

## Table of Contents

Study Issues .....	1
Subject Id Number Assignment .....	1
Linking Hemophiliac to Household Member .....	2
Hemophiliac Centers by Study Id Code .....	3
Adding a New Hemophiliac Center .....	3
Source Code File .....	4
ASSGN    Assign Study Id from Subject Id .....	4
CONTFMT  Formats Study Id's into U.S and European .....	4
FMT      Formats the Study Id .....	5
HEMSUBJ  Selects Hemophiliacs by Subject Id .....	5
SELCTHEM Select Hemophiliacs by Study/Subject .....	6
SELCTSTY Select the Hemophiliac Study Ids .....	6
SUBSTRNG Separate Initials from Subject Ids .....	7
Subject Id Numbers with Initials .....	7
File Name .....	7
File of Linked Couples .....	8
Analysis File .....	9
Variables .....	9
Issues Concerning HIV Date Variables .....	10
MID_YN .....	10
Issues Concerning Status Variables .....	11

Updating the File .....	12
Phil Rosenberg's Sero-conversion Date/Age .....	12
Current File .....	12
Origin of the Data .....	13
'AIDS' File (AIDS Info,LKA,TYPE,SEV) .....	14
Type and Severity .....	15
Updating the File .....	15
Collapsed Screen and Blot Files .....	15
Collapsed Screen Algorithm .....	16
Samples to be Blotted .....	17
File Formats .....	17
Updating the Files .....	17
HLA Files.....	18
Analysis File .....	18
Population File .....	19
Report File .....	20
Flag File .....	21
Programs to Create Inventory File.....	21
By Sample for Serum/Plasma .....	22
By Sample/Material Type for Serum/Plasma/Lymphocytes .....	22
Factor VIII and Factor IX Exposures .....	23
Clinical File Variables .....	23

Product Exposure Summary File .....	24
Defaults set on the exposures.....	25
Updating the File .....	25
F8/F9 Yearly Dose Flag File .....	25
T-Cell Results .....	26
Overview.....	26
Complete Hemophilia T-Cell File .....	27
Laboratory Files .....	30
RTI File.....	30
Updating the Files.....	31
Rolling Means T-Cell File .....	31
'CAT' Date File .....	32
P-24 Antigen File.....	33
Edits .....	33
Variables .....	34
Error Reports.....	34
Input Files .....	34
Updating the File .....	34
Neopterin and Beta-2 Microglobulin.....	34
Variables .....	35
Updating the File .....	35
Drug File (AZT and Others).....	35

Variables .....	35
Input Files .....	36
Updating the File .....	36
Hepatitis B Status File .....	36
Input Files .....	37
Updating the File .....	37
Variables .....	38
Hepatitis B 'Immune' and 'Infected' Codes .....	40
Interferon File .....	41
Variables .....	41
Input Files .....	42
Updating the File .....	42
Sibling Study.....	42
Analysis File .....	42
AIDS Analysis Programs.....	43
CD4 Analysis File.....	43
CD4 Analysis Programs .....	44
PCR Data .....	44
Transactions .....	44
File Format.....	44
Backup Tapes.....	46
Samples tested for HCV bDNA, HCV Genotyping, HCV Serotyping, and/or HIV Monitor .....	46

**Study Issues****Subject Id Number Assignment**

HER001TRQ            A00101001

All hemophiliacs belonging to the same hemophilic center will have the first three characters of their subject id's in common. All hemophilic household members belonging to the same hemophilic center will have subject id numbers beginning with the last letter of the study id number for that center. The following is the subject id number assignments:

STUDY	HEMOPHILIAC	HOUSEHOLD MEMBER
HEA	HER	A
HEC	COR	C
HED	DEN	D
HEE	PHI	E
HEG	GRE	G
HEH	PIT	H
HEJ	MTS	J
HEK	TEX	K
HEL	CHI	none
HEM	MUN	M
HEP	HOP	none
HES	SWI	S
HET	TUL	T
HEU	UNC	U
HEV	CLE	V
HEW	GWU	W
HEZ	VIE	Z

The subject id numbers for the household members have the center related first character followed by 8 digits. The subject id numbers for the hemophiliacs have the three characters followed by a three digit number, and 0 - 3 letters representing their first, middle and last initials.

**WARNING:** For years, the subject's initials changed constantly, with no warning. Therefore, no file under NTT1SWY include these initials. The only ARC maintained files which include the initials are the system files such as the Background and Results data bases. Therefore before merging any non system file with any non-system or RTI maintained file, the subject ID numbers on the non-SWY file must be substringed for the hemophiliacs to include only the first six characters.

**Linking Hemophilic to Household Member**

The following shows the relationship between ID number 'HER004' (hemophiliac) and 'A00401001' (his female partner).

A004-01-001

- A - Designates the hemophilia center, in this case Hershey. The leading letter will be the same as the last letter of the Study Id, in this case 'HEA'.
- 004 - The '004' links to the hemophiliac's number '004'.
- 01 - Represents the relationship to the hemophiliac. In this case, the hemophiliac's female sexual partner.
- 001 - Sequential number indicating the number of individuals in the study with the same relationship to hemophiliac. In this case, sexual partner # 1.

**HOUSEHOLD CODES:**

- 01 - Female sexual partner, usually wife.
- 02 - Male sexual partner
- 03 - Sister
- 04 - Brother
- 05 - Mother
- 06 - Father
- 07 - Daughter
- 08 - Son
- 09 - Aunt
- 10 - Uncle
- 11 - Niece
- 12 - Nephew
- 13 - Grandmother
- 14 - Grandfather
- 15 - Granddaughter
- 16 - Grandson
- 17 - Female cousin
- 18 - Male cousin
- 19 - Other blood relative
- 20 - Non-blood relative
- 21 - Baby of 01 fathered by a non-hemophiliac (stepchild of hemophilia subject)
- 22 - Non-hemophiliac father of 21, sexual partner of 01 (ex-partner of 01)
- 23 - Half-brother

**Hemophiliac Centers by Study Id Code**

The following is a list of the Hemophiliac Centers participating in the study and the correlated Study Id code.

HEA	-	Hershey
HEC	-	Cornell
HED	-	Denver
HEE	-	Cardeza (Philly)
HEG	-	Greece
HEH	-	Pittsburg
HEJ	-	Mount Sinai
HEK	-	Houston
HEL	-	Children's DC
HEM	-	Munich
HEP	-	Philadelphia Children's Hospital (CHOP)
HES	-	Switzerland
HET	-	Tulane
HEU	-	University of North Carolina
HEV	-	Cleveland
HEW	-	GWU
HEZ	-	Vienna

### **Adding a New Hemophiliac Center**

Periodically, new Hemophiliac centers will join the study. When this occurs, the RTI study manager, Barbara Kroner of RTI, will contact ARC with a request for a new study ID designator. In terms of the Background and lab Results systems, a new center translates into a new study ID.

The following information need to be gotten from the client or the RTI study manager:

1. The STUDY ID NUMBER - Be sure this number was cleared by ARC ( Phil Virgo, or whoever is presently in charge of new designator ). It may happen that RTI requested a STUDY ID that has already been assigned.
2. What will the SUBJECT ID's start with.
3. Will there be female partners included in the study.

### **CHANGES:**

All members of the source code library NTT1SWY.CS5.SOURCE.L80 need to be changed to

include the new information.

### Source Code File

This file contains the source code to perform commonly needed functions.

\*\* Subsetting should ALWAYS be done using the subset members of this file.

#### **ASSGN Assign Study Id from Subject Id**

```
IF SUBJ_ID =: 'HER' OR
  ('A0' <= SUBSTR(SUBJ_ID,1,2) <= 'A9') THEN STUDY_ID = 'HEA';
```

etc

#### **CONTFMT Formats Study Id's into U.S and Europeana)**

```
PROC FORMAT;
  VALUE $CONTNT
    'HEA', 'HEC', 'HED', 'HEE',
    'HEH', 'HEJ', 'HEK', 'HEL',
    'HEP', 'HET', 'HEU', 'HEV',
    'HEW' = 'US'
    'HEG', 'HEM', 'HES', 'HEZ' = 'EU'
  ;
```

#### **FMT Formats the Study Id**

This code formats the Study Id into a less formalized format than the HIV7 system format library.

```
PROC FORMAT;
  VALUE $HEMSTUD
    "HEA" = "Hershey"
    "HEC" = "Cornell"
    "HED" = "Denver"
    "HEE" = "Cardeza"
    "HEG" = "Greece"
    "HEH" = "Pittsburg"
    "HEJ" = "Mount Sinai"
    "HEK" = "Texas"
```

```
"HEL" = "Children's DC"
"HEM" = "Munich"
"HEP" = "Phila Children's"
"HES" = "Switzerland"
"HET" = "Tulane"
"HEU" = "Univ. of N Caro"
"HEV" = "Cleveland"
"HEW" = "GWU"
"HEZ" = "Vienna"
;
```

### **HEMSUBJ Selects Hemophiliacs by Subject Id**

This code selects hemophiliacs only (eliminating the household members). It should not be used when in a system file which can be subsetted by Study and Subject.

```
IF SUBJ_ID=: 'HER' OR
SUBJ_ID=: 'COR' OR
SUBJ_ID=: 'PIT' OR
SUBJ_ID=: 'PHI' OR
SUBJ_ID=: 'MTS' OR
SUBJ_ID=: 'CHI' OR
SUBJ_ID=: 'GWU' OR
SUBJ_ID=: 'HOP' OR
SUBJ_ID=: 'TUL' OR
SUBJ_ID=: 'CLE' OR
SUBJ_ID=: 'UNC' OR
SUBJ_ID=: 'DEN' OR
SUBJ_ID=: 'GRE' OR
SUBJ_ID=: 'MUN' OR
SUBJ_ID=: 'SWI' OR
SUBJ_ID=: 'TEX' OR
SUBJ_ID=: 'VIE';
```

### **SELCTHEM Select Hemophiliacs by Study/Subject**

This code selects hemophiliacs only (eliminating the house-hold members by Study and Subject Id). This should be used when subsetting system files such as the Background or Results files.

```
IF (STUDY_ID = 'HEA' AND SUBJ_ID=: 'HER') OR
(STUDY_ID = 'HEC' AND SUBJ_ID=: 'COR') OR
```

(STUDY\_ID = 'HEH' AND SUBJ\_ID=: 'PIT') OR  
 (STUDY\_ID = 'HEE' AND SUBJ\_ID=: 'PHI') OR  
 (STUDY\_ID = 'HEJ' AND SUBJ\_ID=: 'MTS') OR  
 (STUDY\_ID = 'HEK' AND SUBJ\_ID=: 'TEX') OR  
 (STUDY\_ID = 'HEL' AND SUBJ\_ID=: 'CHI') OR  
 (STUDY\_ID = 'HEW' AND SUBJ\_ID=: 'GWU') OR  
 (STUDY\_ID = 'HEP' AND SUBJ\_ID=: 'HOP') OR  
 (STUDY\_ID = 'HET' AND SUBJ\_ID=: 'TUL') OR  
 (STUDY\_ID = 'HEV' AND SUBJ\_ID=: 'CLE') OR  
 (STUDY\_ID = 'HEU' AND SUBJ\_ID=: 'UNC') OR  
 (STUDY\_ID = 'HED' AND SUBJ\_ID =: 'DEN') OR  
 (STUDY\_ID = 'HEG' AND SUBJ\_ID =: 'GRE') OR  
 (STUDY\_ID = 'HEM' AND SUBJ\_ID =: 'MUN') OR  
 (STUDY\_ID = 'HES' AND SUBJ\_ID =: 'SWI') OR  
 (STUDY\_ID = 'HEZ' AND SUBJ\_ID =: 'VIE');

**SELCTSTY Select the Hemophiliac Study Ids**

IF (STUDY\_ID='HEA') OR  
 (STUDY\_ID='HEC') OR  
 (STUDY\_ID='HEH') OR  
 (STUDY\_ID='HEE') OR  
 (STUDY\_ID='HEJ') OR  
 (STUDY\_ID='HEK') OR  
 (STUDY\_ID='HEL') OR  
 (STUDY\_ID='HEW') OR  
 (STUDY\_ID='HEP') OR  
 (STUDY\_ID='HET') OR  
 (STUDY\_ID='HEV') OR  
 (STUDY\_ID='HEU') OR  
 (STUDY\_ID='HED') OR  
 (STUDY\_ID='HEG') OR  
 (STUDY\_ID='HEM') OR  
 (STUDY\_ID='HES') OR  
 (STUDY\_ID='HEZ');

**SUBSTRNG Separate Initials from Subject Ids**

IF SUBJ\_ID =: 'HER' OR SUBJ\_ID =: 'MTS' OR  
 SUBJ\_ID =: 'COR' OR SUBJ\_ID =: 'PHI' OR  
 SUBJ\_ID =: 'UNC' OR SUBJ\_ID =: 'CLE' OR  
 SUBJ\_ID =: 'TUL' OR SUBJ\_ID =: 'PIT' OR

```

SUBJ_ID =: 'GWU' OR SUBJ_ID =: 'DEN' OR
SUBJ_ID =: 'GRE' OR SUBJ_ID =: 'MUN' OR
SUBJ_ID =: 'SWI' OR SUBJ_ID =: 'VIE' OR

SUBJ_ID =: 'TEX' OR
SUBJ_ID =: 'HOP' OR SUBJ_ID =: 'CHI' THEN DO;
    _INITIAL = SUBSTR(SUBJ_ID,7,3);
    SUBJ_ID = SUBSTR(SUBJ_ID,1,6);
END;

```

### Subject Id Numbers with Initials

**File Name:** NTT1SWY.CS5.IDS.SSD(.HEMOPH)

There is often inconsistencies in the initials that the centers put on the end of the hemophiliac's ID numbers. For that reason, all NTT1SWY files have the initials stripped off of the SUBJ\_ID variable. However, there are frequently cases where the initials are needed, usually on listings going to the centers where the staff recognize the initials rather than the ID number. This file contains the id numbers for every hemophiliac in the background file in both formats.

#### Variables

SUBJ_ID	Character 11	Subject ID Number without initials
WHOLE_ID	Character 11	Subject ID Number with initials

#### Updating the file

Automatic update is part of the Analysis File update process.

### File of Linked Couples

**NTT1SWY.CS5.FAMLINK.SASDS(.HEMOPH)**

Female Partner analysis usually involves selecting on the status of both the hemophiliac and the partner. This file is one record per partner, sorted by FAMID.

**\*\*\* NOTE:** Records of hemophiliacs with more than one female partner will have the same FAMID's

#### Variables

FAMID	-	ID that links the family records. The first character and three numerals of the partner. ie (A001) links A00101001 and HER001.
FEMID	-	Female Partner Id Number
FEMSTAT	-	Analysis HIV status of the female partner
HEMID	-	Hemophiliac's subject Id Number (No Initials)
HEMSTAT	-	Analysis HIV status of the hemophiliac

**Updating the file:** Run from &NTT1SWY.CS5.ADHOC.FILES.SAS(FAMLINK)

## Analysis File

### NTT1SWY.CS5.ANALYSIS.SASDS(HEMOPH)

#### Variables

\*\*\* One record per subject, sorted by SUBJ\_ID

This file contains the HIV status and the variables for the sero-conversion related dates. It derives most of its variables from data in the VES Results File and the CS5 collapsed screen file. It contains three sets of sero-conversion dates:

- \* the real dates of first positive, last negative, middate
- \*\* the dates estimated for analysis using default last negs and first pos's
- \*\*\* dates estimated for analysis by Phil Rosenberg.

The variables are as follows:

COH	-	Cohort, 'H'= Hemophiliac 'W'= Wife or female partner 'O'= Other
FRST_POS	-	First HIV-1 positive test results; *
FSTPSANL	-	First positive for analysis purposes; ** for seroconverters this is the same as frst_pos; for prevalent positives this has an upper limit of 1/1/87
LAST_NEG	-	Last HIV-1 negative test results; *

LSTNGANL	-	Last negative for analysis purposes; ** for seroconverters this is the same as last_neg; for prevalent positives this is assigned - 7/1/78 or birth date, if after 7/1/78
MDATE	-	Seroconversion date; * mid point between LAST_NEG and FRST_POS
MDATEANL		Date of seroconversion; ** midway between FSTPSANL and LSTNGANL
MIDAGE	-	Age at ANALYSIS Seroconversion ** (MDATEANL)
MID_YN	-	seroconversion status: 1 = seroconverter 0 = prevalent positive . = negative
PHILAGE	-	Age at Phil R. seroconversion; *** (SERODT)
SERODT	-	Date of Seroconversion; *** Phil R. imputed date
STATUPDT	-	HIV status (POS NEG REV IND PND) for update purposes only
STATANL	-	HIV status used for analysis purposes
STUDY_ID	-	Study Id number
SUBJ_ID	-	Subject Id number

### Issues Concerning HIV Date Variables

FRST\_POS and LAST\_NEG represent the true draw dates of the subject's first positive and last negative tests. These subjects are considered true sero-converters. MDATE is the midpoint between these two dates.

FSTPSANL and LSTNGANL are estimated last neg and first positive dates applied to the prevalent positive hemophiliacs (NOT FEMALE PARTNERS), based on what is known about the range of dates of contaminated blood products in the U.S. The main function of these variables is to prevent the skewing of statistics which would occur if recent enrollees report a first positive of 1990, for example, when they have probably been positive since the early to mid eighties. MDATEANL is the mid-point between these dates.

SERODT is the estimated seroconversion date imputed by Phil Rosenberg using age, geographic location and exposure to contaminated product.

\*\*\* For **analysis** purposes use either FSTPSANL/LSTNGANL/MDATEANL OR SERODT. Check with client.

\*\*\* For other purposes check with client, or with Jim if client is unaware of issue

First Positive is set if:

- Subject Blots Positive
- Subject screened positive before 1990 and has no subsequent conflicting blot
- Subject screens positive and has AIDS

Last Negative is set if:

- Subject screens negative, with no conflicting blot

### **MID\_YN**

- 1 Seroconverter; Has true first positive and last negative dates
- 0 Prevalent Positive; Does not have true last negative date
- Missing Subject is HIV negative

### **Issues Concerning Status Variables**

STATUPDT is the variable that gives detailed information concerning the subject's HIV status.

- Negative Subject is negative with no conflicts  
NO FIRST POSITIVE, HAS LAST NEGATIVE
- Positive Positive by Blot or Positive by screen with AIDS  
HAS FIRST POSITIVE, MAY HAVE LAST NEGATIVE
- Indeterminate Positive with no blot possible and no AIDS OR  
Negative person last sample blotting '/' OR last screen '/' and no blot  
MAY HAVE FIRST POSITIVE - if screens positive with no way to blot and no AIDS  
MAY HAVE LAST NEGATIVE - if negative first
- Pending Subject screens positive for first time and waiting for blot result OR  
Positive subject screens negative and waiting for blot result  
MAY HAVE FIRST POSITIVE - if status was IND with a first positive date (see above) and waiting for blot result  
MAY HAVE LAST NEGATIVE
- Reverter Positive subject that subsequently blots negative  
HAS FIRST POSITIVE - set before reversion

**MAY HAVE LAST NEGATIVE**

STATANAL is the variable used for analysis purposes. It is not as detailed as STATUPDT but selects the subjects in a way consistent with analysis performed for years. The rules for setting are as follows:

- Positive                      First positive not equal missing
- Negative                      First positive equal missing and last neg not equal missing
- Missing                      Neither Positive nor Negative

**Updating the File**

This is done by running the following command:

**Execute from NTT1SWY.CS5.UPDATE.SOURCE(UPDJCL)**

and choosing the option '(A) - ANALYSIS UPDATE'

JCL Stream: NTT1SWY.CS5.UPDATE.SOURCE(JCL)

Source Code: NTT1SWY.CS5.UPDATE.SOURCE(HEMOPH)

**Phil Rosenberg's Sero-conversion Date/Age**

**Current File NTT1SWY.CS5.PHILR.SERODATE(.PHILDATE)**

\*\*\* One record per subject of imputed seroconversion dates. Contains subjects that are not in our traditional analysis. Most of these are not on the background file, have no birthdate and therefore have no seroage.

\*\*\* Since the time that Phil imputed his dates old draw dates have been retested, and additional historical data have been added. Therefore, we found subjects with imputed dates that were no longer accurate - dates that came before subsequent samples that tested negative. We have temporarily set these serodates to the date of the last neg + 6 months. The plan is for Phil Rosenberg to recalculate all dates once the new Texas center is on-line.

**Variables**

PHILAGE     -     Seroconversion age

REALPHIL - Flag indicating the origin of the date  
 0 - Jim's defaults based on country and hemophilia type/severity  
 1 - Phil's imputation  
 2 - Jim's assignment using Phil's technique  
 SERODT - Seroconversion date (SAS date)  
 SUBJ\_ID - Our subject id (no initials)

### Input Files

1. NTT1SWY.CS5.PHILR.ORIG.SASDS(.PHILDATE)
2. NTT1SWY.CS5.ANALYSIS.SASDS(.HEMOPH)
3. NTT1SWY.CS5.RTI.ORI.VARS.SASDS(.HEMOPH)

### Updating the File

Run from NTT1SWY.CS5.PHILR.SAS(CRSERODT)

This program does the following:

- Keeps all subjects that are in files 1 or 2.
- Keeps the serodate from file 1 for all subjects that are on that file.
- Assigns default seroconversion dates for the other subjects
- Calculates age at seroconversion
- Sets the REALPHIL flag based on the origin of the data
- If the imputed or defaulted seroconversion date is earlier than the last negative draw (LAST\_NEG from analysis file) the serodate is set to the Last Negative + 6 months.

### Origin of the Data

\*\*\* Phil Rosenberg wrote statistical programs that imputed a seroconversion date for all the HIV positive hemophiliacs that had sufficient data at the time (1992). For his calculation he used the HIV test results on the subject's blood samples, the amount of non heat-treated Factor VIII and IX the subject was exposed to, as well as the hemophilia center the subject received it at (since different manufacturers had different rates of contaminated product). The files containing the output of his programs was edited into data - subject id and imputed seroconversion date (n=1245). Jim Goedert then assigned 29 additional subjects seroconversion dates using the criteria of Phil's program. This file contains these 1274 subjects.

**NTT1SWY.CS5.PHILR.ORIG.SASDS(.PHILDATE)**

### Variables:

INORIG - flag indicating the origin of the data  
 'P' - Phil's imputation  
 'J' - Jim's assignment using Phil's technique  
 SERODT - Seroconversion date (SAS date)  
 SUBJ\_ID - Our subject id (no initials)

### **Input Files**

NTT1SWY.CS5.PHILR.ORIG.L16  
 Temporary flat file from the output of  
 Phil's PC programs. On diskette as  
 \PHILR\SEROFIELD.DAT

NTT1SWY.CS5.PHILR.OCT93(.PHILDATE)  
 First SAS file created to hold this information

### **Create Program**

NTT1SWY.CS5.PHILR.SAS(CREORIG)

The program merges the two input files together. It keep all the subjects from the flat file (n=1245) and all the subjects from the OCT94 file that nmjmh had the flag (REALPHIL) set to '1' (n=29). These 29 had their seroconversion date set by Jim by hand, using the technique developed by Phil Rosenberg.

### **'AIDS' File (AIDS Info,LKA,TYPE,SEV)**

NTT1SWY.CS5.RTI.ORI.VARS.SASDS(.HEMOPH)

\*\*\* One record per subject, sorted by SUBJ\_ID

This file contains important summary variables created from information that comes from more than one source, often from both ARC and RTI maintained files. This file is particularly helpful to use for Kaplan Meier analysis. Most variables commonly used for KM events and cutoffs have been added to the file.

AIDS - AIDS  
 DAIDS - Date AIDS dx (first manifestation) SAS  
 DEAD - Dead (0/1)  
 DT\_BIRTH - Birth Date mmddyy  
 DT\_AIDS - Date AIDS dx (first manifestation) mmddyy

DT_DEATH	-	Date of death mmddyy
HIVNR	-	HIV Neurologic Reaction AIDS dx
HIVNR_DT	-	Date HIVNR dx (SAS)
INHIB_DT	-	Date of inhibitor onset
INHLOST	-	Inhibitor ever lost(0/1)
INHLVL	-	Inhibitor level (H/L)
INHLVLBU	-	Highest inhibitor level (Bethesda units)
KS	-	Kaposi Sarcoma AIDS dx
KS_DT	-	Date KS dx (SAS)
LK_ALIVE	-	Last Alive ('999999' if dead) mmddyy.
		LKA - Date Death (if dead) Last Known Alive (if alive) (SAS)
LYMPHOMA		Lymphoma - AIDS dx
LYMPH_DT	-	Date LYMPHOMA dx (SAS)
NEWTYP	-	Type of hemophilia
OTHOI	-	Other Opportunistic Infection AIDS dx
OTHOI_DT	-	Date OTHOI dx (SAS)
PCP	-	Pneumocystis AIDS dx
PCP_DT	-	Date PCP dx (SAS)
RCHLDT	-	Date of first factor rechallenge
RCHLYN	-	Factor rechallenge (0/1)
RDVINH	-	Inhibitor redeveloped (0/1)
RACE	-	Race
SEV	-	Severity of hemophilia
SEX	-	Sex

### Type and Severity

The type and severity variables are the only unfamiliar or non 0/1 variables on the file. The breakdown of the codes are as follows:

NEWTYP		SEV
0	=	Other
0.1	=	Other with Inhibitor
1	=	Von Willebrands
1.1	=	VW with Inhibitor
2	=	B
2.1	=	B with Inhibitor
3	=	A
3.1	=	A with Inhibitor

**Updating the File**

Run from NTT1SWY.CS5.UPDATE.SOURCE(CREAIDS)

**Collapsed Screen and Blot Files****NTT1SWY.CS5.NOV12.SCRN on CAT**

The collapsed screen and blot file is to be used by any programs needing the collapsed screen value, or information concerning submission for western blotting on all samples for the Hemophiliac Studies. This file will be updated every time the VES results data base is updated with new lab results.

**Variables**

The file will contain one record per sample ID and the following variables:

Subject Id	1-11	Subject ID
Sample Id	15-21	Sample ID
Date Drawn	22-27	Date Drawn YYMMDD
Screen Result	28	Set to value Based on licensed results:
	'+' ->	Positive
	'-' ->	Negative
	'/' ->	CONF results of the licensed screens conflict or licensed test result = '/'.
Blot Conf	37	Confirmation of Western Blot
Blot Conf2	39	Confirmation 2 of Western Blot
Rcvd_DT	41	Date into CSC's Results File

**NTT1.CS5.TOBLLOT.L7 PDS on CAT**

The update program also creates a file containing the SAMPLE ID's chosen to be Western Blotted.

**Collapsed Screen Algorithm****Samples tested before 01/01/90**

The old screen algorithm was applied and the screen flag was set accordingly. If the sample was blotted the blot flag was set accordingly. These samples will not be evaluated again unless they are retested.

**Samples tested after 01/01/90**

The program will ignore all non-licensed tests. If only one licensed test was done or

more than one was done and the results were concordant, the screen result will be set to the 'CONF' field(s). In the event that more than one licensed test was performed and the CONF values conflict, the screen value will be set to '/'.

**Samples to be Blotted**

1. First positive screen for subject  
     Sample flagged to be sent  
     Previous sample flagged to be sent if not already blotted
2. Screen result differs from previous blot (or previous screen if not blotted)  
     Sample flagged to be sent  
     Previous sample flagged to be sent if not already blotted.
3. More than one licensed screen was done and the results were not concordant.  
     Sample flagged to be sent

**File Formats**

NTT1SWY.CS5.NOV12.SCRN

Variable	Position	Description
SUBJ_ID	1-11	Subject ID
SAMPL_ID	15-21	Sample ID
DT_DRAWN	22-27	Draw date YYYYMMDD
SR	28	Screen result
TOBLOT	29	Flagged to be blotted
BLOTCONF	37	Blot Confirmation
BLOTCONF2	39	Blot Conf 2
RCVD_DT	41	Date result entered VES database

NTT1SWY.CS5.TOBLOT.L7(BTCHnnn)

Variable	Start Position	Length	Description
SAMPL_ID	1	7	Sample ID

**Updating the Files**

- o Execute from NTT1SWY.CS5.UPDATE.SOURCE(UPDJCL)

Choose the option '(B) TOBLOT UPDATE'

Source code: NTT1SWY.FS18.ADHOC.SAS(SCREEN)

- o Review the listing and add any samples to DONTBLOT if they belong. \*\*\* Note the number of new seroconverters (especially among the hemophiliacs) and reverters. There should be VERY few. If more than two or so, check the test dates and centers. A large number happening on the same test date suggests a lab error. A large number happening within a center suggests sample switching in the field. If it seems fishy notify Jim and Barbara immediately, and hold off notifying Violet (next step).
- o Notify Violet and Barbara of TOBLOT samples by WYLBUR mail.

## HLA Files

### Analysis File

#### NTT1SWY.CS5.HLA.ANALYSIS.SSD(.HLA)

**This file should be used for all HLA analysis.** It contains one record per subject Id. The program puts the subject id numbers through the 'dup id' file since Debbie Lomb is not always up-to-date on ID changes. It differs from the HLA files used for analysis thus far, in that it resets alleles from testing that was done before more sensitive testing was developed. Alleles that were considered as one, are now seen to be split into two or more alleles that the more sensitive tests can distinguish.

**Example of a splitout:** A9 splits into A23 and A24. Therefore if A9 is 1, A23 and A24 are set to missing if they had been set to 0, because we dont know which it split out into. But if A9 is 0 and (A23 or A24 is 1) then A9 is set to 1 since it had split from A9 to whichever it is.

This splitouts are as follows:

Original Allele	Splitouts
A9	A23 A24
A10	A25 A26 A34
A19	A29 A30 A31 A32 A33
A28	A68 A69
B5	B51 B52
B12	B44 B45
B14	B64 B65
B15	B62 B63
B16	B38 B39
B17	B57 B58
B21	B49 B50
B22	B54 B55 B56

B40	B60 B61
C3	C9 C10
DR2	DR15 DR16
DR3	DR17 DR18
DR5	DR11 DR12
DR6	DR13 DR14
DQ1	DQ5 DQ6
DQ3	DQ7 DQ8 DQ9

**Input files:**

UBJ1WWV.FS6.HLA.SASDS(.HLAHEM)  
 NQG1BYK.DUPIDS

**To update the file:**

Run from &NTT1SWY.CS5.HLASCORE.SAS(SPLITIT)

**Variables:**

A1	A68	B39	B60	DQ5	DR16
A2	A69	B40	B61	DQ6	DR17
A3	B4	B41	B62	DQ7	DR18
A9	B5	B42	B63	DQ8	DR52
A10	B6	B44	B64	DQ9	DR53
A11	B7	B45	B65	DR1	SUBJ_ID
A19	B8	B46	B70	DR2	
A23	B12	B47	C1	DR3	
A24	B13	B48	C2	DR4	
A25	B14	B49	C3	DR5	
A26	B15	B50	C4	DR6	
A28	B16	B51	C5	DR7	
A29	B17	B52	C6	DR8	
A30	B18	B53	C7	DR9	
A31	B21	B54	C9	DR10	
A32	B22	B55	C10	DR11	
A33	B27	B56	DQ1	DR12	
A34	B35	B57	DQ2	DR13	
A36	B37	B58	DQ3	DR14	
A43	B38	B59	DQ4	DR15	

**Population File**

**NTT1SWY.CS5.HLA.POPS.L70**

This file contains the percentage of the white U.S. population that has each individual allele. **The statistice are only valid therefore, when applied to the white U.S. subset of our hemophiliacs.**

## File format

Starting	Length	Format	Description
1	4	Char	Allele
11	5	5.2	Percent of the tested population with the allele
19	4	4.0	Total tested
24	2	Char	Allele tested for in our population. Blank means 'yes', 'NO' means no.

An example of the use of this file for analysis weighted by population is in the program NTT1SWY.CS5.HLASCORE.SAS(USEPOPS)

**Report File****NTT1SWY.CS5.HLA.SASDS(.HLA)**

\*\*\* One record per subject, sorted by SUBJ\_ID.

Input Files: UBJ1WWV.FS6.HLA.SASDS(.HLAHEM)  
(maintained by IMS)

## Variables:

SUBJ_ID	-	Subject Id Number
QUAL	-	Quality of Typing
		A - Complete well defined typing
		B - Well defined Class I (A,B,C)
		G - Well defined Class II (DR, DQ)
HLATYPE	-	Composite of All Typings

The variable HLATYPE can be substringed to get at individual alleles. The breakdown is as follows:

<u>Allele</u>	<u>Columns within HLATYPE</u>
A1	1-3
A2	5-7
A3	9-11
B1	13-15
B2	17-19
B3	21-23
BX1	29-30
BX2	32-33
C1	35-37
C2	39-41
C3	43-45
DR1	47-50
DR2	52-55
DR3	57-60
DRX1	62-65
DRX2	67-70
DQ1	72-74
DQ2	76-78
DQ3	80-82

Updating the File:     Run from &UAT1FMY.HIV4.PROGS(HEMTYPE)

The program checks the input file to make sure there is only one record per subject. It takes the individual allele flags and makes one variable listing the individual's Class I and/or Class II HLA antigens.

### **Flag File**

**UBJ1WVW.FS6.HLA.SASDS(.HLAHEM)**

\*\*\*     One record per subject, maintained and updated by IMS.

### Variables

PID             - Subject Id Number  
 HLA            - Quality of Typing  
 A1...DQ9       - Flags (0/1) for Individual HLA antigens

### **Programs to Create Inventory File**

It is often necessary to incorporate BSI Inventory information in analyses projects. These program can be used to create a SAS file containing the most common information needed. **PLEASE DO NOT CHANGE THE PROGRAM SAVED OR WRITE OVER THE OUTPUT FILE. MODIFY IT TO SUIT YOUR NEEDS AND SAVE YOUR OWN VERSION WHERE IT BELONGS.**

### **By Sample for Serum/Plasma**

#### **NTT1SWY.CS5.INVENTORY.SAS(TOTVOL)**

This program reads the BSI, selects Serum/Plasma samples for the hemophiliacs and creates a file containing:

- One record per sample WITH BLOOD in the BSI with the following variables:

DRAW	-	Date Drawn (SAS)
NUM_VIAL	-	Number of Vials
SAMPL_ID	-	Sample Id
SUBJ_ID	-	Subject Id
TOTVOL	-	Total volume (ml)

### **By Sample/Material Type for Serum/Plasma/Lymphocytes**

#### **NTT1SWY.CS5.INVENTORY.SAS(TOTVLMAT)**

This program reads the BSI, selects Serum/Plasma/Lymphocyte samples for the hemophiliacs and creates a file containing:

- One record per sample/material type WITH BLOOD in the BSI with the following variables:  
\*\*\*\*\*:

DRAW	-	Date Drawn (SAS)
MAT_CODE	-	Material Type
NUM_VIAL	-	Number of Vials
SAMPL_ID	-	Sample Id
SUBJ_ID	-	Subject Id
TOTVOL	-	Total volume (ml)

\*\*\*\*\* Biotech Respository now stores 'pellets', which are dead lymphocytes that can be used for PCR/DNA testing. These pellets have the same material code 'C1' as viable lymphocytes, but should not be included in the same volume/vial counts as viable lymphocytes since the uses are not the same. The pellets are identified by the code 'P3' in either of the modifier fields. THE PROGRAM

GIVES THESE PELLETS A NEW MATERIAL CODE 'DL' (dead lymphocytes) AND CALCULATES THE VIAL AND VOLUME TOTALS SEPARATELY.

### **Factor VIII and Factor IX Exposures**

Analysis runs and prints are often done using the subject's exposure to Factor VIII (F8) and Factor IX (F9) non heat-treated (NT) blood products.

**\*\*\* The NON-HEAT TREATED and HEAT TREATED variables on the Clinical File are coded differently.**

**The Non-heat treated values are on only 1 of the persons records BUT THEY ARE NOT ALWAYS ON THE FIRST RECORD, SO ALL RECORDS MUST BE CHECKED. This is historical data since non heat-treated products are no longer used.**

#### Clinical File Variables

F8NT = Factor VIII Non-heat treated

F9NT = Factor IX Non-heat treated

F8HT = Factor VIII Heat treated

F9HT = Factor IX Heat treated

The **NON HEAT TREATED** values on the clinical file are coded:

- |   |                 |  |
|---|-----------------|--|
| 1 | HIGH            | ( 50,000+ form 6; 1,000,000+ form 1)         |
| 2 | MEDIUM          | (20,000-50,000 form 6; 100,000-1 mil form 1) |
| 3 | LOW             | (1-20,000 form6; 1-100,000 form 1)           |
| 4 | None            |  |
| 5 | Unknown         |  |
| 6 | Born after 1984 |  |

**It is important to recode 6 to be the same as '4' if you are only looking at actual exposure. All blood products were heat-treated in the the U.S. by 1985, so those born after 1984 were not exposed. If doing runs exclusively on European subjects ASK.**

The **HEAT TREATED** values are on multiple records. Use the Summary File (see below) for a summarization of the information. The variables are coded:

0	None
1	LOW
2	HIGH
3	Unknown/None (must be treated as unknown)

### **Product Exposure Summary File**

#### **NTT1SWY.CS5.F8F9.SUMFILE.SASDS(HEMOPH)**

This file contains one record per subject and a summary of the subject's Factor VIII and Factor IX exposure.

Input: NTT1SWY.CS5.CLINFILE.DATA(.CLINFILE)

#### Variables

SUBJ_ID	-	Subject Id
FACT8NT	-	Non-heat treated Factor VIII
FACT8HT	-	Heat treated Factor VIII
FACT9NT	-	Non-heat treated Factor IX
FACT9HT	-	Heat treated Factor IX
LASTF8NT	-	Date last took Non-heat treated F8 (SAS)
LASTF9NT	-	Date last took Non-heat treated F9 (SAS)
LASTF8HT	-	Date last took Heat treated F8 (SAS)
LASTF9HT	-	Date last took Heat treated F9 (SAS)

A person's exposure categories are set as:

'H'	Ever HIGH = person's exposure is high
'M'	Else Ever MEDIUM = person's exposure is medium
'L'	Else Ever LOW = person's exposure is low
'U'	Else ever UNKNOWN = person's exposure is unknown
'N'	Else ever NONE = person's exposure is none

#### **Defaults set on the exposures**

\*\*\* To decrease the amount of 'unknown' codes, they were replaced with 'NONE' under the following circumstances:

Subjects with Hemophilia 'B' had unknown Factor VIII codes (both heat treated and non heat

treated) set to NONE.

Subjects with Hemophilia 'A' (NO INHIBITOR) had unknown Factor IX codes (both heat treated and non heat treated) set to NONE.

### Updating the File

Run from &NTT1SWY.CS5.F8F9.SAS(SUMFILE)

### F8/F9 Yearly Dose Flag File

NTT1SWY.CS5.DOSE.BYYEAR.SASDS" UNIT=FILE

This file contains F8 and F9 dosage flags for all years we have data. Subjects with no dosage information are not included on the file. For the time being, the past history flags of 1989, 1990 and 1991 are not used. When there is data from multiple files for the same year the hierarchy is: Believe Current Evaluation, then Past History, then Abstract. Only the 'past 12 month' questions from the abstract are used. All flags are coded as:

1, 2	= 'High'
3	= 'Medium'
4	= 'Low'
5	= 'None'
6	= 'Unknown'

### Variables

F8HT83 -- F8HT95	-	Factor VIII Heat Treated Dose 1983-1995
F9HT83 -- F9HT95	-	Factor IX Heat Treated Dose 1983-1995
F8NT78 -- F8NT86	-	Factor VIII Non Heat Treated Dose 1978-1986
F9NT78 -- F9NT86	-	Factor IX Non Heat Treated Dose 1978-1986
SUBJ_ID	-	Subject ID

Create Program: SAVED AS NTT1SWY.CS5.F8F9.SAS(CREBYR)

### Input Files:

NQG1JK2.HIV.HEMEDIT.SASFILE.DATA(CURREVAL, HISTORY)  
NTT1SWY.CS5.ABSTRACT.F8F9.SASDS(.ABSTRACT) \*\*

NTT1SWY.CS5.RTI.ORI.VARS.SASDS(.HEMOPH)

\*\* The abstract file flags on CS5.ABSTRACT.F8F9.SASDS(.ABSTRACT) were created by a separate program: NTT1SWY.CS5.F8F9.SAS(CREABSSS)

Since the abstract file is no longer used, this file should never need to be updated

The variables on this file are coded using the same format

ABF8HT83 -- ABF8HT88 Factor VIII Heat Treated 1983-1988

ABF9HT83 -- ABF9HT88 Factor IX Heat Treated 1983-1988

ABF8NT82 -- ABF8NT86 Factor VIII Non Heat Treated 1982-1986

ABF9NT82 -- ABF9NT86 Factor IX Non Heat Treated 1982-1986

## **T-Cell Results**

### **Overview:**

T-Cell data on the hemophiliacs has historically come in many different forms. The bulk of the data on the subjects from the Hershey, Pa. center (study\_id 'HEA') has come directly from the centers and were made into keyed files. Some additional Hershey data has been sent on floppy disks.

RTI has recently begun including T-cell data on the hemophiliac abstract forms. This for the most part contains results for the same T-cell tests that have come from Hershey (CD4 and CD8). Like the Hershey data, it also contains some CBC results. The CBC results, particularly the Lymphocyte counts, are needed to calculate the number of t-cells on samples for which we are only given percent.

In those cases the number of CD4 cells, for example, is the total number of lymphs times the percent of CD4. In some instances the lymph count is not percent per say. This can be calculated if the percent of lymphs and the white blood count (WBC) are present. the lymph count in that case is the percent of lymphs time the WBC.

The hemophiliac centers are doing less T-cell testing of their own. In order that the results are more standardized, most new samples are being sent to the Braton FACS lab. The results of these tests are included in the main HIV4 T-cell file. This file is much more complete, in terms of the number of different tests per sample that are run. The results of all these different sources are combined into the file NTT1SWY.CS5.TCELLS.LAB.SASDS.

### **Complete Hemophilia T-Cell File**

**NTT1SWY.CS5.TCELLS.LAB.SASDS(.TCELLS)**

\*\*\* One recode per sample, sorted by SAMPL\_ID DRAW.

PROGRAM: UAT1PIV.HIV4.FACS.SOURCE(CREATEH)

INPUT FILES: 1 - UAT1PIV.HIV4.FACS.SAMP.SD  
 2 - NTT1SWY.CS5.HEA86.SASDS(.FACS)  
 3 - NQG1JK2.HIV.HEMEDIT.SASFILE.DATA (.ABSTRACT,.CBC)  
 4 - NTT1SWY.CS5.RTI.TCELL.SASDS(.FACS)  
 5 - NTT1SWY.CS5.INHIB.LAB.SSD(.LAB)  
 6 - NTT1SWY.CS5.ANALYSIS.SASDS(.HEMOPH)

## VARIABLES

BILIRUBN - Bilirubin  
 # 1.4 = Normal  
 1.5-1.9 = Increased, < 2 mg/dl  
 \$ 2.0 = Increased, \$ 2 mg/dl

BIL\_ELEV - Bilirubin elevation level  
 1 = Normal  
 2 = Increased, < 2 mg/dl  
 3 = Increased, \$ 2 mg/dl

CBC\_JOIN - Date the CBC data entered file

DRAW - Draw date (SAS)

HCT - Hematocrit

HGB - Hemoglobin

LYMPH - Lymphocyte count

NUMT4 - CD4 count

NUMT8 - CD8 count

PC\_T4 - Percent (of lymphocytes) are CD4 cells

PC\_T8 - Percent (of lymphocytes) are CD8 cells

PLTCNT - Platelet count

RBC - Red blood count

SUBJ\_ID - Subject Id

TC\_JOIN - Date the Tcell data entered file

T4T8RAT - Ratio of CD4/CD8

WBC - White blood count

Transamenase Variables

AST	-	AST Transaminase elevation actual value
AST_ELEV	-	AST Transaminase elevation level (AST/AST normal range) # 2.0 = Normal > 2 = Elevation
ALT	-	ALT Transaminase elevation actual value
ALT_ELEV	-	ALT Transaminase elevation level (ALT/ALT normal range) # 2.0 = Normal > 2 = Elevation
SGGT	-	SGGT Transaminase elevation actual value
TRANSAM	-	Overall Transaminase elevation 1 = Normal 2 = Elevation, one episode 3 = Elevation, two or more episodes
HTRANSDT	-	Earliest history date of Transaminase elevation

The program reads the Hershey and RTI files. They both contain the percent and number of T-cells. Therefore, after determining that there are no draw dates in common, defined as within seven days of each other, they are combined without additional calculations.

The order of combining files is important, as data from some files are trusted more than from other files. First, the SAMP.SD file is read in. This file contains percentages of T-cells only. The variable names are standardized to the names of the previously read files. For example, with draw date of 1984, then the percent CD4 variable PC\_T4\_84 becomes PC\_T4. This is done for all years from 1984 through the present. These records then overwrite the Hershey and RTI TCELL records.

Second, in order to add CBC data, RTI's CBC file is merged with, and overwrites, RTI's Abstract file. This data is then merged onto the above T-cell file. The T-cell counts are then calculated as described in the overview.

Next, the program then examines the new completed file for records which have very close draw dates. These records are printed out for evaluation.

Finally, the program collapses the information contained the close draw date records into one record, with as few missing values as possible. These records are then added back into the T-cell file and the file is complete.

The printouts of this program should ALWAYS BE READ. At every merge, records with suspiciously close draw dates are printed out. The reason for suspicion is that the hemophiliac centers tend to send the same information more than once with close but not exact draw dates. Every effort is being made to force standardization of these dates.

\*\*\* One time creation - need not be run or considered again \*\*\*

HERSHEY FILE: NTT1SWY.CS5.HEA86.SASDS(.FACS)

PROGRAM: NTT1SWY.CS5.FACS.OLD.SAS(CREATE)

INPUT: 1 NTT1SWY.CS5.HERSHEY.MAY86.FACS.SASDS(.FACS)  
2 - NTT1SWY.CS5.HERSHEY.MARCH86.FACS.SASDS(.FACS)

This program combines the keyed data received from Hershey in March and May 1986. It puts out one record per sample and contains both percents and numbers of T-cells as well as CBC data.

This program should never need to be run again, since there will be no additions to the March and May data.

ADD-TO PROGRAM: NTT1SWY.CS5.FACS.OLD.SAS(ADDTO)

INPUT: 1 - NTT1SWY.CS5.HERSHEY.ADDTL.TCELL.L67  
2 - NTT1SWY.CS5.HERSHEY.ADDTL.OCT87.L67

This program formats additional data received from Hershey on floppy disk consistent with the previously received data and updates the Hershey file.

As new additional data is received from Hershey, this program can be used to update the file. The input files named above are the files which have arrived so far. They are only input files one time, since after the update their data is part of the NTT1SWY.CS5.HEA86.SASDS(.FACS) file.

FIX PROGRAM: NTT1SWY.CS5.FACS.OLD.LIB(FIX)

INPUT FILE: NTT1SWY.CS5.HEA86.SASDS(.FACS)

this program "fixes" PC\_T4 and PC\_T8 values that are not in the correct format. For example, a value of .30 Would be changed to 30. Two subjects' values had to be hardcoded. This program should not be run again, since PC\_T4 and PC\_T8 values would then all be multiplied by one hundred. This program was a one time run.

### Laboratory Files

**UAT1PIV.HIV4.FACS.SASDS(.FACS)** One record per subject.

**UAT1PIV.HIV4.FACS.SAMP.SD(.FACS)** One record per sample.

CREATE PROGRAM: UAT1PIV.HIV4.FACS.SOURCE(CREATE1)

INPUT FILES 1 - UAT1PIV.HIV4.FACS84.SASDS(.FACS)  
 2 - UBJ1WW.FS25.SASDS(.FACS) ON CAT  
 3 - NMG1PGY.FSBCK.SASDS(.BACKGRND) ON CAT

This program combines the results of all samples tested at the Braton (FACS) and now the Frederick Facility of Dean Mann's lab. The FACS lab was not in complete operation for nearly a year in 1984-1985. During that time testing was proceeding, but the results were not able to be electronically transmitted to ORI. The HIV4.FACS84.SASDS file is the result of keyed data from that time. It is read into the program and used if those samples have not been retested. If they have been retested, the newer test is used. The FS25 file (maintained by IMS) contains all FACS data transmitted electronically. This file contains one record per sample/stain/test. It is merged with the Background file to pick up the subject and study ID'S. The program then reads the results of the markers and assigns them to the T-cell variables with the names on CS5's other T-cell files. Two output files are produced. The first output file contains one record per subject, with the results of the last test per sample/stain/year. The second contains one record per sample with the results of the last test per sample/stain. Both files contain percents of T-cell only.

### **RTI File**

OUTPUT FILE: NTT1SWY.CS5.RTI.TCELL.SASDS(.FACS)  
 CREATE PGM: UAT1PIV.HIV4.FACS.SOURCE(RTIDATA)  
 INPUT FILES: NQG1JK2.HIV.HEMEDIT.SASFILE.DATA(.TCELLS)

This program reads the RTI file that contains T-cell and CBC data. It writes out one record per subject/draw date with percent T-cell, calculated number of T-cells and lymphocyte count

### **Updating the Files**

The T-Cell files should be updated every Friday, and whenever IMS notifies us that their file has been updated. To update the files issue the following command:

RUN FROM UAT1PIV.HIV4.FACS.SOURCE(JCL) on CAT.

This JCL stream updates all the hemophiliac files documented here, as well as the DCG analysis file UAT1PIV.HIV4.ANALYSIS.SASDS.

**Rolling Means T-Cell File****NTT1SWY.CS5.ROLLMEAN.SSD(.TCELLS)**

\*\*\* One record per 2 samples, sorted by SAMPL\_ID MIDDRAW.

**Create Program:** NTT1SWY.CS5.TCELL.MISC.SAS(ROLLMEAN)**Input Files:** NTT1SWY.CS5.TCELLS.LAB.SASDS(.TCELLS)

The program creates a file of 'rolling mean' CD4 and percent CD4 values and dates. For each 2 TCELL records the following is determined and written out: The mean CD4 and mean Percent CD4 values; The mid date between the two draw dates; the first of the 2 values that are used to calculate the means; the last of the 2 values that are used to calculate the means; and the first and last dates that are used to calculate the mid date between the two. Only subjects with at least two values are included.

**Variables:**

SUBJ_ID	-	Subject Id Number
DRAW_F	-	First of draws that calculate mid draw (SAS)
MIDDRAW	-	Date midway between first and last draw (SAS)
DRAW_L	-	Last of draws that calculate mid draw (SAS)
CD4_F-		First CD4 used to calculate mean
MEANCD4	-	Mean of each two CD4 values
CD4_L	-	Last CD4 used to calculate mean
CD8_F-		First CD8 used to calculate mean
MEANCD8	-	Mean of each two CD8 values
CD8_L	-	Last CD8 used to calculate mean
PCT4_F	-	First percent CD4 used to calculate mean
MEANPCT4	-	Mean of each two percent CD4 values
PCT4_L	-	Last percent CD4 used to calculate mean
PCT8_F	-	First percent CD8 used to calculate mean
MEANPCT8	-	Mean of each two percent CD8 values
PCT8_L	-	Last percent CD8 used to calculate mean

**'CAT' Date File****NTT1SWY.CS5.CATDT.SSD(.HEMOPH)**

\*\*\* One record Subject sorted by SUBJ\_ID.

**Create Program:** NTT1SWY.CS5.TCELL.MISC.SAS(CRECATDT)

**Input Files:** NTT1SWY.CS5.ROLLMEAN.SSD(.TCELLS)  
 NTT1SWY.CS5.PHILR.SERODATE(.PHILDATE)  
 NTT1SWY.CS5.ANALYSIS.SASDS(.HEMOPH)

The program creates a file of CD4 Category Dates indicating the dates that the subject dropped into a new CD4 category. The categories are:

CATEGORY	CD4	RANGE
1	0	- 49
2	50	- 99
3	100	- 199
4	200	- 499
5	500	- High

**RULES:**

- Only includes HIV Positive Hemophiliacs
- If both draw dates making up the mean date are before seroconversion the records are dropped.
- If the last date making up the mean date is after seroconversion, but the mean date is before seroconversion, the mean date is set to then seroconversion date.
- All values are based on the ROLLING MEAN values
- Once a subject is in a low category by virtue of a low enough count, that subject can never SUBSEQUENTLY get assigned to a higher category even he has a subsequent higher CD4 count.

**Variables:**

SUBJ_ID	-	Subject Id Number
CAT1DT	-	Date subject first entered category 1
CAT2DT	-	Date subject first entered category 2
CAT3DT	-	Date subject first entered category 3
CAT4DT	-	Date subject first entered category 4

CAT5DT - Date subject first entered category 5

## P-24 Antigen File

### NTT1SWY.CS5.ANTIGEN.RESULTS.SASDS(.ANT)

\*\*\* One record per sample, subject's status on every record, sorted by SUBJ\_ID DRAWDT.

#### Edits

The following edits have been included in the update program for all records received after September 1992. The following conditions will force the record into an error file and report and will not affect the update

- The record does not contain EITHER a confirmation or both a ratio and real field
- The ratio is greater than the cutoff (REAL2) and the confirmation is '-'
- The ratio is less than or equal to the cutoff and the confirmation is '+'
- The ratio is less than 0.1 and the confirmation is '+'

#### Variables

This file contains both sample and subject P-24 Antigen information. It should be updated at least once every month, or at any time it is known that antigen data has arrived. Antigen data consists of assay types 'AS', 'AC' and 'AQ'. The antigen file contains the following variables:

ANTGN	-	The SAMPLE antigen status (+,-,/)
ANTPOS	-	SAS Date of first positive antigen, on every record
ANTNEG	-	SAS Date of last negative antigen, on every record
DRAWDT	-	SAS Draw Date
STATUS	-	SUBJECT'S Antigen Status (+,-,/), on every record
SAMPL_ID	-	Sample Id
SUBJ_ID	-	Subject Id

#### Error Reports

Two error reports will be generated. The first will list all records that failed the edits described above. These records were not included in the update and did not affect the output file. The second

error report lists all the records on subjects that have a negative sample occurring after a positive sample. The latest 'receive date' per person is included to facilitate the reading of the report for new cases. These records ARE in the data base.

### **Input Files**

NMG1PGY.FSRSLT.SASDS(.H3) - Complete HIV test results

### **Updating the File**

Run from NTT1SWY.CS5.ANTIGEN.SAS(CREATE)

## **Neopterin and Beta-2 Microglobulin**

### **NTT1SWY.NEOPT.SASDS(.NEOPT)**

\*\*\* One record per sample, sorted by SAMPL\_ID

#### **Variables**

This file is used by reports and by most TDC files. The input data is maintained by RTI. It **does not** contain subject or study information and **does** contain non-hemophiliacs (gays), therefore it must be merged by sample id with the VES Background file to select the hemophiliacs. It contains the following variables:

NEOPTER	-	Neopterin value
MOD	-	Preface '<' or '>' to neopterin value
B2	-	Beta-2 Microglobulin value
BATCH	-	Batch Number

### **Input Files**

NVC1JA5.NEOPTER.SASDS(.NEOPTER, .NEOALL)

### **Updating the File**

Run from NTT1SWY.CS5.NEOP.SAS(CREATE) on Cat

RTI will contact ARC when new Neopterin data is ready. In the past they merely updated the .NEOALL file. One time however, they created a new member .NEOPTER. Find out where the new data is. If they added it to one of the existing files, the CREATE program can be run as is. If

they added a new member, the CREATE program needs to be modified to include it.

### Drug File (AZT and Others)

\*\*\* One record per subject, sorted by SUBJ\_ID.

FILE NAME: NTT1SWY.CS5.DRUGS.SASDS(.DRUGS)

#### Variables

This file contains subjects who have used AZT and other drugs and indicates the drug treatments started. If no date was indicated, a date was derived from the date on the form and the number of weeks they were on the drug. There are some subjects which will not have a start date or a derived date because the forms indicated they were no longer using a drug; therefore, a date could not be derived.

\*\*\* Derived Flags    0 indicates date was on the form  
                           1 indicates date was derived

Variables	Description
ACTDERV	Acyclovir(oral or IV) Date Derived
ACTDT	Date Started ACYCLOVIR(ORAL OR IV)
ALFDERV	ALPHA INTERFERON Date Derived
ALFDT	Date Started ALPHA INTERFERON
AMPLDERV	AMPLIGEN Date Derived
AMPLDT	Date Started AMPLIGEN
AZTDERV	AZT Date Derived
AZTDT	Date Started AZT
A721DERV	AL 721 Date Derived
A721DT	Date Started AL 721
DDCDERV	DDC (DIDEOXYCYTIDINE) DERIVED
DDCDT	Date Started DDC (DIDEOXYCYTIDINE)
DDIDERV	DDI Date Derived
DDIDT	Date Started DDI
PAIRDERV	AERO PENTAMIDINE Date Derived
PAIRDT	Date Started AERO PENTAMIDINE
PINJDERV	PENTAMIDINE - IV Date Derived
PINJDT	Date Started PENTAMIDINE - IV
RIBADERV	IS RIBAVIRIN Date Derived
RIBADT	Date Started RIBAVIRIN
SUBJ_ID	Subject Identification

SULFDERV TRIMETHOPRIM-SULFA (Bactrim) Date Derived  
 SULFDT Date Started TRIMETHOPRIM-SULFA

### **Input Files**

NQG1JK2.HIV.HEMEDIT.SASFILE.DATA(.CUREVAL)  
 NQG1JK2.HIV.HEMEDIT.SASFILE.DATA(.HISTORY)

### **Updating the File**

Run NTT1SWY.CS5.DRUGS.SAS(CREATE)

## **Hepatitis B Status File**

NTT1SWY.CS5.HBV.SASDS(.HBV)

\*\*\* One record per subject, HBV\_DT. Sorted by subject, HBV\_DT

### **Input Files**

NQG1JK2.OLDINHIB.RAWDATA(ENT & EDITFUP)  
 NMG1PGY.FSBCK.SASDS  
 NQG1ECD.HEMO.CLINFILE.DATA  
 NTT1SWY.CS5.HBSAB.L15  
 NTT1SWY.CS5.HBSAG.L15  
 NTT1SWY.CS5.HBVAX.L8  
 UAT1PIV.HIV9.HEP.SASDS  
 NQG1RAD.HEM.FORM11.DATA

### **Updating the File**

NTT1SWY.CS5.HBV.SAS(CRESAS)

This data set is sorted by SUBJ\_ID, HBV\_DT and contains one record per subject, HBV\_DT. The records were collapsed so the HBSAG, HBSAB, HBCAB, and HBVAC values all correspond to the HBV\_DT on that record. Keep in mind that all dates are estimates, and in most cases cannot be linked to a specific sample. All dates are SAS dates.

This data set was created by incorporating all hepatitis B data from all available sources. Because so many sources were involved, an order of source reliability or precedence was developed to ensure that the best or most current data was used. SAS Updates were used instead of SAS Merges so data

values would not be overwritten by missing values. The update order (from least to most reliable data) is as follows: inhibitor file, Hershey supplemental files, clinical file, Gordon Bray's form 11 file, and the results file. Form 11 data is trusted over clinical data because it is believed that the data in form 11 has been verified by Gordon Bray.

Depending on the data source, the HBV\_DT is derived from the DT\_DRAWN, INFODT, HBVACDT, HBEVALDT etc. In cases where there were multiple records with duplicate SUBJ\_ID, HBV\_DT's, all the records were looked at then positive values were kept over negative values, and negative values were kept over missing values.

Currently, the only source of dates associated with hepatitis B vaccinations is the clinical file and only the month and year of vaccination were collected. Therefore, to create complete vaccinations dates, it was necessary to insert a value of '15' for day. In cases where the month was missing, both month and day were inserted using a value of '0701'. The dates associated with HBVAC was calculated differently depending on whether a person indicated being vaccinated on that date or not.

On records indicating a vaccination, only the HBVACDT is considered a reliable estimate to be used as the HBV\_DT and may result in missing HBV\_DT's. However, on records which the person responded no vaccination at that time, the HBV\_DT may come from the HBEVALDT or if that is missing the INFODT.

There are several variables which were created after looking at all the HBV data available for each subject. These variables are carried on each record and include: dates of last positive and negative HBSAG, HBSAB, HBCAB; dates of first and last vaccine; ever vaccinated flag; HBSAB acquired naturally or through vaccine flag; form 11 chronic hepatitis flag; and three calculated hepatitis variables - HBV\_STAT, INFECTED, and IMMUNE variables. These three summary variables were created, using hierarchical algorithms developed by Jim Goedert and Laura Diamondstone, to categorize each persons hepatitis B status.

**\*\*\*\* These variables that do not relate to the record draw date are the same on every record: The first/last date variables, HBV Status variables (including immine infected), Ever and Nator Vaccination flags, and Gordon Bray's chronic flag and onset date (CHRNBRAY and CHRNBRDT).**

### Variables

SUBJ_ID	-	Subject ID
HBV_DT	-	Date associated with hepatitis B variables - may originate from DT_DRAWN, INFODT, HBEVALDT, HBVACDT etc. depending on the data source
HBSAG	-	Hepatitis B surface antigen test - flag
FSTPSSAG	-	Date of first positive surface antigen test

FSTNGSAG	-	Date of first negative surface antigen test
LSTPSSAG	-	Date of last positive surface antigen test
LSTNGSAG	-	Date of last negative surface antigen test
HBSAB	-	Hepatitis B surface antibody test - flag
FSTPSSAB	-	Date of first positive surface antibody test
FSTNGSAB	-	Date of first negative surface antibody test
LSTPSSAB	-	Date of last positive surface antibody test
LSTNGSAB	-	Date of last negative surface antibody test
HBCAB	-	Hepatitis B core antibody test - flag
FSTPSCAB	-	Date of first positive core antibody test
FSTNGCAB	-	Date of first negative core antibody test
LSTPSCAB	-	Date of last positive core antibody test
LSTNGCAB	-	Date of last negative core antibody test

The HBSAG, HBSAB, HBCAB flag variables can have the follow values:

- 0 - Tested negative
- 1 - Tested positive
- . - Missing

HBVAC		Hepatitis B vaccination - flag Relates to record draw date
	0 -	Did not receive hepatitis B vaccine
	1 -	Received hepatitis B vaccine
	. -	Missing

EVERVAC	-	EVER vaccinated for Hepatitis B - flag Same on every record
	0 -	Did not receive hepatitis B vaccine
	1 -	Received hepatitis B vaccine
	. -	Missing

FRSTVAC - Date of first hepatitis B vaccination

LASTVAC - Date of last hepatitis B vaccination

NATORVAC - Hepatitis surface antibody acquired naturally or through vaccination - flag

	Note: only subjects from Part 7 of Gordon Form 11 file will have this data	Bray's
	N - Acquired HBSAB naturally V - Acquired HBSAB through vaccine . - Missing	
CHRNBRAY	- Chronic Hepatitis B - flag Note: only subjects from Part 6 of Gordon Form 11 file will have this data	Bray's
	1 - Chronic hepatitis . - Missing	
CHRNBRDT	- Date of onset of HBV infection Note: only subjects from Part 6 of Gordon Bray's Form 11 file will have this data	
HBV_STAT	- Hepatitis B summary variable - contains subjects calculated hepatitis status	
	CHRONIC (at least two consecutive HBSAG+ tests at least six months apart or CHRNBRAY = 1)	
	HBSAG+ (ever)	
	HBCAB+ (ever)	
	HBVAC+ (ever)	
	HBSAB+ (ever)	
	HBCAB- (ever)	
	HBSAB- (ever)	
	UNKNOWN	

### **Hepatitis B 'Immune' and 'Infected' Codes**

Hepatitis B Immunity Status (in order of source code hierarchy)

- 1 = 'Last HBsAb +, all + HBsAb subsequent to last HB vax'
- 2 = 'All HBsAb values subsequent to last HB vax missing'
- 3 = 'Last HBsAb +, vax missing'
- 4 = 'Last HBsAb +, vax -'
- 12 = 'Last HBsAB +, vax +'
- 5 = 'All HBsAb values missing, vax missing'
- 6 = 'All HBsAb values missing, vax -'

- 7 = 'Last HBsAb - and subsequent to last HB vax, vax +'
- 8 = 'Last non-missing HBsAb -, vax missing'
- 9 = 'Last non-missing HBsAb -, vax -'
- 10 = 'HB vax positive, missing vax date- Forever unknown'
- 11 = 'Other'

#### Hepatitis B Infection Status

- 1 = 'HB status=chronic, last HBsAg +'
- 2 = 'HB status=chronic, last HBsAg -'
- 3 = 'Last HBsAg +, no previous - HBsAg'
- 4 = 'Last HBsAg +, a previous - HBsAg'
- 5 = 'HB infected but not chronic/ recently antigeniac'
- 6 = 'HBsAb ever + and no vax or HBsAb + prior to vax'
- 7 = 'HB vax + but date missing - Forever unknown'
- 8 = 'HB vax. + and HBsAg - or missing prior to vax'
- 9 = 'HB status= HBCAb -'
- 10 = 'Other'

#### Hepatitis B Combined Infection and Immunity Status

- 1 = 'HBV status= Now Chronic, Immunity= HBsAb+'
- 2 = 'HBV status= Now Chronic, Immunity= HBsAb-'
- 3 = 'HBV status= Was Chronic, Immunity= VAX- HBsAb+'
- 4 = 'HBV status= Now HBsAg+, Immunity= VAX- HBsAb-'
- 5 = 'HBV status= New HBsAg+, Immunity= VAX- HBsAb-'
- 6 = 'HBV status= Positive, Immunity= VAX- HBsAb+'
- 7 = 'HBV status= Positive, Immunity= VAX- HBsAb-'
- 8 = 'HBV status= Positive, Immunity= VAX+ HBsAb+'
- 9 = 'HBV status= Positive, Immunity= VAX+ HBsAb-'
- 10 = 'HBV status= Negative, Immunity= VAX+ HBsAb+'
- 11 = 'HBV status= Negative, Immunity= VAX+ HBsAb-'
- 12 = 'HBV status= Negative, Immunity= VAX- HBsAb-/?'
- 13 = 'Other'

#### Interferon File

NTT1SWY.CS5.HEMOPH.INTERFRN.SASDS (.INT)

**\*\* One record per sample, Sorted by SAMPL\_ID \*\***

### **Variables**

This file contains Interferon codes and test results for Hemophilia, DCG, GAM, and a few other studies. It represents results from test years 1983-1989. Where test dates were previously missing, approximate dates have been hard-coded into the file. The file also includes subject, study, and sample ids for data selection purposes.

The classification codes were calculated from the average of two well values, or the actual well value if only one well was tested. The code is assigned according to the following groups of values:

<u>Average</u>	<u>Code</u>
0-3.99	=> 0
4-7.99	=> 1
8-11.99	=> 2
12-19.99	=> 3
20-49.99	=> 4
>= 50 =>	5

### **Input Files**

NTT1SWY.CS5.INTERFRN.DATA.L17(INTnn)  
 NTT1SWY.CS5.INTERFRN.DATA.L21(INT2)  
 NTT1SWY.CS5.INTERFRN.DATA.L13  
 UAT1FMY.HIV4.INTERFRN.L20

### **Updating the File**

As new interferon data becomes available, member CREATE of the PDS NTT1SWY.CS5.INTERFRN.LIB.SAS must be altered in order to accommodate the new data set.

### **Sibling Study**

**Analysis File**            NTT1SWY.CS5.BYK93.SASDS

**\*\*\* One record per subject, sorted by SUBJ\_ID**

SAS version 5 data set containing all HIV positive hemophilacs. It contains the variables

used in most of the proportional hazard ANOVA and Kaplan Meier analysis programs. It must be a version 5 file in order to run PROC PHGLM.

### Variables

SUBJ_ID	-	Subject id number
SIBPAIR	-	Sibling Pair number (Sibs only)
SHARE	-	Haplotypes shared (Sibs only 0,1 or 2)
AIDS87	-	AIDS flag
AIDS87DT	-	AIDS dx date
AIDS93	-	AIDS flag 1993 definitian
AIDS93DT	-	AIDS93 dx date
SERODT	-	Seroconversion date
SEROAGE	-	Age at seroconverion
LKA	-	Last known alive

### To update the file

Run from NTT1SWY.CS5.BYK93.SAS(CREANAL)

### AIDS Analysis Programs

This command procedure runs the basic AIDS proportional hazard and ANOVA analysis. The user is prompted for the cutoff desired by censoring.

Execute from the following command procedure;  
NTT1SWY.CS5.BYK93.PHMAY93(AIDSCP)

This program runs a Kaplan Meier using the haplotype file for defining groups

### CD4 Analysis File

NTT1SWY.CS5.CD4PHGLM.SSD

This is a file to be used for the CD4 PHGLM runs. It contains flags showing if the subject is eligible for four different events: Percent CD4 < 20, Percent CD4 < 30, CD4 under 200 and CD4 under 300. To be eligible a subject must:

- have a CD4 count (either percent or number depending on the event).
- the first value cannot be under the chosen cutoff

In addition to the eligibililty flags the file contains:

- A default end date - the last percent draw date and last number draw date
- The draw date of the first value under each threshold

Variables

GETIN15	-	Eligible for percent < 15 run
GETIN20	-	Eligible for percent < 20 run
GETIN200	-	Eligible for CD4 < 200 run
GETIN300	-	Eligible for CD4 < 300 run
LASTPCT	-	Last percent CD4 draw date
LASTNUM	-	Last absolute CD4 draw date
NUMPCT	-	Number of percent CD4 values
NUMCD4	-	Number of absolute CD4 values
UNDER15	-	First draw date with CD4% under 15
UNDER20	-	First draw date with CD4% under 20
UNDER200	-	First draw date with CD4 under 200
UNDER300	-	First draw date with CD4 under 300

**To update this file**

Run from NTT1SWY.CS5.BYK93.PHGLM(CRECD4)

**CD4 Analysis Programs**

Execute fr NTT1SWY.CS5.BYK93.PHMAY93(STRATCP)

**\*\*\* When performing CD4 analysis, the rules need to be clarified with Barbara. For many runs, both siblings are dropped if either is ineligible.**

**PCR Data**

Transactions

**NTT1SWY.CS5.PCR.L30**

This PDS contains the hemophilia PCR transactions received thus far.

File Format

Variable	Cols	Type/Length	Description
----------	------	-------------	-------------

Sample Id	1-7	Char 7 Sample Id
Vial #	9-11	Num 3 Vial Id
Material	13	Char 1 Material Tested
Result	15	Char 1 Score (+, -, /)
Virus	17	Char 1 Virus Sought
Lab	19-20	Char 2 Testing Laboratory
Dilution	22-23	Num 2 Dilution ***
Date Rcvd	25-30	Char 6 Date added to database

\*\*\* The dilution indicates the lowest dilution at which the sample tests positive. A dilution value of 0 or greater is considered positive, 1 indicates positive at a dilution of 10, 2 = 100, 3 = 1000 etc. A dilution of 'NA' indicates the sample is negative or indeterminate.

### SAS File

NTT1SWY.CS5.PCR.SASDS

The file is sorted by subject Id number and contains one record per sample.

### To Update the File

Run NTT1SWY.CS5.PCR.SAS(CRESAS) on CAT

Update the PCR transaction file, date stamping the records with the current date. The update program will add the background information, and write out the SAS data set.

**Backup Tapes**

There are two backup tapes for the hemophilia files. Backups will be made each Friday on alternating tapes. The access for these tapes is NTT1SWY.

Tape Numbers:

107520  
109579

Files to be backed up:

NTT1SWY.CS5.ANALYSIS.SASDS(.HEMOPH)  
NTT1SWY.CS5.RTI.ORI.VARS.SASDS(.HEMOPH)  
NTT1SWY.CS5.CORE.SASDS(.CORE)  
NTT1SWY.CS5.TCELLS.LAB.SASDS(.TCELLS)  
NTT1SWY.CS5.NOV12.SCRN

**Samples tested for HCV bDNA, HCV Genotyping, HCV Serotyping, and/or HIV Monitor****NTT1SWY.CS5.SAMPLES.TESTED.SSD(HCVHIV)**

This file includes all MHCS and DCG samples that were ever sent to be tested for HIV Monitor, HCV BDNA, HCV Genotyping and/or HCV Serotyping. Samples are included in the file even if test results have not yet been received.

The file has one record per sample and contains the following variables:

- 1- Variables indicating whether a sample was sent to be tested for any of the assays mentioned above:

HCBDNAST	Sample sent for HCV BDNA
HIVMNTST	Sample sent for HIV Monitor
HCGNTST	Sample sent for HCV Genotyping
HCSEOST	Sample sent for HCV Serotyping

- 2- Variables indicating whether results from a list of selected laboratories are available :

HCBDNADB	HCV BDNA results from Di Bisceglie
HCBDNASA	HCV BDNA results from SAIC
HIVMNTDB	HIV Monitor results from Di Bisceglie
HIVMNTPA	HIV Monitor results from MHCS form 04
HIVMNTSA	HIV Monitor results from SAIC

These variables are created based on data in the VES PCR file. If results from a laboratory other than the ones listed above are to be incorporated into this file, the program that creates it will have to be modified to include the new code(s).

**Input files:**

NYU1IBW.BSI.BIOTECH.ACTIVE.REQUESTS  
NTT1SWY.CS5.HCVGENOS.SSD(HCV)  
UAT1PIV.HCVSER.SSD(HCVSER)  
UAT1PIV.HIV9.PCR.SSD(PCR)