Pediatric Heart Network (PHN)

VENTRICULAR VOLUME VARIABILITY STUDY
PUBLIC USE DATASET

ABOUT THE STUDY

The NHLBI Ventricular Volume Variability (VVV) Study was conducted by the Pediatric Heart Network (PHN) at 8 centers from May 2005 to July 2007. The PHN screened a total 275 infants, children, and young adults (<22 years) with known or suspected dilated cardiomyopathy: There were 169 enrolled (131 as full protocol participants; and 38 as partial protocol participants—those who did not meet criteria for left ventricular dilation and/or dysfunction but who did not meet any exclusion criteria).

From the 169 patients enrolled, a total of 646 echocardiograms were collected and analyzed. At each visit, there was a primary and secondary echocardiogram acquisition (325 primary; and for technical reasons, 321 secondary). Of the 169 patients, at least one follow-up echocardiogram was submitted for 107. Analysis was performed by a core laboratory, as well as at the local study centers by a single designated site sonographer. The 646 echocardiograms generated 1,629 readings at the core laboratory (see Table 1). Furthermore, these readings resulted in nearly 4,887 datasets - each echocardiogram had measurements made 3 times each (3 cardiac cycles x 1,629 = 4,887).

The standardized core laboratory measurement protocol included a total of **150 measurements and calculated variables, performed for each of 3 cardiac cycles (450 measurements total)** for each set of echocardiographic images. Measurements were categorized as areas (9 variables), calculated variables derived from 2, 3 and 4 measured variables (19, 25 and 21 variables respectively), dimensions (16 variables), ECG time intervals and heart rates (7 and 9 variables respectively), integrals (1 variable), slopes (4 variables), Doppler and M-mode time intervals (22 variables) and Doppler velocities (17 variables). The dataset constructed at the local centers was comprised of 119 of the 150 measurements.

The aims of the study were:

Primary aim:

- To determine the interstudy variability of echocardiographically-derived LV end-diastolic volume z-score, mass z-score, and ejection fraction z-score in pediatric patients with DCM; more specifically, the variance at a single point in time as well as the variance of change in measurements over time.

Secondary aims:

- To determine the relative magnitude of the various sources of variability in echocardiographic outcomes in order to optimize operational procedures that can minimize variance.
- To determine the interstudy variability of echocardiographically-derived indices of LV systolic and diastolic function.
- To determine the relationship of clinical status, including treatment, to the interstudy variability and repeatability of echocardiographic measurements.

To achieve these aims, as noted above, there was a primary and secondary acquisition of each study echocardiogram during a study visit. There was a primary and secondary core lab reader who performed measurements on each of these acquisitions, in addition to the local sonographer who performed measurements on the primary acquisition only. Furthermore, for the baseline echocardiograms, the primary core lab reader repeated measurements in a blinded fashion of the primary acquisition at 1 month and 1 year after the image was obtained. A summary of the image analysis sets are shown in Table 1.

Table 1. VVV Study Echocardiographic image analysis sets.

#	Visit	Image acquisition	Observer	Timing of analysis
1	Initial	Primary	Study center	Initial
2	Initial	Primary	Primary core lab	Initial
3	Initial	Primary	Secondary core lab	Initial
4	Initial	Primary	Primary core lab	Repeat at 1 month
5	Initial	Primary	Primary core lab	Repeat at 1 year
6	Initial	Secondary	Primary core lab	Initial
7	Initial	Secondary	Secondary core lab	Initial
8	Followup	Primary	Study center	Initial
9	Followup	Primary	Primary core lab	Initial
10	Followup	Primary	Secondary core lab	Initial
11	Followup	Secondary	Primary core lab	Initial
12	Followup	Secondary	Secondary core lab	Initial

For any given echocardiographic parameter, there were 3x7=21 or 3x5=15 sets of measurements from each study visit (3 sequential cardiac cycles × 7 different readings for a baseline visit and 5 different readings for a follow-up visit). A total of 1,403 readings were performed on 571 echocardiograms from the 131 full protocol patients. These readings resulted in 4,209 datasets - each echocardiogram had measurements made 3 times each (3 cardiac cycles x 1,403 = 4,209).

The study design has been described in great detail in Colan et al. (*JASE* 2012; 25:842-854) and in the study protocol (available to users with approved logins). Due to the complexity of the fractional factorial design, it is strongly recommended that users read that article in full. Table 2 provides key subject characteristics. Additional information can be found in the published articles on specialized topics (see posted Bibliography at http://pediatricheartnetwork.org/ResourcesPublications/Publications/CardiomyopathyVVVStudyObservational.aspx).

Table 2. Key VVV Study Analytic Cohort (N=169) Characteristics

Characteristic	All enrolled	Partial Participation	Full Participation
N	169	38	131
Male	78 (46.2%)	20 (52.6%)	58 (44.3%)
Race			
White	112 (66.3%)	26 (68.4%)	86 (65.6%)
Black or African-American	45 (26.6%)	9 (23.7%)	36 (27.5%)
Asian	9 (5.3%)	2 (5.3%)	7 (5.3%)
Other	3 (1.8%)	1 (2.6%)	2 (1.5%)
Hispanic	22 (13.8%)	3 (8.6%)	19 (15.2%)
Age at baseline echo, yr			
Mean±SD	9.4 ± 5.9	9.8 ± 5.4	9.3 ± 6.0
Median (IQR)	9.5 (4.1, 14.9)	10.6 (5.0, 14.2)	9.2 (4.0, 15.0)
Age at cardiomyopathy diagnosis, yr			
Mean±SD	5.7 ± 6.1	6.5 ± 6.0	5.5 ± 6.1
Median (IQR)	2.7 (0.3, 11.4)	6.1 (0.3, 11.4)	1.5 (0.3, 11.5)
Time from cardiomyopathy diagnosis to enrollment, yr			
Mean±SD	3.7 ± 4.1	3.3 ± 3.9	3.8 ± 4.2
Median (IQR)	2.3 (0.4, 4.8)	2.0 (0.5, 4.0)	2.3 (0.4, 5.2)
Primary cause of dilated cardiomyopathy			
Metabolic disorder	4 (2.4%)	2 (5.3%)	2 (1.5%)
Mitochondrial disorder	2 (1.2%)	0 (0%)	2 (1.5%)
Neuromuscular disease associated with CM	6 (3.6%)	2 (5.3%)	4 (3.1%)
Single gene defect	5 (3.0%)	1 (2.6%)	4 (3.1%)
Adriamycin-associated cardiotoxicity	25 (14.8%)	11 (28.9%)	14 (10.7%)
Idiopathic	104 (61.5%)	17 (44.7%)	87 (66.4%)
Other	23 (13.6%)	5 (13.2%)	18 (13.7%)
Number of follow-up echos			
0	62 (36.7%)	38 (100%)	24 (18.3%)
1	69 (40.8%)	0 (0%)	69 (52.7%)
2	29 (17.2%)	0 (0%)	29 (22.1%)
3	7 (4.1%)	0 (0%)	7 (5.3%)
4	2 (1.2%)	0 (0%)	2 (1.5%)

Table 2, continued

Characteristic	All enrolled	Partial Participation	Full Participation
M-mode LV end-diastolic dimension (EDD)			
Mean±SD z-score	3.5 ± 2.8	1.2 ± 1.1	4.2 ± 2.8
Median (IQR) z-score	2.9 (1.6, 4.8)	1.2 (0.7, 1.7)	3.5 (2.2, 5.7)
EDD > 5.5 cm	54 (32.3%)	2 (5.3%)	52 (40.3%)
M-mode LV shortening fraction (SF)			
Mean±SD z-score	-6.2 ± 4.8	-2.2 ± 2.3	-7.4 ± 4.7
Median (IQR) z-score	-5.0 (-8.3, -2.5)	-1.9 (-3.8, -0.8)	-6.3 (-10.2, -3.7)
Mean±SD SF (%)	22.1 ± 8.2	29.7 ± 6.9	19.8 ± 7.1
SF < 28%	124 (74.3%)	13 (34.2%)	111 (86.0%)
5/6 AL LV ejection fraction (EF)			
Mean±SD z-score	-4.4 ± 2.7	-2.2 ± 1.8	-5.0 ± 2.6
Median (IQR) z-score	-4.1 (-5.9, -2.3)	-2.1 (-3.5, -0.8)	-4.8 (-6.5, -2.7)
Mean±SD EF (%)	43.0 ± 12.6	53.6 ± 8.3	39.9 ± 12.0
EF < 50%	110 (65.5%)	13 (34.2%)	97 (74.6%)

SD=standard deviation; IQR=interquartile range; LV=left ventricular; AL=area length

VENTRICULAR VOLUME VARIABILITY STUDY
PUBLIC USE DATASET

DATA AND 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 January 10, 2014. 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 any related created variables.

2. SAS version 9.4 datasets

3. The file *vvvformats.sas7bcat* – Include this file in your program using:
```
options fmtsearch = (fmtlib.vvvformats64);
```
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 website include:

- VVV Study Design paper (see https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3568492/)

- VVV Study bibliography (see http://pediatricheartnetwork.org/ResourcesPublications/Publications/CardiomyopathyVVVStudyObservational.aspx)

- VVV 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 Ventricular Volume Variability Study dataset was used in preparation of this work. Data were downloaded from http://pediatricheartnetwork.org/ForResearchers/PHNPublicUseDatasets.aspx 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 in bullet 1.
- 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

1. Identification numbers for study subjects and study sites have been re-assigned for privacy protection.
    - *blind_id*: Subject ID ranging from 1 to 275
    - *blind_site*: Site ID ranging from 1 to 8

2. To analyze the Full Protocol patient data only, restrict the data to *eligible*='F'. The created variable 'eligible' is found on Form V101 and in KEYINFO. To analyze all Full and Partial Protocol patients, restrict the data to *eligible* not equal to 'N'. There are 4 subjects who are not eligible (eligible='N') due to LV myocardial non-compaction.

3. Notation: In the Proc Contents documents, a variable in UPPER CASE letters represents an original variable from the case report form (CRF), and a variable in lower case letters represents a created variable. Prior to analysis, original variables must have any 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.

4. Annotated CRFs: For ease of reading, all variable names in the annotated CRFs are presented in upper case letters. Those highlighted in yellow are original variables; those highlighted in blue are created variables (for example, echocardiographic z-scores; 3-beat averages; age at event or age at echo, since all dates are removed from the database).

5. Local vs. Core Lab measurements: Echo variables measured at the local site have variable names beginning with L. Echo variables measured at the core laboratory have variable names beginning with C (with the exception of those in KEYINFO, see item 7 below).

6. The study data are contained in multiple individual forms. These forms may be used jointly by merging on *blind_id,* in combination with VISITNUM (0 to 4, for local echo reads) or VISIT (VSCR, VFV1 to VFV4, for core lab echo reads) AND the type of echo reading (ITEMCODE). The most commonly used reading is ITEMCODE='CD1_PSPR_IM'. This is the primary

acquisition echo (CD1) performed by the primary sonographer at the site (PS) and read by the primary reader at the core lab (PR), immediate reading (IM).

7. A single dataset called KEYINFO contains 73 of the most commonly used raw and created variables. It contains 173 records. Using *this dataset only*, the baseline data may be used without indexing on a visit or itemcode variable. These measurements all have variable names preceded by "b_", for example, *b_edd_mm* is the baseline measurement by the primary core lab reader of LV end-diastolic dimension from m-mode echocardiography (equivalent to the measurement *cmmedsad_avg* on the core lab form restricted to VISIT='VSCR' and ITEMCODE= 'CD1_PSPR_IM'.

8. The KEYINFO dataset also contains the created variable 'progression' related to disease progression as described in Molina et al., Circ Heart Failure 2014 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4111626/). It has values of 0=None/Unknown, 1=Mild, 2=Severe. Variables 62 to 73 in KEYINFO are all measures related to the calculation of disease progression.

9. To select for core laboratory echocardiogram readings (Forms V305 to V308) that have data acceptable for analysis, use ACCEPTABLE=1 (Form V305) or ACEPTECHO=1 (Forms V306, V307, V308). Unacceptable echocardiograms have no qualitative or quantitative measurements recorded.

10. It should be noted that in the main core laboratory datasets (Forms V305 to V308), there are multiple measurement and calculation approaches: M-mode, 2D, 5/6AL, Devereaux-2D. For many measurements, there are also raw values, and echo z-scores indexed to either body surface area calculated by the Haycock method, or age.

11. The core laboratory measurements were made on 3 consecutive cardiac cycles. For most analysis purposes, users will be interested in analyzing the beat-averaged data. These created variables all have names ending in "_avg". These averages use up to 3 beats; a small number of readings have fewer than 3 beats recorded on selected variables and the average is still calculated.

12. Echocardiographic z-scores are calculated only on the 3-beat-averaged core laboratory data. These variables end in "_z". The "_avg" suffix is not used.

13. All anthropometric z-score calculated variables (weight-for-age, length-for-age, weight-for-length, etc.) in the VVV Study datasets are based on the WHO standard (see de Onis M, Garza C, Onyango AW, Borghi E. Comparison of the WHO child growth standards and the CDC 2000 growth charts. J Nutr 2007; 137(1):144-148).

14. The VVV Study code lists are posted along with the datasets. Some variables are formatted (such as Complications), but with short versions of the names, and for some, the code numbers are retained and the code list must be referenced (e.g., medications, cardiac cath interventional procedures).

ADDITIONAL ASSISTANCE

If you have questions about the study dataset that this documentation and the above resources (protocol, articles) have not answered, please contact Felicia Trachtenberg (ftrachtenberg@neriscience.com) at the PHN Data Coordinating Center (617-923-7747).

APPENDIX A

Implementation of Privacy Protection Rules for Public Use of the PHN VVV 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 *blind_id*, a random consecutive numbering ranging from 1 to 275), and removal of acrostic. Of note, no names, addresses, zip code, or medical record numbers were ever contained in the original study dataset.

2. All dates in the original datasets were removed, and replaced with "Age at echo/event/intervention/procedure" in years (to 2 decimal places).

3. There was only one adverse event reported in the VVV study (Form V203) – cardiac arrest while in-hospital for treatment of heart failure. These data are not provided in the public dataset.

4. Free (write-in) text variables remain in the public dataset, with the exception of specification of type of “other” race and multiple races and (on V204) the reason for no follow-up echo, which often contained indentifying information. 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.

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