



FONTAN CROSS-SECTIONAL STUDY PUBLIC USE DATASET

ABOUT THE STUDY

The NHLBI Fontan Cross-Sectional Study was conducted by the Pediatric Heart Network (PHN) at 7 centers in 2003-2004. The PHN screened a total 1,078 patients and enrolled 546 children aged 6 to 18 years old. Study measurements were specified to be made in a 3-month window following enrollment. The primary aim of the study was to examine associations between functional health status (measured by parent- and child-report questionnaires) and ventricular state and performance (measured by 2D and Doppler echocardiography, maximal exercise testing, ECG, cardiac MRI and resting B-type natriuretic peptide [BNP] concentration). Core laboratories were used for interpretation of echocardiograms and MRIs, and to perform the BNP assay. Test completion rates in the enrolled cohort ranged from 536 echocardiograms and 511 parent-report questionnaires to 161 MRIs that were acceptable for analysis. The study design has been summarized in Sleeper et al. (AHJ 2006) and in the study protocol (available to users with approved logins). Tables 1 and 2 provide selected subgroup sizes for the Full Protocol Fontan Study cohort. A great deal of additional information on available sample sizes for measurements may be found in Anderson et al. (JACC 2008; http://www.pediatricheartnetwork.org/publications/Fontan_Mainresultspaper.pdf), as well the published articles on specialized topics (see posted Bibliography at <insert link>).

Table 1. Enrollment Age Distribution of the Fontan Study Cohort by Age at Fontan Surgery

Age at Fontan, yr	Age at enrollment, yr				Total
	[6,9)	[9,11)	[11,14)	[14,19)	
[0,2)	27	28	38	20	113
[2,3)	59	44	46	42	191
[3,4)	29	27	20	28	104
≥4	23	21	30	64	138
Total	138	120	134	154	546

FONTAN CROSS-SECTIONAL STUDY PUBLIC USE DATASET

Table 2. Distribution of Cardiac Anatomic Diagnosis and Ventricular Morphology

Ventricular Morphology (vent_type)	Pre-Fontan Cardiac Anatomic Diagnosis (preop_dxcat)								Total
	A1.01: SV, DILV	A1.02: SV, DIRV	A1.03: SV, MA	A1.04: SV, TA	A1.05: SV, Unbalanced AVCD	A1.06: SV, Heterotaxia syndrome	A2: HLHS	A3: Other functional SV not fitting any other categories	
LV	79	0	0	119	9	2	0	56	265
RV	1	8	31	0	13	16	109	6	184
MIXED	0	0	0	0	0	24	3	70	97
Total	80	8	31	119	22	42	112	132	546

SV=single ventricle; LV=left ventricular; DI=double inlet; MA= mitral atresia; TA=tricuspid atresia; HLHS=hypoplastic left heart syndrome; AVCD=Atrioventricular canal defect

STUDY DOCUMENTATION

The following datasets and descriptor files are available for download. A login and password (request access via <http://www.pediatricheartnetwork.org>) are required for download capability. The lock date used for creation of the public dataset is August 1, 2011. Privacy protection of these data is described in Appendix A.

1. Annotated study data collection forms (PDF) – These contain the SAS variable names next to each data field on the form. These form documents also include some created variables and their definitions.
2. SAS version 9.2 datasets
3. The file *fontanformats.sas7bcat* – Include this file in your program using:

```
options fmtsearch = (fmtlib.fontanformats);
```

 where *fmtlib* is specified using a *libname* statement as the path name.
4. SAS Proc Contents for each dataset (PDF)
5. Excel datasets (with variable formats applied) – These data have a .csv extension, which means that the file may also be opened either in Excel, OR in a text editor, appearing as a comma-delimited file.

STUDY RESOURCES

Resources posted on the [pediatricheartnetwork.org](http://www.pediatricheartnetwork.org) website include:

- Fontan Cross-Sectional Study bibliography (see <http://www.pediatricheartnetwork.org/<NEWLINK>>)

FONTAN CROSS-SECTIONAL STUDY PUBLIC USE DATASET

- Fontan Study Design paper and Main Summary paper (see http://www.pediatricheartnetwork.org/publications/Design%20Paper%20AHJ%209_06.pdf and http://www.pediatricheartnetwork.org/publications/Fontan_Mainresultspaper.pdf)
- Fontan Cross-Sectional Study protocol (with login access)

DATA USE POLICY

- **REQUIRED ACKNOWLEDGEMENTS:** All presentations and publications using these data must include the following statement: *“The NIH/NHLBI Pediatric Heart Network Fontan Cross-Sectional Study dataset was used in preparation of this work. Data were downloaded from https://www.pediatricheartnetwork.org/pud_login.asp?study_id=FCSD on mm/dd/yyyy.”*
- **PAPER, ABSTRACT, and PRESENTATION TITLES:** Titles may, at the authors’ discretion, mention the PHN database but should not imply that the work is from the PHN. An example of an acceptable phrase would be, “an analysis of the Pediatric Heart Network public database.” Whether or not the title makes mention of the PHN, acknowledgement should be made as described above.
- All users are requested to send a copy of published abstracts and articles to the PHN Data Coordinating Center at New England Research Institutes (PHNpubs@neriscience.com) within one month of publication. This will allow the PHN and the NHLBI to document the continued impact of this study on the field.
- The login and password provided to each user are valid for 6 months. If a user decides to complete analyses leading to more than one presentation or publication in that time period, it is requested that they notify the PHN Data Coordinating Center at New England Research Institutes of their additional analysis topics, solely for the purposes of tracking.
- The login and password to access the public dataset is provided to a single user. If a colleague would like to access the public dataset for a different analysis topic, a separate request for login and password should be submitted via the www.pediatricheartnetwork.org website.
- As an approved user, you agree that you will not attempt to establish the identities of research participants through use of this dataset.
- As an approved user, you agree to not place these data in other public locations.

TIPS ON USING THESE DATA

- Identification numbers for study subjects and study sites have been re-assigned for privacy protection.
 - subj_id:* Subject ID ranging from 1 to 1078;
 - site_id:* Ranges from 1 to 7
- The study data are contained in a large number of individual forms. These may be used jointly by merging on *subj_id*. No dataset has more than one record per subject. A single dataset of 168 of the most commonly used raw and created variables, FXS_KEYINFOPUB, is also provided. It contains 1078 records.

FONTAN CROSS-SECTIONAL STUDY PUBLIC USE DATASET

- It is important to keep in mind that some forms include records for more than the 546 Full Protocol subjects. Forms F001, F02A, and F02B contain some screening records and the functional health status forms contain records for the 60 Partial Protocol patients. The Full Protocol cohort may be selected with *consented=1* (found in FXS_KEYINFOPUB and F02B) or *disp=7* (FXS_KEYINFOPUB).
- The raw data collected are contained in the original variables (denoted by upper case variable names). Prior to analysis, these variables must have special values (typically negative numbers, see Appendix B) set to missing. Created variables (denoted by lower case variable names) already contain a SAS missing value if the measurement is unavailable.
- To select for echocardiograms and MRIs that have data acceptable for analysis, use *ACPTECHO=1* and *ACCPTMRI=1*. Unacceptable echocardiograms and MRIs have no qualitative or quantitative measurements recorded.
- Anatomy: A key grouping variable that has been used for many study analyses is ventricular morphology (left, right, mixed). This variable is determined according to cardiac anatomic diagnosis and is called *vent_type* (Form 001 and dataset FXS_KEYINFOPUB). This variable is not the same as the echo core laboratory (Form 12B) variable *vent_dom*, which does not take into account anatomic diagnoses involving reversed location of cardiac structures. All Fontan Study publications have utilized *vent_type*.
- The core laboratory echocardiographic measurement dataset Form F12B contains many created variables that are most commonly used in analysis. These variables express total ventricular size (e.g., *echoedv*) and function (e.g., *echoef*) and overall regurgitation grades (*oavvregurg*, *slvregurg*). They incorporate, where available, measures from both ventricles, and measurements from left, right, and common atrioventricular valves. Echocardiographic z-scores (e.g., *echoedv_z*, *echoef_z*) accounting for body surface area or age that provide a reference relative to normal (two-ventricle) children are also included in this dataset.
- Exercise testing was conducted in 411 subjects. This cohort is often divided into those who did and did not achieve maximal effort. All Fontan Study publications defined maximal effort as a respiratory exchange ratio (peak respiratory quotient) ≥ 1.1 . The variable on Form F10B called *MAXEFF* (Question F1) was not used. The created variable *rer_ge1_1* was used.
- Anthropometric z-scores are calculated using the 2000 CDC standard and are stored in FXS_KEYINFOPUB. The raw measurements used for z-score calculation are weight and height on Form F12B (Core Laboratory echocardiography data).
- There has been one noticeable change in the study dataset that was made after the majority of Fontan Study papers were published. The change is with regard to the Type of Fontan Procedure. The corrected data have a much lower proportion of extracardiac lateral tunnel procedures (*fontan_sgcat="B2.02.04"*) and a higher proportion of extracardiac conduits (*fontan_sgcat="B2.02.05"*). The variable *fontan_sgcat* is found in both the *fxs_keyinfopub* and *f04bpub* datasets. In the enrolled cohort of 546 subjects, the frequencies are: 72 atriopulmonary connections; 327 total cavopulmonary connection (TCPC) intracardiac lateral tunnels; 3 TCPC extracardiac lateral tunnels; 133 TCPC extracardiac conduits; and 11 other.

**FONTAN CROSS-SECTIONAL STUDY
PUBLIC USE DATASET**

ADDITIONAL ASSISTANCE

If you have questions about the study dataset that this documentation and the above resources (protocol, articles) have not answered, please email the PHN Mailbox at PHNmailbox@neriscience.com.

FONTAN CROSS-SECTIONAL STUDY PUBLIC USE DATASET

APPENDIX A

Implementation of Privacy Protection Rules for Public Use of the PHN Fontan Study Dataset

Variables that could lead to subject identification were eliminated in the public dataset. Steps included:

1. Removal of original study ID number (replaced with *subj_id*, a random consecutive numbering ranging from 1 to 1078), and removal of acrostic. Of note, no names, addresses, zip code, or medical record numbers were ever contained in the original study dataset.
2. Seven centers contributed data to the Fontan Cross-Sectional Study. A new center identifier (*site_id*), which represents a random consecutive numbering ranging from 1 to 7 was created, without formats (i.e., without center names).
3. All dates in the original datasets were removed, and replaced with "Age at event/intervention/procedure" in years (to 2 decimal places). Therefore, time intervals may be calculated by subtraction of two ages. The one exception to this convention is that the calendar year of the most recent Fontan procedure performed remains in the dataset for the 546 enrolled (Full protocol) subjects, in order for analyses to adjust as needed for era effects. This variable is called FONTAN_YEAR and is located in the Form 04B dataset and the FXS_KEYINFOPUB dataset. Provision of a year only without day or month conforms to HIPAA requirements.
4. Free (write-in) text variables remain in the public dataset. These often provide highly relevant information for interpretation of the data. However, any write-in string that referred to a specific date, a particular medical center or a particular MD was blinded or omitted.
5. Outliers for continuous variables and small group sizes for categorical variables were retained in the dataset for public use due to their importance in interpretation of the data and low likelihood of unblinding any user to a subject identity unless the user already had access to the particular medical center's data for valid reasons.

**FONTAN CROSS-SECTIONAL STUDY
PUBLIC USE DATASET**

APPENDIX B

Special Value Codes

-9 = missing

-8 = don't know/indeterminate

-7 = refused to answer

-6 = not recorded

-5 = measurement could not be reliably recorded or is not interpretable (study technically inadequate)

-4 = illegible

-2 = programmed skipped field based on results of or response to a previous question

-1 = not applicable/structure not present

-77 = Not detectable below 4.0 pg/ml (BNP concentration)