Dataset Integrity Check for the Pegylated Interferon+/-Ribavirin for Children with Hepatitis C Virus (PEDS-C) Study

Prepared by

RTI International
3040 Cornwallis Road
Research Triangle Park, NC 27709-2194
March 1, 2013

Revision History

Version	Author/Title	Date	Comments
1.0	Susan M Rogers	November, 2011	Original
2.0		April, 2012	Revision based on add'l files provided by B. Barton
2.1		June, 2012	Revisions based on query response, B. Barton
3.0		June, 2012	QC review
3.1		January, 2013	Revision following add'l analyses by B. Barton
4.0		March 2013	Final DSIC

Table of Contents

Contents

1 Standard Disclaimer

The PEDS-C study was a multicenter randomized prospective trial to evaluate the safety and efficacy of pegylated interferon (PEG) with and without ribavirin (RV) in children 5 to 17 years of age with chronic hepatitis C. As a partial check of the PEDS-C data archived in the NIDDK data repository, a dataset integrity check (DSIC) was performed to verify that selected published results from the PEDS-C study can be reproduced using the archived dataset. This DSIC consists of several analyses performed to duplicate selected results reported by Schwarz et al [1] in Gastroenterology in 2011.

The intent of this DSIC is to provide confidence that the data distributed by the NIDDK repository is a true copy of the study data. Our intent is not to assess the integrity of the statistical analyses reported by study investigators. As with all statistical analyses of complex datasets, complete replication of a set of statistical results should not be expected on a first (or second) exercise in secondary analysis. This occurs for a number of reasons including differences in the handling of missing data, restrictions on cases included in samples for a particular analysis, software coding used to define complex variables, etc. Experience suggests that most discrepancies can ordinarily be resolved by consultation with the study data coordinating center (DCC); however, this process is labor-intensive for both DCC and Repository staff. We do not attempt to resolve minor or inconsequential discrepancies with published results or discrepancies that involve complex analyses, unless NIDDK Repository staff suspect that the observed discrepancy suggests that the dataset may have been corrupted in storage, transmission, or processing by repository staff. We do, however, document those instances in which our secondary analyses produced results that were not fully consistent with those reported in the target publication.

2 Study Background

HCV RNA-positive children were randomly assigned to receive either PEG alfa-2a (PEG-2a) and ribavirin (n=55) or PEG-2a and placebo (n=59) for 48 weeks. Subjects were enrolled at 11 U.S. medical sites from December 2004 to May 2006. The primary endpoint, sustained virologic response (SVR), was defined as the lack of detectable HCV RNA at least 24 weeks after stopping therapy. Patients without detectable HCV RNA at 24 weeks were continued on treatment for another 24 weeks; patients with detectable HCV RNA at 24 weeks were considered treatment failures. Patients who failed to respond to treatment with PEG-2a plus placebo were offered "open-label" therapy with PEG-2a plus RV for 48 weeks (stopping after 24 weeks if HCV RNA remained positive). The paper by Schwarz et al. published in Gastroenterology provides the main study results for the efficacy trial.

3 Archived Datasets

The DCC submitted 80 "forms" datasets which correspond with study data collection forms and 10 "additional datasets" that provide information including laboratory results, histology reports, and viral kinetics data. Contents of the archived forms datasets match dataset descriptions provided in the Word document, *datacontents.doc*, provided by the DCC.

Analysis datasets to correspond with published results by the PEDS-C study group also were provided. The DCC provided 6 SAS analysis datasets (see Appendix 1 for brief description) and 7 SAS program files for replication of the results published in *Gastroenterology*. For this DSIC, the analysis datasets -- endpoint_out, hcv_all, oc_tab1, pathfinal, and patients were used. We also used the additional dataset of histology results, vk_jhu_fine,[1] the laboratory results data, Covance_lab, and the "forms" datasets pdc020 and pdc021. All 7 SAS programs -- Main Outcomes Table 1, Main Outcomes Table 1 path, Main Outcomes Table 2, Main Outcomes Table 3, HCV_Compare, Main_Outcomes_Table 1 12-30-2012 and Main Outcomes Table 2 12-30-2012 were reviewed. For this DSIC, the SAS program files are included as Appendix 2.

[1] The dataset, vk_jhu_fine, includes two HCV RNA lab results (with different lab dates) at visit 1 for IDNs 030550 and 030960 and for IDN 051240 at visit 24. For our DSIC analyses, we used the lab results at the latter visit date.

4 Statistical Methods

DSIC Analysis. We compared our DSIC results to the published results in Table 1 (Baseline Characteristics according to Treatment Group), Table 2 (Virologic Results by Treatment Group and Baseline Features), Figure 2 (Percent of Patients with No Detectable Virus by Time on Study (weeks) and Treatment Group), and Table 3 (Patterns of Viral Response as Predictors of SVR in Children with Genotype 1).

Our initial DSIC analyses found a number of discrepancies between the DSIC results and the published manuscript that required consultation with the DCC to attempt to resolve the differences. Consequently, our DSIC analyses were conducted in both SAS and Stata. SAS analyses were conducted by S. Tan, a NIDDK statistician, applying the SAS program files provided by the DCC. For the Stata analyses, SAS datasets were converted to Stata using Stat/Transfer (Circle Systems Inc). Stata v12 was used for all analyses (Appendix 2).

Users of the analysis datasets should note that the SAS data files and program files provided by the DCC for the manuscript analyses include the variable "IDN" for the subject ID, whereas the study forms and supplementary files archived in the repository use "MaskedID." A linkage file to match the two IDS has been supplied: link.sas7bdat.

SAS Analyses

The DCC provided 7 SAS program files to replicate results published in Tables 1 – 3 of the manuscript: Main Outcomes Table 1; Main Outcomes Table 1 path (Histology results:HAI, Steatosis, Fibrosis score); Main Outcomes Table 2; Main Outcomes Table 3; hcv_compare (to compare the primary outcome -- SVR, svr_wk72, calculated per protocol from the original Covance Laboratory data to the HCV data, hcv_all, used in the analysis file for Table 2); Main Outcomes Table 1 12-30-2012 (an updated Table 1 program file); and, Main Outcomes Table 2 12-30-2012 (an updated program file for Table 2). An Excel file, prop_ci.xls, provides the 95% CIs for the values presented in Table 2. These programs replicated the published results in Tables 1 – 3 with exceptions noted below (See Appendix 2 for SAS program files). Data users should note that these SAS files recode several variables, including main outcomes, based on the subject ID (IDN).

A few discrepancies remain between the results provided by the SAS program files and the published manuscript:

Table 1: Baseline Characteristics. The estimated mean durations of infection reported in Table 1 do not match the published results (although the results from the SAS output are the same as this DSIC). Similarly, the ALT and AST results match the DSIC tabulations but are slightly different from the manuscript.

Table 2: Virologic Results. Likewise, ALT and HCV RNA tabulations in Table 2 are somewhat similar to but do not match published results.

Output from the SAS program hcv_compare provides different values for HCV at visit 24 and 48 for two subjects in the two datasets being compared (Covance_lab and hcv_all). Output from the Covance Lab Data indicates that subject IDN 020200 has positive HCV quantitative results at visits 24 and 48, i.e., HCV_24=1 and HCV_48==1, and that subject IDN 071120 has negative HCV results at visit 24, HCV_24==0. In contrast, these values are missing in the data hcv_all, i.e., subject IDN 020200 is missing HCV_24 and HCV_48 and subject IDN 071120 is missing HCV_24. It is also noted that the data file hcv_all has a total of 114 subjects (with HCV RNA results from visits 12, 24, 48, and/or 72) whereas the file computed from the Covance_lab data has 112 subjects (subject IDNs 030960 and 080560 are missing).

Table 3: Patterns of Viral Response. The program Main Outcomes Table 3.sas produces output consistent with the published report. However, based on a finding of discrepant results between our Stata DSIC and output from the SAS program for the 1-log decreases at week 1, we note an inconsistency in SAS programming. The discrepancy appears to be related to differences in the calculation of HCV RNA values at visit 1 and in the handling of missing values for the primary outcome variable, svr_wk72 (sustained virologic response, SVR; lack of detectable HCV RNA at least 24 weeks after stopping therapy). Additional details are provided below in the discussion of Table 3 DSIC results.

As noted above, users should note that the SAS program files recode several key outcome variables by subject ID. Specifically, special handling for early drop-outs is noted in the SAS programs, Main Outcomes Table 2 and Main Outcomes Table 2 12-20-2012, as well as recodes of HCV RNA levels at various visits for selected subjects.

5 Results

DSIC Results: Baseline Characteristics according to Treatment Group. Table A lists the data variables and labels/algorithms for analysis variables that were used in our replication of Table 1. We used the analysis files, oc_tab and pathfinal, as well as the "forms" and "additional" datasets archived in the repository for our DSIC tabulations.

Table A: Data variables used in replicating Schwarz et al (2011), Table 1. Baseline Characteristics according to Treatment Group.

Variable	PEDS-C Dataset	Data Variable	Variable Label/Algorithm
Treatment group	Oc_tab1	Trtgrp	1"PEG+RV" 2"PEG only"
Patient Characteristics			
Age (years)	Oc_tab1	Agrp	1"5-11" 2"12-17"
Sex – Female	Oc_tab1	Gender	1=male, 2=female
Race – Non-Caucasian	Oc_tab1	Race	1"non white" 2"white"
BMI Z-scores	Oc_tab1	BMIZ	
Total CDI raw score	Oc_tab1	Cdi_tot	
Mode of acquisition	Oc_tab1	Hcvroute	1"maternal" 2"transf" 3"other'"
Genotype	Oc_tab1	GENOTYPE	1,2,3,6
Baseline Lab Measures			
ALT (U/L)	Covance_lab	Tst_name, visit_pvc, con_result, con_unit	Tst_name="ALT (SGPT)" & visit_pvc="BL"
ALT > Upper Limit of Normal	Covance_lab	Tst_name, visit_pvc, con_high	Defined as baseline ALT >35U/l
AST (U/L)	Covance_lab	Tst_name, visit_pvc, con_result, con_unit	Tst_name="AST (SGOT)" & visit_pvc="BL"
AST>Upper Limit of Normal	Covance_lab	Tst_name, visit_pvc, con_high	Defined as baseline AST >40U/l
Baseline HCV RNA			
HCV RNA $\log_{10}$IU/mL	Covance_lab	Tst_name, visit_pvc, con_result, con_unit	=log(10)base[HCVRNA]
HCV RNA ≥600,000 IU/mL	Covance_lab	Tst_name, visit_pvc, con_result, con_unit	tst_name=="HCVRNAQuant Cobas ED-RUO-CL-QT" & visit_pvc=="BL"
Histology			
Histology Activity Index	Histology [also pathfinal]	Kno_pot, kno_pri, kno_pcy	=kno_pot+kno_pri+kno_pcy (1/3=1)(4/6=2)(7/9=3)(10/12=4)
Steatosis	Histology [also pathfinal]	Fat	
Fibrosis Score	Histology [also pathfinal]	kno_fib	

Note: Oc_tab1 and Histology are analysis files provided by the DCC. Covance_lab and pathfinal are "additional" datasets of the PEDS-C NIDDK archive.

The published manuscript results and the DSIC results for Table 1 are shown below. The base Ns and mean values for the baseline patient characteristics and histology results calculated by the DSIC generally correspond to published values. The estimated duration of infection differed between the DSIC and published results (see highlighted results below). However, as noted above [SAS Analyses] our DSIC estimates corresponded exactly with output provided by the SAS analysis programs, Main Outcomes Table 1 and Main Outcomes Table 1 12-30-2012, provided by the DCC. The geometric means of the baseline lab measures (ALT and AST) calculated by the DSIC were similar to published results; the DCC notes that multiple measures were collected at baseline and these measures could not be exactly reproduced in the analysis files provided (B. Barton, personal communication, April 2012). Baseline HCV RNA levels were replicated using the SAS program, Main Outcomes Table 1 12-30-2012. Baseline HCV RNA levels ≥600,000 IU/ml [2] and the calculation of the log 10 HCV RNA value differed from published results.

[2] The percent of patients with baseline HCV RNA≥600,000 IU/mL in the PEG+RV group (70%) appears to be significantly different from the percent in the PEG+placebo group (82%) in the published manuscript.

Table 1. Baseline characteristics according to treatment group from *Schwarz et al., 2011 Gastroenterology 140:450-458*.

	Schwarz et al. 2011			DSIC calculations			2012-13 DCC calculations *	
	PEG + RV	PEG + placebo		PEG + RV	PEG + placebo		PEG + RV	*PEG + placebo
	(n=55)	(n=59)		(n=55)	(n=59)		(n=55)	(n=59)
Patient Characteristics								
Age (years)	10.7 (±3.3)	10.8 (±3.6)		10.7 (±3.3)	10.8 (±3.6)		10.7 (±3.3)	10.8 (±3.6)
11-May	30 (54%)	30 (51%)		30 (54%)	30 (51%)		30 (54%)	30 (51%)
17-Dec	25 (46%)	29 (49%)		25 (46%)	29 (49%)		25 (46%)	29 (49%)
Sex – Female	28 (51%)	23 (39%)		28 (51%)	23 (39%)		28 (51%)	23 (39%)
Race – Non-Caucasian	12 (22%)	17 (29%)		12 (22%)	17 (29%)		12 (22%)	17 (29%)
BMI Z-scores	0.8 (±1.0)	0.7 (±1.1)		0.7 (±1.3)	0.7 (±1.1)		0.7 (±1.3)	0.7 (±1.1)
Total CDI raw score	5.9 (±4.2)	5.9 (±4.6)		5.9 (±4.2)	5.9 (±4.6)		5.9 (±4.2)	5.9 (±4.6)
Mode of acquisition								
Maternal-infant	39 (71%)	47 (80%)		42 (76%)	46 (78%)		39 (71%)	47 (80%)
Transfusion	6 (11%)	2 (3%)		6 (11%)	3 (5%)		6 (11%)	2 (3%)
Other	10 (19%)	10 (18%)		7 (1%)	10 (17%)		10 (19%)	10 (18%)
Est. Duration of Infection (mo.)	105 (±56)	111(±55)		121 (±55)	114 (±57)		121 (±55)	114 (±57)
Genotype								
1	45 (82%)	47 (80%)		45 (76%)	47 (80%)		45 (76%)	47 (80%)
2	4 (7%)	3 (5%)		4 (7%)	3 (5%)		4 (7%)	3 (5%)
3	6 (11%)	7 (12%)		6 (11%)	7 (12%)		6 (11%)	7 (12%)
6	0	2 (3%)		0	2 (3%)		0	2 (3%)
Baseline Lab Measures								
ALT (U/L)	49 (±59)	49 (±59)		48	49		48	49
ALT > Upper Limit of Normal	32 (58%)	38 (64%)		33 (60%)	39 (66%)		33 (60%)	39 (66%)
AST (U/L)	45 (±40)	45 (±29)		43	46		43	46
AST>Upper Limit of Normal	28 (51%)	28 (47%)		29 (53%)	30 (51%)		29 (53%)	30 (51%)
Baseline HCV RNA								
HCV RNA $\log_{10}$IU/mL	6.2 (±0.8)	6.3 (±0.9)		5.6	5.6		6.2 (±0.8)	6.3 (±0.9)
HCV RNA ≥600,000 IU/mL	32 (70%)	46 (82%)		35 (64%)	43 (73%)		32 (70%)	46 (82%)
Histology results								
Histology Activity Index [a]								
Minimal	23 (43%)	24 (43%)		23 (43%)	24 (43%)		Results not provided	
Mild	10 (19%)	10 (18%)		10 (19%)	10 (18%)			
Moderate (7-9)	19 (35%)	21 (38%)		19 (35%)	21 (38%)			

	Schwarz et al. 2011			DSIC calculations			2012-13 DCC calculations *
Marked (10-12)	2 (4%)	1 (2%)		2 (4%)	1 (2%)		
Steatosis							
None	29 (54%)	34 (61%)		29 (54%)	34 (61%)		
Minimal (<=5% of tissue)	21 (39%)	17 (31%)		21 (39%)	17 (31%)		
Mild (6-33%)	4 (7%)	5 (9%)		4 (7%)	5 (9%)		
Fibrosis score							
None	7 (13%)	8 (14%)		7 (13%)	8 (14%)		
Portal-periportal (Ishak 1-2)	43 (80%)	46 (82%)		43 (80%)	46 (82%)		
Bridging fibrosis (Ishak 3-4)	4 (7%)	1 (2%)		4 (7%)	1 (2%)		
Cirrhosis (Ishak 5-6)	0	1 (2%)		0	1 (2%)		

Notes: N(%) or mean± SD. Results for ALT, AST, and HCV RNA presented as geometric mean.

[a]Sample Ns for pathology variables, PEG+RV, n=54; PEG+placebo, n=56.

*Results generated from "Main Outcomes Table 1_12-30-2012" provided by B. Barton and run by S. Tan 24 Jan 2013.

** N=46 PEG+RV, N=56 PEG

Values in parentheses represent the percentage of SVR, and values in brackets represent 95% CIs.

Tabulations highlighted in blue do not sum to "Total"

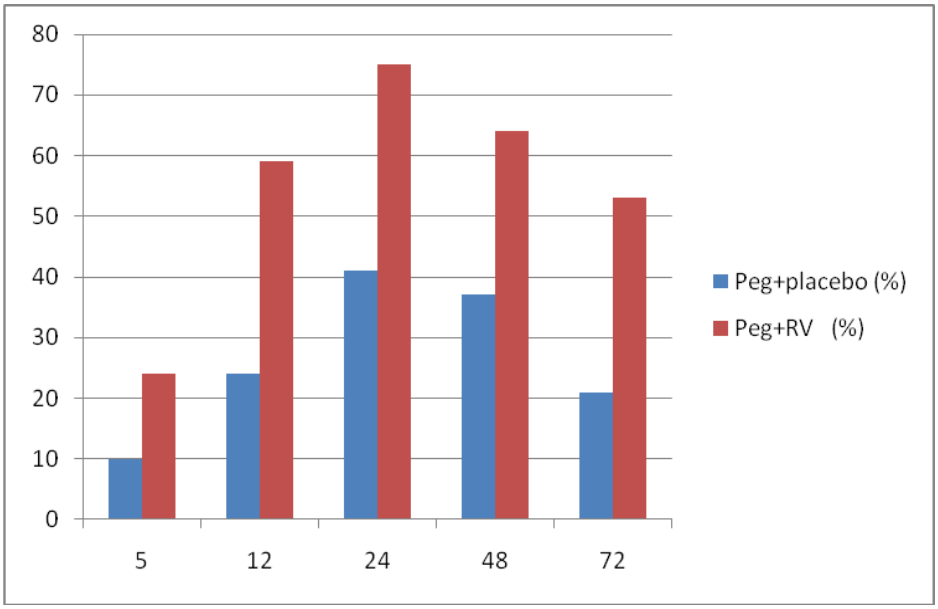
DSIC Results: Sustained Viral Response by Time on Study and Treatment Group. Figure 2 of the manuscript presents the percent of patients with no detectable virus at each time point from week 5 to week 48 of therapy, as well as 24 weeks after stopping treatment (week 72). HCV quantitative results were derived from the SAS analysis file *hcv_all*. The primary outcome, SVR at week 72, *svr_wk72*, was derived from the SAS analysis file *endpoint*. Specific coding of the HCV quantitative results and SVR at week 72 was modeled on the SAS program, Main Outcomes Table 2, provided by the DCC.

According to study protocol and the published manuscript, the primary outcome, *svr_wk72*, was the proportion of subjects with a SVR, defined as nondetectable HCV RNA (<10 IU/mL) at least 24 weeks after stopping treatment. All subjects randomized (n=114) were included in the primary efficacy analysis. Two subjects were lost to follow-up and considered treatment failures. All other dropouts were nonresponders at 24 weeks.

We note that our DSIC analysis recodes missing values of *svr_wk72* to 0 – as noted in the analysis program Main Outcomes Table 2 "if SVR is missing, the patient is a drop-out and assumed no SVR." Two individual exceptions – subjects who dropped out of the study before week 24 – and were lost to follow-up are coded as missing. This coding was retained throughout the DSIC.

DSIC calculations were similar to the published results reflecting higher levels of no detectable virus in the PEG+RV group than in the PEG+placebo group at each time point. PEG+placebo estimates in the DSIC were higher than published estimates at weeks 12, 24, and 48 (results highlighted below). DSIC results for SVR at week 72 matched the published results for both treatment groups.

Figure 2. Percent of patients with no detectable virus by time on study (wks) and treatment group from *Schwarz et al., 2011 Gastrenterology 140:450-458*.



DSIC Calculation from PEDS-C data in NIDDK repository.

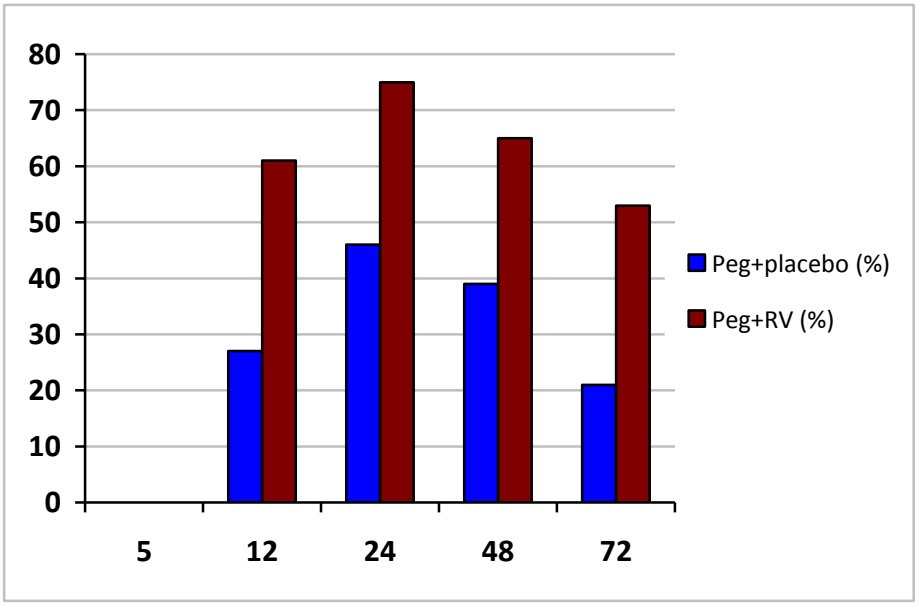


Table B. Calculated percentages of patients with no detectable virus by time on study and treatment group.

Schwarz et al, 2011				DSIC calculations		
Time on Study	Peg+placebo (%)	Peg+RV (%)	*P*	Peg+placebo (%)	Peg+RV (%)	*P*
5	10	24	0.06	nc	Nc	
12	24	59	<0.001	27	61	<0.001
24	41	75	0.003	46	75	0.002
48	37	64	0.007	39	65	0.004
72	21	53	<0.001	21	53	0.001
N				56	52-55	

DSIC Results: Virologic Results by Treatment Group and Baseline Features. Table 2 of the manuscript tabulates the Ns and percent of patients with no detectable virus at 72 weeks by treatment and patient genotype, gender, age, race, ALT, HCV RNA levels, and severity of liver histology. Variables used in these tabulations were similar to those used in Table 1 and Figure 2. The study publication reported a SVR was achieved in 29 of 55 PEG + RV subjects in comparison to 12 of 57 PEG + placebo children (p=0.0005). The DSIC calculations matched the published values for the overall Total and for most baseline features.

We note that the SAS analysis programs for Table 2, Main Outcomes Table 2 and Main Outcomes Table 2 12-30-2012, included special handling of several observations due to early dropout that included recoding of HCV RNA levels at weeks 24, 48, and 72 (see Appendix 2, Main Outcomes Table 2 and Main Outcomes Table 2 12-30-1012); we incorporated these recodes into our DSIC analyses.

As in Table 1, we were unable to exactly replicate the ALT and HCV RNA results (see highlighted DSIC results). We note that in the published manuscript, the base Ns for HCV RNA levels <600,000 (n=13) and ≥600,000 (n=46) for the PEG + placebo group do not sum to the "Total" (n=57).

Table 2. Virologic Results by Treatment Group and Baseline Features.

	Schwarz et al, 2011						DSIC calculations						2012-13 DCC calculations **					
	PEG/RV		PEG/placebo				PEG/RV		PEG/placebo				PEG/RV		PEG/placebo			
	(n=55)		(n=57)		*P* value		(n=55)		(n=57)				(n=55)		(n=57)		*P* value	
Total	29/55	53%	12/57	21%	0.0005		29/55	53%	12/57	0.21			Results identical to published ms. and DSIC calculations					
	[40-66%]		[10-32%]				[40-66%]		[10-32%]									
Genotype 1	21/45	47%	8/46	17%	0.0027		21/45	47%	8/46	17%								
	[32-61%]		[6-28%]				[32-61%]		[6-28%]									
Genotype 2-6	8/10	80%	4/11	36%	0.0563		8/10	80%	4/11	36%								
	[55-100%]		[8-65%]				[55-100%]		[8-65%]									
Female	15/28	54%	8/23	35%	0.1797		15/28	54%	8/23	35%								
	[35-72%]		[15-54%]				[35-72%]		[15-54%]									
Male	14/27	52%	4/34	12%	0.0007		14/27	52%	4/34	12%								
	[33-71%]		[1-23%]				[33-71%]		[1-23%]									
Age 11 or younger	15/30	50%	7/29	24%	0.04		15/30	50%	7/29	24%								
	[32-68%]		[9-40%]				[32-68%]		[9-40%]									
Age 12 y or older	14/25	56%	5/28	18%	0.0038		14/25	56%	5/28	18%								
	[37-75%]		[4-32%]				[37-75%]		[4-32%]									
White	22/43	51%	8/40	20%	0.0003		22/43	51%	8/40	20%								
	[36-66%]		[8-32%]				[36-66%]		[8-32%]									
Nonwhite	7/12	58%	4/17	24%	0.0651		7/12	58%	4/17	24%								
	[30-86%]		[3-44%]				[30-86%]		[3-44%]									
Normal ALT	16/23	70%	6/20	30%	0.0096		14/22	64%	7/19	37%			13/20	66%	6/21	24%	0.008	
	[51-88%]		[10-50%]															
ALT > upper limit of normal	13/32	41%	6/37	16%	0.0246		15/33	45%	5/38	13%			16/35	46%	7/36	19%	0.018	
	[24-58%]		[4-28%]															
HCV RNA <600,000 IU/ml	16/23	70%	10/13	78%	0.73		16/20	80%	9/15	60%			10/14	71%	7/9	78%	0.74	
HCV RNA ≥600,000 IU/ml	16/32	50%	5/46	11%	0.0002		13/35	37%	3/42	7%			16/32	50%	5/45	11%	0.0002	
Inflammation (HAI)																		
Minimal 1-3	10/23	43%	5/24	21%	0.0959		10/23	43%	5/24	21%			Results identical to published ms. and DSIC calculations					
	[23-64%]		[5-37%]				[23-64%]		[5-37%]									
Mild-marked 4-12	18/31	56%	6/30	20%	0.0023		18/31	56%	6/30	20%								

	Schwarz et al, 2011					DSIC calculations					2012-13 DCC calculations **			
	PEG/RV		PEG/placebo			PEG/RV		PEG/placebo			PEG/RV	PEG/placebo		
	(n=55)		(n=57)		*P* value	(n=55)		(n=57)			(n=55)	(n=57)	*P* value	
	[41-75%]		[6-34%]			[41-75%]		[6-34%]						
Fibrosis (Ishak stage)														
None	3/7	43%	3/8	38%	0.8327	3/7	43%	3/8	38%					
	[6-80%]		[4-71%]			[6-80%]		[4-71%]						
Stage 1-6	25/47	53%	8/48	17%	0.0003	25/47	53%	8/48	17%					
	[39-67%]		[6-27%]			[39-67%]		[6-27%]						
Steatosis														
Present	9/25	36%	1/21	5%	0.0105	9/25	36%	1/21	5%					
	[17-55%]		[0-14%]			[17-55%]		[0-14%]						
Absent	19/29	66%	10/33	30%	0.0056	19/29	66%	10/33	30%					
	[48-83%]		[15-46%]			[48-83%]		[15-46%]						

Note: ** Results generated from "Main Outcomes Table 2_12-30-2012" provided by B. Barton and run by S. Tan 24 Jan 2013 .

DSIC Results: Patterns of Viral Response as Predictors of SVR in Children with Genotype 1. Table 3 of the manuscript presents patterns of viral response during the first 12 weeks of the study as predictors of SVR in children with genotype 1. DSIC calculations are provided for the percent achieving response, the percent with response achieving SVR, and negative predictive values. The positive predictive values (expressed as values from 0 to 1) are equal to the percent with response achieving SVR.

DSIC calculations for Table 3 are similar to published results for patterns of RVR and EVR, but do not match published reports of the 1-log decrease at week 1.

Table 3. Patterns of Viral Response as Predictors of SVR in Children with Genotype 1 ***from Schwarz et al, 2011 Gastroenterology 140:450-458.***

	Schwarz et al, 2011			DSIC Calculations		
Response	**% Achieving Response**	**% with Response achieving SVR**	**NPV**	**% Achieving Response**	**% with Response achieving SVR**	**NPV**
PEG + RV						
1-log decrease at week 1	61	50	.43	37	85	.68
RVR (no detec virus) at week 5	15	100	.64	15	100	.66
EVR (2-log decrease) at week 12	71	65	.78	71	59	.78
PEG plus placebo						
1-log decrease at week 1	66	21	.85	27	27	.93
RVR (no detec virus) at week 5	4	100	.85	4	100	.86
EVR (2-log decrease) at week 12	40	43	.95	40	43	.95

PPV=probability of SVR given earlier response is equal to percentage with response achieving SVR; NPV=probability of no SVR given no earlier response

We note that the SAS program file, Main Outcomes Table 3.sas, provided by the DCC produced output that is consistent with the published results. Two inconsistencies in the SAS programming appear to cause the discrepancy between the DSIC and published results. First, vk1, is defined twice in the SAS program, Main Outcomes Table 3. Initially, vk1 is defined as the root of the HCV RNA value for each visit (and used in the calculation of the viral load, vk, and the log10 of the viral load, logvk). See bolded statement below:

```
if substr(results,1,6) = 'Target' then vk = 0.0;
  if substr(results,1,1) = '<' then vk = 12.5;
  if vk = 0.0 then logvk = 0.0;
  if vk = . then do;
    vk1 = input(substr(results,1,4),4.2);
    vk2 = input(substr(results,7,1),2.);
    vk = vk1*10**vk2;
   logvk = log10(vk);
```

The variable vk1 is subsequently defined in the program as the log10 of viral load at visit 1; the variable vk1 is then subtracted from the log10 of the viral load at baseline to calculate the 1-log decrease at week 1, vkdiff1:

```
 vk1 = .;
 vk3 = .;
 …
  end;
   if visit = 'BL' then vkbl = logvk;
   if visit = 'BL' then vkbl_nolog = 10**logvk;
   if visit = '1' then vk1 = logvk;
   if visit = '3' then vk3 = logvk;
   …;
   if last.idn then do;
     vkdiff1 = vkbl - vk1;
     vkdiff5 = vkbl - vk5;
     vkdiff12 = vkbl - vk12;
```

Based on the SAS programming, the initial definition overrides the definition of vk1 as the log10 of the viral load where visit is equal to one (which was the definition applied in the DSIC). As a result, the tabulation of the 1-log decrease at week 1 differs between the SAS program and published results and our DSIC.

Second, there is an additional discrepancy between the SAS program output and DSIC results regarding the number of cases in Table 3 with a missing value of svr_wk72, the primary outcome variable reflecting sustained viral response at week 72. The DSIC retains the coding of svr_wk72 that was applied in calculations of Figure 2 and Table 2, i.e., all missing values are coded as 0 – "if SVR is missing, the patient is a drop-out and assumed no SVR"-- with 2 individual exceptions who dropped out of the study before week 24, as noted in the SAS program Main Outcomes Table 2. In the SAS program Main Outcomes Table 3, missing values for svr_wk72 are handled somewhat differently. Missing values are

recoded as 0 only if the viral response at Week 24 is 0. As a result, the DSIC estimates and published results do not match.

6 Conclusions

Initial discrepancies between our DSIC calculations and the published manuscript were addressed by coordination with the PEDS-C DCC. The DCC provided multiple SAS program files to supplement our analyses in order to replicate the manuscript results, however in some instances neither the SAS program files nor our DSIC replicated the published results. The DSIC analysis matched most published results of baseline characteristics and viral response by treatment group, however we were unable to fully replicate the published ALT or HCV RNA levels and patterns of viral response (1-log decrease at week 1) differed from the published manuscript. Variation in computations involving complex analyses of repeated laboratory measurements and in the assignment of missing values for key outcome variables may explain the difference between the published data and DSIC results for the primary efficacy analysis. While we have no reason to believe that the data were compromised in storage or processing, users of the data may wish to review the SAS program files.

7 References

1. Schwarz K, Gonzalez-Peralta RP, Marray KF, Molleston JP, Haber BA, Jonas MM, Rosenthal P, Mohan P, Balistreri WF, Narkewicz MR, Smith L, Lobritto SJ, Rossi S, Valsamakis A, Goodman Z, Robuck PR, and Barton BA for the PEDS-C Clinical Research Network. The combination of ribavirin and peginterferon is superior to peginterferon and placebo for children and adolescents with chronic hepatitis C, Gastroenterology 2011;140:450-458.

8 Appendices

[1] Brief description of analysis files provided by the DCC for the Schwarz et al (2011) manuscript.

[2] Supplemental SAS programming code and output for Schwarz et al (2011) manuscript analysis provided by the DCC.

[3] STATA v11 log of programming code for DSIC analysis of Schwarz et al (2011) manuscript in *Gastroenterology*.

Appendix 1

SAS Analysis Dataset	No. observations	Description
oc.tab1	114	Baseline variables from Table 1: demographic, clinical
Pathfinal	121	Histology results
Endpoint	114	Virologic results (includes svr_wk72)
hcv_all	703	HCV quantitative results, visit
Demo	114	Demographic, virologic, histology results
Patients	114	Basic demographic variables, treatment group

Appendix 2

Main Outcomes Table 1path.sas

```
/* Program to Generate the Baseline Pathology Frequencies in PEDS-C */
/* Used in Table 1 of PEDS-C Main Results Paper                     */
/* Uses SAS data sets: PATHFINAL and OC_TAB1                        */

libname dir 'c:\peds-c\main outcomes';
proc format;
 value txf
 1 = 'Combo'
 2 = 'Mono';
 value fatf
 0 = 'None'
 1 = 'Minimal'
 2 = 'Mild';
 value haif
 1-3 = 'Minimal'
 4-6 = 'Mild'
 7-9 = 'Moderate'
 10-12 = 'Marked';
 value fibf
 0 = 'None'
 1-2 = 'Portal-periportal fibrosis'
 3-4 = 'Bridging fibrosis'
 5-6 ='Cirrhosis';
 run;

data temptrt (keep=idn trtgrp genotype);
 set dir.oc_tab1;
 run;
proc sort data=dir.pathfinal; by idn; run;
proc sort data=temptrt; by idn; run;

data temp;
 merge temptrt(in=in1) dir.pathfinal(in=in2);
 by idn;
 if in1;
 label hai = 'Histology Activity Index'
                        fat = 'Steatosis'
                        ish_fib = 'Ishak Fibrosis Score';
 run;
proc freq data=temp;
 table trtgrp*(hai fat ish_fib) / chisq;
 title 'Table 1 Pathology Variables';
 format trtgrp txf. hai haif. fat fatf. ish_fib fibf.;
 run;
```

Main Outcomes Table 2.sas

```
/* SAS Program to produce Table 2 for PEDS-C Main Outcomes Paper */

libname dir 'c:\peds-c\main outcomes';
proc format;
```

```
value $txf
'A' = 'Combo'
'B' = 'Mono';
value hcvf
0 = 'Negative'
1 = 'Positive';
value vrf
0 = 'No VR'
1 = 'VR';
value svrf
0 = 'No SVR'
1 = 'SVR';
value agef
0 = '5-11 yrs'
1 = '12-17 yrs';
value bmif
0 = '< +0.8'
1 = '>= +0.8';
value $sexf
1='Male'
2='Female';
value $hispf
1='Yes'
2='No'
3='Unk';
value $hisp2f
1 = 'Yes'
2,3 = 'No/Unk';
value racef
1='Am Indian'
2='Asian'
3='Black/AA'
4='Hawaiian'
5='White'
6='More than one'
7='Unk'
.=' ';
value race2f
1,2,3,4,6,7 = 'Other'
5='White';
value routef
1='Vertical'
2='Sex contact'
3='IV Drug'
4='Transfusion'
5='Unk'
6='Other'
.=' ';
value route2f
1 = 'Vertical'
other = 'Other';
value exempf
1='OK'
2='Refuse'
.=' ';
value pathf
.=' ';
value $gtype2f
```

```
 1 = 'Type 1'
 other='Other';
 value age11f
 0 = '< 11 yrs'
 1 = '>= 11 yrs';
 value ifibf
 0 = 'None'
 1,2 = 'Portal-Periportal'
 3 = 'Bridging'
 5 = 'Cirrhosis';
 value ifib2f
 1,2 = 'Portal-Periportal'
 0,3,5 = 'Other';
 value ifib3f
 0 = 'None'
 1,2,3,5 = 'Other';
 value kinflf
 1,2,3 = 'Minimal'
 5 = 'Mild'
 7,8,9 = 'Moderate'
 10,11 = 'Marked';
 value kinf2f
 1,2,3,5 = 'xMin-Mild'
 7,8,9,10,11 = 'Mod-Marked';
 value haif
 1,2,3 = 'Minimal'
 4,5,6,7,8,9,10,11,12 = 'Other';
 value fatf
 0 = 'None'
 1 = 'Minimal'
 2 = 'Mild';
 value fat2f
 0 = 'None'
 1,2 = 'xMin-Mild';
 value hcv7f
 0 = '<700k'
 1 = '700k+';
 value hcv6f
 0 = '<600k'
 1 = '>600k+';
 value cdif
 0 = '< 6.0'
 1 = '6.0+';
 value durf
 0 = '< 10 yrs'
 1 = '10+ yrs';
 value altf
 0 = '< 47'
 1 = '47+';
 run;

/* Demographics data set including and some other basic information */
/* Drop HCVRNA as truncated at 700000 */
data tempdemo (drop=age agrp hcvrna);
  set dir.demo;
 run;
/* Baseline Data for Table 1 of Manuscript - and used in other tables as well */
data temptab1 (keep=idn age agrp hcv700 duration bmiz cdi_tot);
```

```
 set dir.oc_tab1;
 duration = duration / 12;
run;
/* Histology / Pathology data set */
data temppath (keep=idn hai fat ish_fib k_infl);
 set dir.pathfinal;
 hai = sum(kno_pri,kno_pcy,kno_pot);
 k_infl = hai;
 run;
proc sort data=tempdemo;
 by idn;
 run;
proc sort data=dir.hcv_all;
 by idn visit;
 run;

/* Data set with Endpont SVR (Sustained Viral Response) at Week 72        */
/* Per Protocol - SVR only if Viral Response = VR at Weeks 24, 48, and 72  */
/* or only HCV Quant Results = 'Negative' at Weeks 24, 48, and 72         */
/* Based on Covance HCV Quant Results, not on Hopkins Qualitative Results  */

data tempsvr;
 set dir.endpoint_out;
 keep idn svr_wk72;
 if svr_wk72 = . then do; /* If SVR is missing, Patient is drop-out and assume No SVR */
  svr_wk72 = 0;
 end;
 if (idn = 061110 or idn = 110640) then svr_wk72 = .; /* Drop-outs before Week 24 */
 run;

/* Quantitative HCV results from Hopkins Lab                              */
/* Used in Manuscript only to show Level of Viral Response - not to Assess SVR  */

proc sort data=dira.vk_jhu_fine; by idn labdt; run;
data tempvk;
 set dir.vk_jhu_fine;
 by idn labdt;
 retain vk_bl vk_s1;
 if first.idn then do;
  vk_bl = .;
            vk_s1 = .;
 end;
 if substr(results,1,6) = 'Target' then vk = 0.0;
 if substr(results,1,1) = '<' then vk = 12.5;
 if vk = 0.0 then logvk = 0.0;
 if vk = . then do;
  vk1 = input(substr(results,1,4),4.2);
            vk2 = input(substr(results,7,1),2.);
            vk = vk1*10**vk2;
            logvk = log10(vk);
            end;
 if . < vk <= 600000 then hcv600 = 0;
 if vk > 600000 then hcv600 = 1;
 if visit = 'BL';
 run;

proc sort data=tempvk; by idn; run;
proc sort data=temptab1; by idn; run;
```

```
proc sort data=tempsvr; by idn; run;
proc sort data=temppath; by idn; run;

/* Merge all data sets to obtain Tables 2-3 of Manuscript                */
/* HCV_ALL is identical to HCV Quant Results from Covance and SVR is derived  */
/* from those results                                                    */
/* As noted, SVR = 'SVR' only if VR = 'VR' (Viral Response) at Weeks 24, 48,  */
/*    and 72 as per Protocol - if there is no viral response, there is no SVR */

data tempall;
 merge dir.hcv_all(in=in1) tempdemo(in=in2) temptab1(in=in3) temppath(in=in4)
     tempvk(in=in5) tempsvr(in=in6);
 by idn;
 if in2;
 length hcvrna $30;
 retain hcv_24 hcv_48 hcv_72 origtxgrp hcvrna;
 if first.idn then do;
  hcv_24 = .;
            hcv_48 = .;
            hcv_72 = .;
            end;
 if visit = '24' then do; /* Visit 24 if first official check for viral response */
  hcvrna = con_result;   /* Store for later use */
            if hcvrna = 'Negative' then hcv_24 = 0; /* If HCV Quant Result = Negative */
            if hcvrna = 'Positive' then do;       /* If HCV Quant Result = Positive then set later */
            hcv_24 = 1;                                                    /* visits also to Positive - indicating no Viral */
            hcv_48 = 1;                                                    /* Response and no SVR (per Protocol)          */
            hcv_72 = 1;                                                    /* SVR only if Result = Negative at Week 24, 48, */
            end;                                                                  /* and 72                                */
            end;
 if visit = '48' and hcv_48 = . then do;
            if hcvrna = 'Positive' then do;      /* If Week 48 HCV Quant Result is missing and     */
            hcv_48 = 1;                                          /*   Week 24 is Positive (still Virus in specimen)*/
            hcv_72 = 1;                                          /*   then Week 48 is assumed to be positive also  */
            end;                                                        /* Cannot assume negative if missing           */
            end;
 if visit = '72' and hcv_72 = . then do;
            if hcvrna = 'Positive' then hcv_72 = 1; /* Same logic as for Week 48                */
            end;
 if last.idn then do;                /* if last ID, special handling for some early drop-outs */
            if hcv_72 = . and (idn ne '061110' and idn ne '110640') then do;
                        hcv_72 = 1;
                        if hcv_24 = . then hcv_24 = 1;
                        if hcv_48 = . then hcv_48 = 1;
            end;
  if idn = '110750' then hcv_48 = 1;
            if idn = '061290' then hcv_48 = 0;
            if idn = '070970' then do;
                        hcv_24 = 1;
                        hcv_48 = 1;
            end;
            bmiz_c = 0;                  /* Set up some baseline variables for tables  */
            if bmiz >= 0.8 then bmiz_c = 1;
            cdi_c = 0;
            if cdi_tot >= 6.0 then cdi_c = 1;
            if duration ne . then dur_c = 0;
            if duration >= 10.0 then dur_c = 1;
            alt_c = 0;
```

```
            if alt >= 47.0 then alt_c = 1;
            output;
 end;
run;
proc contents data=tempall;
run cancel;

/* Listing of Merged data set for checking SVR Primary Outcome */
proc print data=tempall;
 title 'Listing of HCV_ALL Data Set';
 var idn hcv_24 hcv_48 hcv_72 svr_wk72;
 format hcv_24 hcv_48 hcv_72 hcvf. svr_wk72 svrf.;
 run;

/* Listing of ENDPOINT_OUT data set to check consistency     */
proc print data=dir.endpoint_out;
 title 'Listing of ENDPOINT_OUT Data Set';
 format svr_wk72 svrf. vr_wk24 vr_wk72 vrf.;
 run;

/* Table 2: Basic Results - Primary Outcome: SVR_WK72                */
/* CI for proportions calculated using Excel spreadsheet prop_ci.xls   */

proc freq data=tempall;
 table origtxgrp*(hcv_24 hcv_48 hcv_72 svr_wk72)/ chisq;
 title 'Table 2: Outcome Event Rates by Treatment Group';
 format origtxgrp $txf. hcv_24 hcv_48 hcv_72 hcvf. svr_wk72 svrf.;
 run;

/* Results by Subgroups for Table 2                    */
proc freq data=tempall;
table (genotype gender agrp race hai ish_fib fat)*origtxgrp*svr_wk72/ chisq;
title 'Table 2: SVR Event Rates by Treatment Group within Subgroup';
format origtxgrp $txf. hcv_24 hcv_48 hcv_72 hcvf. svr_wk72 svrf. agrp agef. gender $sexf.
 bmiz_c bmif. race race2f. hcvroute routef. genotype $gtype2f. fat fat2f.
 ish_fib ifib3f. hai haif. hcv600 hcv6f.;
run;
```

Main Outcomes Table 3.sas

```
/* SAS Program to produce Table 3 of PEDS-C Main Outcomes Paper */

libname dir 'c:\peds-c\main outcomes';

proc format;
 value $txf
 'A' = 'Combo'
 'B' = 'Mono';
 value yesnof
 0 = 'No'
 1 = 'Yes';
 run;
proc contents data=dir.vk_jhu_fine;
run;

/* Data step to convert Results variable from character to numeric for numeric viral load */
/*    data from JHU lab                                          */

data temp;
```

```
 set dir.vk_jhu_fine;
 if substr(results,1,6) = 'Target' then vk = 0.0;
 if substr(results,1,1) = '<' then vk = 12.5;
 if vk = 0.0 then logvk = 0.0;
 if vk = . then do;
  vk1 = input(substr(results,1,4),4.2);
            vk2 = input(substr(results,7,1),2.);
            vk = vk1*10**vk2;
            logvk = log10(vk);  /* Calculations below done in Log10 for Viral Load (VK) */
            end;
 run;

/* Print statement to check conversion of results */

proc print data=temp;
 var idn visit results vk vk1 vk2 logvk;
 run cancel;
proc sort data=temp;
 by idn visit;
 run;

/* Data Step to calculate and store VK results in appropriate weeks for easier analysis */
/* All calucations/storage done within IDN (patient ID)  */

data tempx;
 set temp;
 by idn;
 retain vk3 vkbl vk1 vk5 vk12 vk24 vkbl_nolog;
 if first.idn then do;
  vk1 = .;
  vk3 = .;
            vk5 = .;
            vk12 = .;
            vk24 = .;
  vkbl = .;
            vkbl_nolog = .;
 end;
 if visit = 'BL' then vkbl = logvk;
 if visit = 'BL' then vkbl_nolog = 10**logvk;
 if visit = '1' then vk1 = logvk;
 if visit = '3' then vk3 = logvk;
 if visit = '5' then vk5 = logvk;
 if visit = '12' then vk12 = logvk;
 if visit = '24' then vk24 = logvk;
 if last.idn then do;
  vkdiff1 = vkbl - vk1;
            vkdiff5 = vkbl - vk5;
            vkdiff12 = vkbl - vk12;
            output;
 end;
run;

/* Print statement to verify appropriate VK storage in weeks */

proc print data=tempx;
 var idn vkbl vk1 vk3 vk5 vk12 vk24;
 run;
```

```
/* Keep genotype from Patients Data Set */

data tempdemo;
 set dir.patients;
 keep idn genotype;
 run;
proc sort data=tempdemo; by idn; run;

/* Data step to calculate changes from baseline for Table 3: 1 log drop at Week 1, */
/*     RVR (no detectable virus) at Week 5, and EVR (2-log drop) at Week 12       */

data tempz;
 length rvr evr vr12 vk600k $ 3 vklog1 $ 5;
 merge dir.endpoint_out(in=in1) tempdemo(in=in2) tempx(in=in3);
 by idn;
 if in1;
 if svr_wk72 = . and vr_wk24 = 0 then svr_wk72 = 0;
 if vk5 ne . then rvr = 'No';
 if vk5 = 0.0 then rvr = 'Yes';
 if vkdiff12 ne . then evr = 'No';
 if vkdiff12 >= 2.0 then evr = 'Yes';
 if (vkbl ne . and vk1 ne .) then vklog1 = 'No';
 if (vkbl ne . and vk1 ne .) then vklog2 = 'No';
 if (vkbl - vk1) >= 1.0 then vklog1 = 'Yes';
 if (vkbl - vk1) >= 2.0 then vklog2 = 'Yes';
 if (vkbl ne . and vk12 ne .) then vklog12 = 'No';
 if (vkbl - vk12) >= 1.0 then vklog12 = '1 Log';
 if (vkbl - vk12) >= 2.0 then vklog12 = '2 Log';
 if vk12 ne . then vr12 = 'No';
 if vk12 = 0.0 then vr12 = 'Yes';
 if vkbl_nolog ne . then vk600k = 'No';
 if vkbl_nolog >= 600000 then vk600k = 'Yes';
 label evr = 'Early Viral Response (Week 5)'
                        rvr = 'Rapid Viral Response (Week 5)'
                        vklog1 = 'Week 1 Log Drop'
                        vklog12 = 'Week 12 Log Drop'
                        vr12 = 'Viral Response Week 12';
run;
proc freq data=tempz;
 table genotype;
 run;
proc contents data=tempz;
run;

/* Table 3 is for Genotype 1 only */

data tempz1;
 set tempz;
 if genotype eq '1';
 run;

/* Tables below are in the Same Order as in Table 3 */
/* For Each Proc Freq, the first table is used for the "Percent Achieving Response" column */
/*   in Table 3 and second/third tables are for the individual tx group column for       */
/*   "Percent with response achieving SVR" - these tables are also used for the last     */
/*   section of the table with the two treatment groups combined with the exception of   */
/*   "Percent with response achieving SVR" which was calculated with a hand calculator   */
/* Columns for "Positive Predictive Value" and "Negative Predictive Value" were also     */
```

```
/*   taken from the second/third tables for each outcome                    */

proc freq data=tempz1;
 table origtxgrp*vklog1 origtxgrp*vklog1*svr_wk72 / chisq;
 title 'Table 3: Week 1 Log Drop Results';
 title2 'Genotype 1 Patients Only';
 format svr_wk72 yesnof. origtxgrp $txf.;
 run;

proc freq data=tempz1;
 table origtxgrp*rvr origtxgrp*rvr*svr_wk72 / chisq;
 title 'Table 3: RVR results';
 title2 'Genotype 1 Patients Only';
 format svr_wk72 yesnof. origtxgrp $txf.;
 run;

proc freq data=tempz1;
 table origtxgrp*evr origtxgrp*evr*svr_wk72 / chisq;
 title 'table 3: EVR results';
 title2 'Genotype 1 Patients Only';
 format svr_wk72 yesnof. origtxgrp $txf.;
 run;
```

HCV_compare.sas

```
/* Program: HCV_COMPARE.SAS
   Program to pull raw data from COVANCE_LAB.SAS7BDAT data set and compare
   HCV Quantitative Results to Analysis File HCV_ALL.SAS7BDAT              */

libname dir 'c:\peds-c\sas data sets';
libname dira 'c:\peds-c\main outcomes';
proc format;
 value hcvf
 0 = 'Negative'
 1 = 'Positive';
 run;
proc freq data=dir.covance_lab;  /* List of Test Names to find Correct HCV Quant Name */
 table tst_name;
 run;
data temphcv;  /* Data Step to pull only HCV Quant Test Results */
 set dir.covance_lab;
 length idn $6 visit $20 labdt $ 9;
 keep idn visit labdt con_result;
 idn = pat_num;                    /* Rename variables to match HCV_ALL data set Names */
 visit = visit_pvc;
 labdt = col_date;
 if substr(tst_name,1,18) = 'HCVPCRQualCobas2.0';
 if substr(con_result,1,9) = 'Equivocal' then delete; /* These are repeated to get Pos/Neg */
 run;
options obs=20;
proc print data=temphcv;
 var pat_num visit_pvc tst_code tst_name con_result col_date;
 run;
options obs=max;

/* Restructure HCV_ALL into one record per subject with indicators of HCV results by visit */

data temp_hcv_all;
 merge dira.hcv_all;
```

```
 by idn;
 keep idn hcv_24 hcv_48 hcv_72;
 length hcvrna $ 30;
 retain hcv_24 hcv_48 hcv_72 hcvrna;
 if first.idn then do;  /* Iniitalize retain variables */
  hcv_24 = .;        /* Retain HCV Quant results for Visit 24 */
            hcv_48 = .;        /* Retain HCV Quant results for Visit 48 */
            hcv_72 = .;        /* Retain HCV Quant results for Visit 72 */
            hcvrna = .;        /* Retain HCV_24 Quant results to help with later visits */
            end;
 if visit = '24' then do;
  hcvrna = con_result; /* Con_Result is actual result for HCV Quant Assay = 'Positive' / 'Negative' */
            if hcvrna = 'Negative' then hcv_24 = 0; /* Result = 'Negative' indicates no virus in specimen */
            if hcvrna = 'Positive' then do;       /* Result = 'Positive' indicates virus in specimen */
            hcv_24 = 1;                    /* Per Protocol, if "Positive" at 24 weeks, assume */
            hcv_48 = 1;                                                    /* to be positive through out (i.e., no SVR)      */
            hcv_72 = 1;
            end;
            end;
 if visit = '48' and hcv_48 = . then do; /* Only do if no HCV result at 48 weeks */
            if hcvrna = 'Positive' then do;      /* Only for Week 24 = Positive per Protocol */
            hcv_48 = 1;                    /* Otherwise, cannot assume to be Negative */
            hcv_72 = 1;
            end;
            end;
 if visit = '72' and hcv_72 = . then do; /* Only do if no HCV result at 72 weeks */
            if hcvrna = 'Positive' then hcv_72 = 1; /* Only for Week 24 = Positive per Protocol */
            end;                          /* Otherwise, cannot assume to be Negative */
 if last.idn then do;
            if hcv_72 = . and (idn ne '061110' and idn ne '110640') then do; /* special handling for drop-outs */
                        hcv_72 = 1;
                        if hcv_24 = . then hcv_24 = 1;
                        if hcv_48 = . then hcv_48 = 1;
            end;
   if idn = '110750' then hcv_48 = 1;
            if idn = '061290' then hcv_48 = 0;
            if idn = '070970' then do;
                        hcv_24 = 1;
                        hcv_48 = 1;
            end;
            output;
 end;
 label hcv_24 = 'HCV Quant Result - Week 24'
                        hcv_48 = 'HCV Quant Result - Week 48'
                        hcv_72 = 'HCV Quant Result - Week 72';
run;
proc sort data=temphcv; by idn visit; run;

/* Same Data Program Flow and Comments as above for HCV_ALL data Set */

data temp_covance;
 merge temphcv;
 by idn;
 keep idn hcv_24 hcv_48 hcv_72;
 length hcvrna $ 30;
 retain hcv_24 hcv_48 hcv_72 hcvrna;
 if first.idn then do;
   hcv_24 = .;
```

```
            hcv_48 = .;
            hcv_72 = .;
            hcvrna = .;
            end;
 if visit = '24' then do;
  hcvrna = con_result;
            if hcvrna = 'Negative' then hcv_24 = 0;
            if hcvrna = 'Positive' then do;
            hcv_24 = 1;
            hcv_48 = 1;
            hcv_72 = 1;
            end;
            end;
 if visit = '48' and hcv_48 = . then do;
            if hcvrna = 'Positive' then do;
            hcv_48 = 1;
            hcv_72 = 1;
            end;
            end;
 if visit = '72' and hcv_72 = . then do;
            if hcvrna = 'Positive' then hcv_72 = 1;
            end;
 if last.idn then do;
            if hcv_72 = . and (idn ne '061110' and idn ne '110640') then do;
                        hcv_72 = 1;
                        if hcv_24 = . then hcv_24 = 1;
                        if hcv_48 = . then hcv_48 = 1;
            end;
  if idn = '110750' then hcv_48 = 1;
            if idn = '061290' then hcv_48 = 0;
            if idn = '070970' then do;
                        hcv_24 = 1;
                        hcv_48 = 1;
            end;
            output;
 end;
 label hcv_24 = 'HCV Quant Result - Week 24'
                        hcv_48 = 'HCV Quant Result - Week 48'
                        hcv_72 = 'HCV Quant Result - Week 72';
run;

proc print data=temp_covance;
 title 'Listing from Covance Data Set';
 format hcv_24 hcv_48 hcv_72 hcvf.;
 run;

proc print data=temp_hcv_all;
 title 'Listing from HCV_ALL Data Set';
 format hcv_24 hcv_48 hcv_72 hcvf.;
 run;

/* Run PROC COMPARE to verify that the two Data Sets give Identical Results         */
/* NOTE: In later analysis, a "Positive" HCV Quant Result indicates NO Viral Response */
/*  and a "Negative" HCV Quant Result indicates a Viral Response                    */
/*  Formats in Later Programs Reflect this coding convention                        */

proc compare base=temp_hcv_all compare=temp_covance;
run;
```

Main Outcomes Table 1 12-30-2012

```
/* SAS Program to generate frequencies for baseline variables in PEDS-C */
/* Generates the non-pathology variables in Table 1 of Main Outcomes Paper */
/* Uses SAS data set: OC_TAB1 vk_jhu_fine covance_lab                    */

libname dir 'c:\peds-c\main outcomes';
libname dira 'c:\peds-c\roche material';

proc format;
 value txf
 1 = 'Combo'
 2 = 'Mono';
 value $genef
 '1' = 'Type 1'
 '2' = 'Type 2'
 '3' = 'Type 3'
 '6' = 'Type 6';
 value agrpf
 0 = 'Age 5-11'
 1 = 'Age 12-17';
 value $sexf
 '1' = 'Male'
 '2' = 'Female';
 value racef
 1,2,3,6,7 = 'Non-white'
 5 = 'White';
 value routef
 1 = 'Maternal-infant'
 4 = 'Transfusion'
 2,3,5,6 = 'Other';
 value hcvf
 0 = 'HCV < 700'
 1 = 'HCV >= 700';
 run;

proc contents data=dir.oc_tab1;
run;

data temp020;
 set dir.pdc020;
 keep idn hcvroute;
 run;

data temp021;
 set dir.pdc021;
 keep idn hcvroute;
 run;

data temp02;
 set temp020 temp021;
 run;

proc freq data=temp02;
 table hcvroute;
 run;
```

```
proc sort data=temp02; by idn; run;

data temptab1;
 merge dir.oc_tab1(in=in1) temp02;
 if in1;
 by idn;
 run;

proc contents data=temptab1;
run;

 /*  T-Test for Continuous Variables in Table 1  */
 /*  Baseline ALT, AST were done later in program using Covance Data */
proc ttest data=dir.oc_tab1;
 class trtgrp;
 var age cdi_tot duration bmiz;
 title 'Table 1 Means';
 format trtgrp txf.;
 run;

/* Frequency/percentages for categorical variables in Table 1 */

proc freq data=temptab1;
 table trtgrp*(genotype agrp gender hcvroute race) / chisq;
 format trtgrp txf. genotype $genef. agrp agrpf. gender $sexf. hcvroute routef. race racef.;
 run;

/* Program segment for HCVRNA Table 1 Results */

data temp;
 set dir.vk_jhu_fine;
 if substr(results,1,6) = 'Target' then vk = 0.0;
 if substr(results,1,1) = '<' then vk = 12.5;
 if vk = 0.0 then logvk = 0.0;
 if vk = . then do;
  vk1 = input(substr(results,1,4),4.2);
            vk2 = input(substr(results,7,1),2.);
            vk = vk1*10**vk2;
            logvk = log10(vk);
            end;
 run;

proc sort data=temp;
 by idn visit;
 run;

data tempx;
 set temp;
 by idn;
 retain vks1 vk3 vkbl vk1 vk5 vk12 vk24;
 if first.idn then do;
  vks1 = .;
  vk1 = .;
  vk3 = .;
            vk5 = .;
            vk12 = .;
            vk24 = .;
  vkbl = .;
```

```
 end;
 if visit = 'S1' then vks1 = logvk;
 if visit = 'BL' then vkbl = logvk;
 if visit = 'BL' then vkbl_nolog = 10**logvk;
 if visit = '1' then vk1 = logvk;
 if visit = '3' then vk3 = logvk;
 if visit = '5' then vk5 = logvk;
 if visit = '12' then vk12 = logvk;
 if visit = '24' then vk24 = logvk;
 if last.idn then do;
  vkdiff1 = vkbl - vk1;
            vkdiff5 = vkbl - vk5;
            vkdiff12 = vkbl - vk12;
            output;
 end;
run;

data tempz;
 merge temptab1(in=in1) tempx(in=in2);
 by idn;
 if in1;

 if vks1 ne . then do;
  nologvks1 = 10**vks1;
  if nologvks1 < 600000 then vks1_600 = 0;
            if nologvks1 >= 600000 then vks1_600 = 1;
 end;
 if vkbl ne . then do;
  nologvkbl = 10**vkbl;
            if nologvkbl < 600000 then vkbl_600 = 0;
            if nologvkbl >= 600000 then vkbl_600 = 1;
 end;

 run;

proc format;
 value vk600f
 0 = '< 600'
 1 = '>= 600';
 run;

proc ttest data=tempz;
 class trtgrp;
 var vkbl;
 title 'Table 1: Means, SD, and T-Test of Baseline Log10 HCV RNA Levels';
 format trtgrp txf.;
 run;

proc freq data=tempz;
 table trtgrp*vkbl_600/chisq;
 title 'Table 1: Frequency of HCV RNA >= 600';
 format trtgrp txf. vkbl_600 vk600f.;
 run;

/*****************************************************************************/
/* Program segment for ALT and AST results                                   */
/* Following code pull the ALT and AST results from the Original Covance Data */
/*   TST_CODE RCT4 is ALT and TST_CODE RCT5 is AST                           */
```

```
/*   RES_TYPE ne 'C' selects original results only                    */
/*************************************************************************/

data tempa (keep=idn con_result con_high con_low tst_code col_date do_coll visit_pvc);
 set dira.covance_lab;
 where (visit_pvc = 'S1' or visit_pvc = 'BL')  and (tst_code = 'RCT4' or tst_code = 'RCT5')
                        and res_type ne 'C';
 idn = pat_num;
 xday = substr(col_date,1,2);
 cmonth = substr(col_date,3,3);
  if cmonth = 'JAN' then xmonth = 1;
            else if cmonth = 'FEB' then xmonth = 2;
            else if cmonth = 'MAR' then xmonth = 3;
            else if cmonth = 'APR' then xmonth = 4;
            else if cmonth = 'MAY' then xmonth = 5;
            else if cmonth = 'JUN' then xmonth = 6;
            else if cmonth = 'JUL' then xmonth = 7;
            else if cmonth = 'AUG' then xmonth = 8;
            else if cmonth = 'SEP' then xmonth = 9;
            else if cmonth = 'OCT' then xmonth = 10;
            else if cmonth = 'NOV' then xmonth = 11;
            else if cmonth = 'DEC' then xmonth = 12;
 xyear = substr(col_date,6,4);
 do_coll = mdy(xmonth,xday,xyear);
 run;

proc sort data=tempa; by idn tst_code do_coll; run;

data temp_first;
 set tempa;
 by idn tst_code;
 if first.tst_code; /* keep first occurence of ALT/AST for each patient */
 run;

/*************************************************************************/
/* Following code sets up low/high flags for ALT/AST for Table1         */
/*    using high and low indicators in Covance File                     */
/*    Generate seperate files for ALT/AST because on separate records in  */
/*    Covance Data Set                                                  */
/*************************************************************************/

data tempalt;
 set temp_first;
 if tst_code = 'RCT4';
            alt_cov = input(con_result,3.);
            alt_low = input(con_low,3.);
            alt_high = input(con_high,3.);
            if alt_low < alt_cov < alt_high then flg_alt_cov = 'OK';
            if . < alt_cov <= alt_low then flg_alt_cov = 'Low';
            if alt_high <= alt_cov then flg_alt_cov = 'Hi';
            logalt_cov = log10(alt_cov);
run;

data tempast;
 set temp_first;
 if tst_code = 'RCT5';
            ast_cov = input(con_result,3.);
            ast_low = input(con_low,3.);
```

```
            ast_high = input(con_high,3.);
            if ast_low < ast_cov < ast_high then flg_ast_cov = 'OK';
            if . < ast_cov <= ast_low then flg_ast_cov = 'Low';
            if ast_high <= ast_cov then flg_ast_cov = 'Hi';
            logast_cov = log10(ast_cov);
run;

proc sort data=tempalt; by idn; run;
proc sort data=tempast; by idn; run;

/*****************************************************************/
/* Merge ALT and AST data sets together to get everything on one  */
/* record                                                        */
/*****************************************************************/

data temp_cov;
 merge tempalt tempast;
 by idn;
 run;


proc sort data=temp_cov; by idn do_coll; run;

data temp_comb;
 merge dir.oc_tab1(in=in1) temp_cov(in=in2); /* merge with OC_TAB1 to get trtgrp */
 by idn;
 if in1;
 run;

proc freq data=temp_comb;
 table trtgrp*(flg_alt_cov flg_ast_cov) / chisq;
 title 'Table 1: Frequencies of HI/LO of ALT/AST';
 format trtgrp txf.;
 run;

proc sort data=temp_comb; by trtgrp; run;

proc ttest data=temp_comb dist=lognormal;
 class trtgrp;
 var alt_cov ast_cov;
 run;
```

Main Outcomes Table 2 12-30-2012

```
/* SAS Program to produce Table 2 for PEDS-C Main Outcomes Paper */

libname dir 'c:\peds-c\main outcomes';
libname dira 'c:\peds-c\roche material';

proc format;
 value $txf
 'A' = 'Combo'
 'B' = 'Mono';
 value hcvf
 0 = 'Negative'
 1 = 'Positive';
 value vrf
 0 = 'No VR'
 1 = 'VR';
```

```
value svrf
0 = 'No SVR'
1 = 'SVR';
value agef
0 = '5-11 yrs'
1 = '12-17 yrs';
value bmif
0 = '< +0.8'
1 = '>= +0.8';
value $sexf
1='Male'
2='Female';
value $hispf
1='Yes'
2='No'
3='Unk';
value $hisp2f
1 = 'Yes'
2,3 = 'No/Unk';
value racef
1='Am Indian'
2='Asian'
3='Black/AA'
4='Hawaiian'
5='White'
6='More than one'
7='Unk'
.=' ';
value race2f
1,2,3,4,6,7 = 'Other'
5='White';
value routef
1='Vertical'
2='Sex contact'
3='IV Drug'
4='Transfusion'
5='Unk'
6='Other'
.=' ';
value route2f
1 = 'Vertical'
other = 'Other';
value exempf
1='OK'
2='Refuse'
.=' ';
value pathf
.=' ';
value $gtype2f
1 = 'Type 1'
other='Other';
value age11f
0 = '< 11 yrs'
1 = '>= 11 yrs';
value ifibf
0 = 'None'
1,2 = 'Portal-Periportal'
3 = 'Bridging'
```

```
 5 = 'Cirrhosis';
 value ifib2f
 1,2 = 'Portal-Periportal'
 0,3,5 = 'Other';
 value ifib3f
 0 = 'None'
 1,2,3,5 = 'Other';
 value kinflf
 1,2,3 = 'Minimal'
 5 = 'Mild'
 7,8,9 = 'Moderate'
 10,11 = 'Marked';
 value kinf2f
 1,2,3,5 = 'xMin-Mild'
 7,8,9,10,11 = 'Mod-Marked';
 value haif
 1,2,3 = 'Minimal'
 4,5,6,7,8,9,10,11,12 = 'Other';
 value fatf
 0 = 'None'
 1 = 'Minimal'
 2 = 'Mild';
 value fat2f
 0 = 'None'
 1,2 = 'xMin-Mild';
 value hcv7f
 0 = '<700k'
 1 = '700k+';
 value hcv6f
 0 = '<600k'
 1 = '>600k+';
 value cdif
 0 = '< 6.0'
 1 = '6.0+';
 value durf
 0 = '< 10 yrs'
 1 = '10+ yrs';
 value altf
 0 = '< 47'
 1 = '47+';
 run;

/* Demographics data set including and some other basic information */
/* Drop HCVRNA as truncated at 700000 */
data tempdemo (drop=age agrp hcvrna);
  set dir.demo;
 run;

/* Baseline Data for Table 1 of Manuscript - and used in other tables as well */
data temptab1 (keep=idn age agrp hcv700 duration bmiz cdi_tot);
  set dir.oc_tab1;
  duration = duration / 12;
run;

/* Histology / Pathology data set */
data temppath (keep=idn hai fat ish_fib k_infl);
  set dir.pathfinal;
  hai = sum(kno_pri,kno_pcy,kno_pot);
```

```
 k_infl = hai;
 run;

proc sort data=tempdemo;
 by idn;
 run;

proc sort data=dir.hcv_all;
 by idn visit;
 run;

/* Data set with Endpont SVR (Sustained Viral Response) at Week 72        */
/* Per Protocol - SVR only if Viral Response = VR at Weeks 24, 48, and 72  */
/* or only HCV Quant Results = 'Negative' at Weeks 24, 48, and 72         */
/* Based on Covance HCV Quant Results, not on Hopkins Qualitative Results  */

data tempsvr;
 set dir.endpoint_out;
 keep idn svr_wk72;
 if svr_wk72 = . then do; /* If SVR is missing, Patient is drop-out and assume No SVR */
  svr_wk72 = 0;
 end;
 if (idn = 061110 or idn = 110640) then svr_wk72 = .; /* Drop-outs before Week 24 */
 run;

/* Quantitative HCV results from Hopkins Lab                              */
/* Used in Manuscript only to show Level of Viral Response - not to Assess SVR  */

proc sort data=dira.vk_jhu_fine; by idn labdt; run;

data tempvk;
 set dir.vk_jhu_fine;
 by idn labdt;
 retain vk_bl vk_s1;
 if first.idn then do;
  vk_bl = .;
            vk_s1 = .;
 end;
 if substr(results,1,6) = 'Target' then vk = 0.0;
 if substr(results,1,1) = '<' then vk = 12.5;
 if vk = 0.0 then logvk = 0.0;
 if vk = . then do;
  vk1 = input(substr(results,1,4),4.2);
            vk2 = input(substr(results,7,1),2.);
            vk = vk1*10**vk2;
            logvk = log10(vk);
            end;
 if . < vk <= 600000 then hcv600 = 0;
 if vk > 600000 then hcv600 = 1;
 if visit = 'BL';
 run;

proc sort data=tempvk; by idn; run;
proc sort data=temptab1; by idn; run;
proc sort data=tempsvr; by idn; run;
proc sort data=temppath; by idn; run;

/* Merge all data sets to obtain Tables 2-3 of Manuscript                 */
```

```
/* HCV_ALL is identical to HCV Quant Results from Covance and SVR is derived  */
/* from those results                                                          */
/* As noted, SVR = 'SVR' only if VR = 'VR' (Viral Response) at Weeks 24, 48,  */
/*    and 72 as per Protocol - if there is no viral response, there is no SVR */

data tempall;
 merge dir.hcv_all(in=in1) tempdemo(in=in2) temptab1(in=in3) temppath(in=in4)
      tempvk(in=in5) tempsvr(in=in6);
 by idn;
 if in2;
 length hcvrna $30;
 retain hcv_24 hcv_48 hcv_72 origtxgrp hcvrna;
 if first.idn then do;
  hcv_24 = .;
            hcv_48 = .;
            hcv_72 = .;
            end;
 if visit = '24' then do; /* Visit 24 if first official check for viral response */
  hcvrna = con_result;   /* Store for later use */
            if hcvrna = 'Negative' then hcv_24 = 0; /* If HCV Quant Result = Negative */
            if hcvrna = 'Positive' then do;        /* If HCV Quant Result = Positive then set later */
            hcv_24 = 1;                                                                    /* visits also to Positive -
indicating no Viral */
            hcv_48 = 1;                                                                    /* Response and no SVR (per
Protocol)          */
            hcv_72 = 1;                                                                    /* SVR only if Result =
Negative at Week 24, 48, */
            end;                                                                           /* and 72
*/
            end;
 if visit = '48' and hcv_48 = . then do;
            if hcvrna = 'Positive' then do;       /* If Week 48 HCV Quant Result is missing and    */
            hcv_48 = 1;                                                                  /*   Week 24 is Positive (still Virus in
specimen)*/
            hcv_72 = 1;                                                                  /*   then Week 48 is assumed to be
positive also  */
            end;                                                                         /* Cannot assume negative if missing
*/
            end;
 if visit = '72' and hcv_72 = . then do;
            if hcvrna = 'Positive' then hcv_72 = 1; /* Same logic as for Week 48                */
            end;
 if last.idn then do;                /* if last ID, special handling for some early drop-outs */
            if hcv_72 = . and (idn ne '061110' and idn ne '110640') then do;
                        hcv_72 = 1;
                        if hcv_24 = . then hcv_24 = 1;
                        if hcv_48 = . then hcv_48 = 1;
            end;
  if idn = '110750' then hcv_48 = 1;
            if idn = '061290' then hcv_48 = 0;
            if idn = '070970' then do;
                        hcv_24 = 1;
                        hcv_48 = 1;
            end;
            bmiz_c = 0;                  /* Set up some baseline variables for tables  */
            if bmiz >= 0.8 then bmiz_c = 1;
            cdi_c = 0;
            if cdi_tot >= 6.0 then cdi_c = 1;
```

```
            if duration ne . then dur_c = 0;
            if duration >= 10.0 then dur_c = 1;
            alt_c = 0;
            if alt >= 47.0 then alt_c = 1;
            output;
 end;
run;

proc contents data=tempall;
run cancel;

/* Listing of Merged data set for checking SVR Primary Outcome */
proc print data=tempall;
 title 'Listing of HCV_ALL Data Set';
 var idn hcv_24 hcv_48 hcv_72 svr_wk72;
 format hcv_24 hcv_48 hcv_72 hcvf. svr_wk72 svrf.;
 run;

/* Listing of ENDPOINT_OUT data set to check consistency     */
proc print data=dir.endpoint_out;
 title 'Listing of ENDPOINT_OUT Data Set';
 format svr_wk72 svrf. vr_wk24 vr_wk72 vrf.;
 run;

/* Table 2: Basic Results - Primary Outcome: SVR_WK72                */
/* CI for proportions calculated using Excel spreadsheet prop_ci.xls    */

proc freq data=tempall;
 table origtxgrp*(hcv_24 hcv_48 hcv_72 svr_wk72)/ chisq;
 title 'Table 2: Outcome Event Rates by Treatment Group';
 format origtxgrp $txf. hcv_24 hcv_48 hcv_72 hcvf. svr_wk72 svrf.;
 run;

/* Results by Subgroups for Table 2                    */
proc freq data=tempall;
table (genotype gender agrp race hai ish_fib fat hcv600)*origtxgrp*svr_wk72/ chisq;
title 'Table 2: SVR Event Rates by Treatment Group within Subgroup';
format origtxgrp $txf. hcv_24 hcv_48 hcv_72 hcvf. svr_wk72 svrf. agrp agef. gender $sexf.
 bmiz_c bmif. race race2f. hcvroute routef. genotype $gtype2f. fat fat2f.
 ish_fib ifib3f. hai haif. hcv600 hcv6f.;
run;


data tempa (keep=idn con_result con_high con_low tst_code col_date do_coll visit_pvc);
 set dira.covance_lab;
 where (visit_pvc = 'S1' or visit_pvc = 'BL')  and (tst_code = 'RCT4' or tst_code = 'RCT5')
                        and res_type ne 'C';
 idn = pat_num;
 xday = substr(col_date,1,2);
 cmonth = substr(col_date,3,3);
   if cmonth = 'JAN' then xmonth = 1;
            else if cmonth = 'FEB' then xmonth = 2;
            else if cmonth = 'MAR' then xmonth = 3;
            else if cmonth = 'APR' then xmonth = 4;
            else if cmonth = 'MAY' then xmonth = 5;
            else if cmonth = 'JUN' then xmonth = 6;
            else if cmonth = 'JUL' then xmonth = 7;
            else if cmonth = 'AUG' then xmonth = 8;
```

```
            else if cmonth = 'SEP' then xmonth = 9;
            else if cmonth = 'OCT' then xmonth = 10;
            else if cmonth = 'NOV' then xmonth = 11;
            else if cmonth = 'DEC' then xmonth = 12;
 xyear = substr(col_date,6,4);
 do_coll = mdy(xmonth,xday,xyear);
 run;

proc sort data=tempa; by idn tst_code do_coll; run;

data temp_first;
 set tempa;
 by idn tst_code;
 if first.tst_code; /* keep first occurence of ALT/AST for each patient */
 run;

/*************************************************************************/
/* Following code sets up low/high flags for ALT/AST for Table1          */
/*    using high and low indicators in Covance File                      */
/*    Generate seperate files for ALT/AST because on separate records in  */
/*    Covance Data Set                                                   */
/*************************************************************************/

data tempalt;
 set temp_first;
 if tst_code = 'RCT4';
            alt_cov = input(con_result,3.);
            alt_low = input(con_low,3.);
            alt_high = input(con_high,3.);
            if alt_low < alt_cov < alt_high then flg_alt_cov = 'OK';
            if . < alt_cov <= alt_low then flg_alt_cov = 'Low';
            if alt_high <= alt_cov then flg_alt_cov = 'Hi';
            logalt_cov = log10(alt_cov);
run;

data tempast;
 set temp_first;
 if tst_code = 'RCT5';
            ast_cov = input(con_result,3.);
            ast_low = input(con_low,3.);
            ast_high = input(con_high,3.);
            if ast_low < ast_cov < ast_high then flg_ast_cov = 'OK';
            if . < ast_cov <= ast_low then flg_ast_cov = 'Low';
            if ast_high <= ast_cov then flg_ast_cov = 'Hi';
            logast_cov = log10(ast_cov);
run;

proc sort data=tempalt; by idn; run;
proc sort data=tempast; by idn; run;

/*****************************************************************/
/* Merge ALT and AST data sets together to get everything on one  */
/* record                                                         */
/*****************************************************************/

data temp_cov;
 merge tempalt tempast;
 by idn;
```

```
 run;


proc sort data=temp_cov; by idn; run;

data temp_comb;
 merge tempall(in=in1) temp_cov(in=in2); /* merge with TEMPALL to get trtgrp and SVR */
 by idn;
 if in1;
 run;

proc freq data=temp_comb;
 table flg_alt_cov*origtxgrp*svr_wk72/ chisq;
 title 'Table 2: SVR Event Rates by Treatment Group within Subgroup';
 format origtxgrp $txf. svr_wk72 svrf.;
 run;
```

Appendix 3.

```
**/TABLES 1, 2, and 3 and Figure 2 in Schwarz et al, Gastroenterology 2011/**

cd "C:\Documents and Settings\smr\My Documents\PEDSC\DSIC\2012 DSIC\"

        /* use linkage file to match subject id in analysis file, IDN, with id in archived files, Masked_ID  */
use "linkage_patient.dta", clear
sort idn
keep idn masked_id
        save "link.dta", replace        /* 128 observations */

        /*  TABLE 1. Use OC_tab1 -- analysis file sent by DCC April 2012, 114 observations */
use "oc_tab1.dta", clear
        /*Tabulations for Table 1 based on SAS program 'Main Outcomes Table1.sas' provided by DCC  */
                        /*      Note: SAS programs use IDN not masked_id  */
recode race 1 2 3 6 7=1 5=2
label define trtgrp 1"PEG+RV" 2"PEG only"
label define GENOTYPE 1"1" 2"2" 3"3" 6"6"
label define agrp 1"5-11" 2"12-17"
label define gender 1"male" 2"female"
label define race 1"non white" 2"white"
recode hcvroute 1=1 4=2 2 3 5=3
label define hcvroute 1"maternal" 2"transf" 3"other"
label val trtgrp trtgrp
label val agrp agrp
label val race race
label val hcvroute hcvroute
sort idn
        /* IDN is used as ID in SAS analysis files -- masked_id is used in all other archive data */

 * Table 1.  Baseline characteristics according to Treatment Group
* t-test for continuous variables
ttest age, by(trtgrp)
ttest BMIZ, by(trtgrp)
ttest cdi_tot, by(trtgrp)
ttest duration, by(trtgrp)
                /* Baseline ALT, AST, HCVRNA do not match published results */
                /*  DCC notes: Baseline ALT, AST, and HCVRNA were done later - results from this step do not match Table 1
*/
sort trtgrp
by trtgrp: ameans alt
by trtgrp: ameans ast
by trtgrp: ameans hcvrna
ttest loghcv, by(trtgrp)
        merge 1:1 idn using "link.dta", generate(_merge2)
        save "oc_link.dta", replace

        *****Pathology Variables for Table 1 **********
        /* Use Histology file from archive file (non-analysis) */
use "histology.dta", clear
tab1 kno_pot kno_pri kno_pcy
gen hai=kno_pot+kno_pri+kno_pcy
tab hai
label var hai "histology activ index"
recode hai 1/3=1 4/6=2 7/9=3 10/12=4, gen(rechai) label(rechai)
label define rechai 1"minimal" 2"mild" 3"moderate" 4"marked"
```

```
tab hai rechai

tab fat
label define fat 0"none" 1"minimal" 2"mild"
label var fat "steatosis"
label val fat fat

tab1 ish_fib kno_fib met_fib
tab kno_fib met_fib
tab ish_fib kno_fib
label var kno_fib "fibrosis"
label define kno_fib 0"none" 1"portal_peri" 2"bridging" 3"cirrhosis", modify
label val kno_fib kno_fib
        sort masked_id
        keep masked_id kno_pot kno_pri kno_pcy hai rechai fat ish_fib kno_fib met_fib
                        /* n = 121  */

        /*merge with oc_link data to include trtgrp and IDN */
        merge 1:1 masked_id using "oc_link.dta", generate(_merge3)
tab rechai trtgrp, col
tab fat trtgrp, col
tab kno_fib trtgrp, col
        /*  Histology results match published results  */
sort idn
                        save "table1_path.dta", replace   /* 128 obs */


                /* Laboratory measures for Table 1 __Calculations of baseline ALT and AST and HCV RNA using archived
(non-analysis) files, covance_lab */
                        /*          DCC notes that analysis files do not replicate published findings */
                        /*  DSIC results using covance_lab similar to published but not exact match */
use "covance_lab.dta", clear
 tab tst_name
keep if visit_pvc=="BL"
keep if tst_name=="ALT (SGPT)" | tst_name=="AST (SGOT)" | tst_name=="HCVRNAQuant Cobas ED-RUO-CL-QT"
                        /* 342 observations */
encode tst_name, gen(newtest)
keep masked_id newtest visit con_result
reshape wide con_result,  i (masked_id) j(newtest)
label var con_result1 "ALT_bl"
label var con_result2 "AST_bl"
label var con_result3 "HCVRNA_bl"
tab1 con_result1 con_result2 con_result3
gen baseALT=real(con_result1)
label var baseALT "ALT derived from covance_lab data"
gen baseAST=real(con_result2)
label var baseAST "AST derived from covance_lab data"
gen basehcvrna=real(con_result3)
replace basehcvrna=599 if con_result3=="<600 IU/mL noHCVRNA detected"
replace basehcvrna=777777 if con_result3==">700000"
label define basehcvrna 599"<600" 777777">700000"
label val basehcvrna basehcvrna
tab basehcvrna
recode basehcvrna (599=1) (967/599999=1) (600000/777777=2), gen(bhcvrna600)
label define bhcvrna600 1 "<600000" 2 "600000+"
label val bhcvrna600 bhcvrna600
label var bhcvrna600 "baseline hcvrna level quan"
tab bhcvrna600
```

```
sort masked_id
        /*merge with table1_path data to include trtgrp and IDN */
        merge m:1 masked_id using "table1_path.dta", generate(_merge4)
tab bhcvrna600 trtgrp, col
ttest baseALT, by(trtgrp)
ttest baseAST, by(trtgrp)
ttest basehcvrna, by(trtgrp)
                        /* DSIC does not replicate Table 1 baseline estimates for ALT, AST, HCV RNA  */
keep if visit_pvc=="BL"
        save "TABLE_1", replace



/* TABLE 2 Virologic Results by Treatment Group and Baseline Features    */
        /* Tabulations for Table 2 below are derived from SAS code 'Main Outcomes Table 2.sas' provided by DCC */
        /* Note that SAS code includes recodes by idn for several variables, including svr_wk72, hcv48 */

        /* Per DCC: As noted, SVR = 'SVR' only if VR = 'VR' (Viral Response) at Weeks 24, 48,  */
        /*    and 72 as per Protocol - if there is no viral response, there is no SVR */

        /* Endpoint is DCC analysis dataset with SVR at week 72 */
use "endpoint.dta", clear
list idn origtxgrp vr_wk24 vr_wk72 svr_wk72
tab svr_wk72, missing                   /* valid measure for 93 obs, 41=1, 52=0, 21=.*/
        keep idn origtxgrp svr_wk72
        /*          DSIC Note:  svr_wk72 is primary outcome variable for Table 2
                    Following code supplied by DCC; recodes missing svr_wk72 to 0 except for 2 cases, early dropouts*/
recode svr_wk72 . = 0
replace svr_wk72 = . if idn=="061110" | idn=="110640"      /* DCC notes these subjects dropped out before week 24  */

        /* DSIC note: variable wd_wk, week at withdrawal, == 0 in endpoint.all */
list idn svr_wk72
        sort idn
                                                /*  N=114, 112 obs with svr measures, 41=1, 71=0, 2=.  */
                merge 1:1 idn using "TABLE_1.dta", generate(_merge5)

                sort origtxgrp idn
by origtxgrp: tab GENOTYPE svr_wk72, row chi2
by origtxgrp: tab gender svr_wk72, row chi2
by origtxgrp: tab agrp svr_wk72, row chi2
by origtxgrp: tab race svr_wk72, row chi2
by origtxgrp: tab rechai svr_wk72, row chi2
by origtxgrp: tab ish_fib svr_wk72, row chi2
by origtxgrp: tab fat svr_wk72, row chi2

                sort origtxgrp idn
by origtxgrp: tab bhcvrna600 svr_wk72, row chi2
recode baseALT 10/35=0 36/254=1, gen(altupper)
by origtxgrp: tab altupper svr_wk72, row chi2

                save "TABLE_2", replace


********** Figure 2.
use "hcv_all.dta", clear
 /* n=703 */
 keep if visit=="BL" | visit=="24" | visit=="48" |visit=="72" | visit=="12"  /* no data on visit at 5 weeks?? */
 by idn, sort: gen nvals4 = _n ==1
```

```
count if nvals4==1
        *n=111 observations
 /*  create rectangular file of hcv results  */
tab visit, missing
sort idn
drop in 296/296
keep idn visit con_result
        reshape wide con_result,  i (idn) j(visit) string       /* IDN 091040 has duplicate visit 72, same result different date

        drop earlier visit measurment */
                                        /*112 observations con_result12 con_result24 con_result48 con_result72 */

        /*  HCV Quantitative results -- the following tabulations as specified by DCC */
gen hcv_12=.
gen hcv_24=.
gen hcv_48=.
gen hcv_72=.
replace hcv_12 = 0 if con_result12=="Negative"   /* if HCV Quant result=negative */
replace hcv_12 = 1 if con_result12=="Positive"
replace hcv_24 = 0 if con_result24=="Negative"   /* if HCV Quant result=negative */
replace hcv_24 = 1 if con_result24=="Positive"
replace hcv_48 = 0 if con_result48=="Negative"   /* if HCV Quant result=negative */
replace hcv_48 = 1 if con_result48=="Positive"
replace hcv_72 = 0 if con_result72=="Negative"   /* if HCV Quant result=negative */
replace hcv_72 = 1 if con_result72=="Positive"


replace hcv_48 = 1 if hcv_24==1
replace hcv_72 = 1 if hcv_24==1 & hcv_48==1
tab1 hcv_24 hcv_48 hcv_72, missing
                                        /* DCC Note: If HCV Quant Result = Positive then set later */
                                                                /* visits also to Positive - indicating no Viral */
                                                                /* Response and no SVR (per Protocol)          */
                                                                /* SVR only if Result = Negative at Week 24, 48, */
                                                                /* and 72                                      */
tab1 con_result24 con_result48 con_result72 hcv_24 hcv_48 hcv_72
list idn con_result24 con_result48 con_result72 hcv_24 hcv_48 hcv_72

         /* PER DCC: special handling for some early drop-outs */
tab hcv_72, missing
  replace hcv_72=1 if hcv_72==.
        replace hcv_72=. if idn=="061110" | idn =="110640"
tab hcv_24, missing
  replace hcv_24=1 if hcv_24==. & hcv_72==1
tab hcv_48, missing
  replace hcv_48=1 if hcv_48==. & hcv_72==1
list idn con_result24 con_result48 con_result72 hcv_24 hcv_48 hcv_72

  replace hcv_48=1 if idn=="110750"
  replace hcv_48=0 if idn=="061290"
  replace hcv_24=1 if idn=="070970"
  replace hcv_48=1 if idn=="070970"

 tab1 con_result24 con_result48 con_result72 con_result12 hcv_24 hcv_48 hcv_72 hcv_12
                merge 1:1 idn using "TABLE_2.dta", generate(_merge7)

        /*  FIGURE 2. PERCENT OF PATIENTS WITH NO DETECTABLE VIRUS  */
tab origtxgrp svr_wk72, row chi2
```

```
tab origtxgrp hcv_48, row chi2
tab origtxgrp hcv_24, row chi2
tab origtxgrp hcv_12, row chi2

        save "Figure_2", replace


        /* FOR TABLE 3. SVR Event Rates per DCC SAS program */
        use "vk_jhu_fine.dta", clear
        /*  Per DCC: Quantitative HCV Results from JHU lab to indicate level of viral response in manuscript */
by idn, sort: gen nvals = _n ==1
count if nvals==1
   *n=114 subjects with multiple observations
gen vk=.
replace vk=0 if results=="Target Not Detected"
replace vk=12.5 if substr(results,1,1)=="<"
tab vk, missing
        /* 636 missing values */
        gen vk1=substr(results,1,4) if vk==.
        gen vk1R=real(vk1)
           format vk1R %4.2f
        gen vk2=substr(results,7,1) if vk==.
        gen vk2R=real(vk2)
                format vk2R %2.0f
list idn visit results vk vk1R vk2R in 1/100

gen vkR=vk1R*10^vk2R if vk==.
          gen logvk=log10(vkR)
            replace logvk=0 if vk==0
          replace logvk=. if vk==12.5
*tab1 vk logvk, missing
list idn visit results vk vk1R logvk in 1/100

gen hcv600=.
replace hcv600=1 if vk > 600000
replace hcv600=0 if vk <= 600000
tab hcv600
sort idn visit
keep if visit=="BL" | visit=="1" | visit=="3" |visit=="5" | visit=="12" | visit=="24"
        /* NOTE: subj idn 051240 has 2 visits at 24 mos--both with 'Target not Detected'
                        for this analysis the earlier measurement dated 08 Nov 06 was dropped
idn          labdt   visit      sampid     batchid   results
051240   08 Nov 06          24         222768678          1040      Target Not Detected
051240   21 Nov 06          24         2315791520         1040      Target Not Detected                    */
drop if idn == "051240" & sampid == "222768678"
        keep idn vkR visit logvk

drop in 175/175             /* multiple visit 1 for subj IDN 030550 and 030960, drop earlier result */
drop in 156/156
        /*      idn         visit      logvk
                030960   1   5.687529
                030960   1          4.778151

                030550   1          6.509202
                030550   1          6.465383
                  */
        reshape wide vk logvk,  i (idn) j(visit) string
        /* creates new variables logvk1 logvk3 logvk5 logvk12 logvkBL
```

```
                        113 obs  */

sort idn

*gen vkbl_nolog=.
*replace vkbl_nolog=10^logvk if visit=="BL"
gen vkdiff1 = logvkBL - logvk1
gen          vkdiff5 = logvkBL - logvk5
gen vkdiff12 = logvkBL - logvk12
*tab1 vkdiff1 vkdiff5 vkdiff12

list idn logvkBL logvk1 vkdiff1 logvk5 vkdiff5 logvk12 vkdiff12


/* Per DCC: Data step to calculate changes from baseline for Table 3: 1 log drop at Week 1, */
/*     RVR (no detectable virus) at Week 5, and EVR (2-log drop) at Week 12       */

  merge 1:1 idn using "TABLE_2.dta", generate(_merge8)
  tab svr_wk72, missing
*replace svr_wk72=0 if svr_wk72==. & logvk24==0              /* NOTE: This statement is not clear--replaces definition of svr in
Table 2*/
gen rvr=.
*tab logvk5, missing
replace rvr = 0 if logvk5 !=.
replace rvr=1 if logvk5==0
tab rvr, missing

gen evr=.
*tab vkdiff12, missing
replace evr=0 if vkdiff12 !=.
replace evr=1 if vkdiff12 >= 2.0
replace evr=. if vkdiff12==.
list idn evr vkdiff12 in 1/20
tab evr, missing

gen vklog1=.
replace vklog1=0 if logvkBL !=. & logvk1 !=.
replace vklog1=1 if logvkBL - logvk1 >= 1.0
replace vklog1=. if logvkBL==. | logvk1==.
list idn vklog1 logvkBL logvk1
tab vklog1, missing

gen vklog2=.
replace vklog2=0 if logvkBL !=. & logvk1 !=.
replace vklog2=1 if logvkBL - logvk1 >= 2.0
replace vklog2=. if logvkBL==. | logvk1==.
list idn vklog2 logvkBL logvk1 in 1/20
tab vklog2, missing

gen vklog12=.
replace vklog12=0 if logvkBL !=. & logvk12 !=.
replace vklog12 = 1  if logvkBL - logvk12 >= 1.0
replace vklog12 = 2  if logvkBL - logvk12 >= 2.0
replace vklog12=. if logvkBL==. | logvk12==.
list idn vklog12 logvkBL logvk12 in 1/20
tab vklog12, missing

gen vr12=.
```

```
*tab logvk12, missing
replace vr12 = 0 if logvk12 !=.
replace vr12 = 1 if logvk12 == 0
replace vr12 = . if logvk12==.
list idn vr12 logvk12 in 1/20
tab vr12, missing

            /*  Stata code above to match SAS Code below from DCC ;
 if in1;
 if svr_wk72 = . and vr_wk24 = 0 then svr_wk72 = 0;
 if vk5 ne . then rvr = 'No';
 if vk5 = 0.0 then rvr = 'Yes';
 if vkdiff12 ne . then evr = 'No';
 if vkdiff12 >= 2.0 then evr = 'Yes';
 if (vkbl ne . and vk1 ne .) then vklog1 = 'No';
 if (vkbl ne . and vk1 ne .) then vklog2 = 'No';
 if (vkbl - vk1) >= 1.0 then vklog1 = 'Yes';
 if (vkbl - vk1) >= 2.0 then vklog2 = 'Yes';
 if (vkbl ne . and vk12 ne .) then vklog12 = 'No';
 if (vkbl - vk12) >= 1.0 then vklog12 = '1 Log';
 if (vkbl - vk12) >= 2.0 then vklog12 = '2 Log';
 if vk12 ne . then vr12 = 'No';
 if vk12 = 0.0 then vr12 = 'Yes';
 if vkbl_nolog ne . then vk600k = 'No';
 if vkbl_nolog >= 600000 then vk600k = 'Yes';
 label evr = 'Early Viral Response (Week 5)'
                        rvr = 'Rapid Viral Response (Week 5)'
                        vklog1 = 'Week 1 Log Drop'
                        vklog12 = 'Week 12 Log Drop'
                        vr12 = 'Viral Response Week 12'; */

label var evr "Early Viral Response (Week 5)"
 label var rvr "Rapid Viral Response (Week 5)"
label var               vklog1 "Week 1 Log Drop"
label var               vklog12 "Week 12 Log Drop"
label var               vr12 "Viral Response Week 12"
tab1 rvr evr vklog1 vklog12 vr12
list idn logvk5 rvr evr vkdiff12 vklog1 logvkBL logvk1 vklog2 vklog12 logvk12 vr12 in 1/20

tab GENOTYPE
                        /*n=114 */

 ** SELECT genotype 1 patients only for Table 3
keep if GENOTYPE=="1"
            /* 92 observations */
 tab1 vklog1 trtgrp origtxgrp svr_wk72 rvr evr
tab origtxgrp vklog1, row
sort origtxgrp
label define origtxgrp 1"PEG+RV" 2"Mono"
list idn logvkBL logvk1 vkdiff1 vklog1
by origtxgrp: tab vklog1 svr_wk72, row col exact
tab origtxgrp evr, row
by origtxgrp: tab evr svr_wk72, row col exact
tab origtxgrp rvr, row
by origtxgrp: tab rvr svr_wk72, row col exact

*** EVR and RVR results similar, but not week 1 log drop
```

```
save "TABLE_3", replace
```