9315

          std. dev: 47.6321

percentiles:          10%      25%      50%      75%      90%

                          42          59          87         131         169

AGE\_D0E:

1. Age in days on Phase 1 entry date

AGE\_D0E2 ----- AGE AT DATE OF PHASE 2 ENTRY (YRS)

          type: numeric (float)

          range: [1,11]                                  units: 1

unique values: 11                                  coded missing: 0 / 467

          mean: 6.16916

          std. dev: 2.70575

percentiles:          10%      25%      50%      75%      90%

                          2          4          6          8         10

AGE\_D0E2:

1. Age in years on Phase 2 entry date

AGE\_D0E3 ----- AGE AT DATE OF PHASE 3 ENTRY (YRS)

          type: numeric (float)

          range: [6,16]                                  units: 1

unique values: 11                                  coded missing: 89 / 467

          mean: 10.5159

          std. dev: 2.74195

percentiles:          10%      25%      50%      75%      90%

                          7          8         10         13         15

AGE\_D0E3:

1. Age in years on Phase 3 entry date

**CODEBOOK FOR CSSCD PATIENT ROSTER**  
CSSCD INFANT COHORT PATIENTS

---

---

AGE\_EOF ----- AGE AT END OF FOLLOW-UP (YRS)

type: numeric (float)

range: [3,19] units: 1  
unique values: 17 coded missing: 0 / 467

mean: 12.6531  
std. dev: 3.2834

|              |     |     |     |     |     |
|--------------|-----|-----|-----|-----|-----|
| percentiles: | 10% | 25% | 50% | 75% | 90% |
|              | 9   | 10  | 13  | 15  | 17  |

\_dta:

1. Created 06/01/01