

**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:

206426Orig1s000

STATISTICAL REVIEW(S)

Statistical Review and Evaluation CARCINOGENICITY STUDIES



IND/NDA Number: NDA 206426
Drug name: Rapivab
Indication(s): Treatment of acute uncomplicated influenza in patients 18 years and older
Applicant: Biocryst Pharmaceuticals Inc.
Documents Reviewed: Electronic submission
Electronically submitted dataset
Dated: 2014-02-28

Biometrics Division: Division of Biometrics 6
Statistical Reviewer: Matthew Jackson, PhD
Concurring Reviewer: Karl Lin, PhD
Medical Division: Division of antiviral products
Reviewing Pharmacologist: Kuei-Meng Wu, PhD
Project Manager: Elizabeth Thompson

Keywords: Animal Rat Carcinogenicity

Contents

1	Summary of findings	5
2	Rat Study	6
2.1	Experimental design	6
2.2	Sponsor's analysis	6
2.2.1	Survival analysis	6
2.2.2	Tumor analysis	6
2.3	Data analysis	6
2.3.1	Survival analysis	6
2.3.2	Tumor analysis	11
2.3.3	Analysis of unexamined and autolytic organs	12
3	Assessment of the validity of a negative study	14
3.1	Issues of concern when selecting the dose levels	14
3.2	Assessment of the validity of the rat study	15
A	Tables from rat study	16
A.1	Survival analysis	16
A.2	Tumor analysis	21
A.3	Unexamined and autolytic organs	38
A.4	Weight changes	43

List of Tables

2.1	Critical p -values used to determine statistical significance	12
A.1	Numbers of animals alive at certain timepoints (rat study)	17
A.2	Results of log-rank tests of survival across all groups (rat study)	18
A.3	Results of pairwise log-rank tests of survival between treated groups and vehicle control (rat study)	19
A.4	Results of log-rank test of survival across control groups	20
A.5	Primary organs in female rat experiment	22
A.6	Primary organs in male rat experiment	23
A.7	Secondary organs in female rat experiment	24
A.8	Customized endpoints analyzed	25
A.9	Tumors reported in female rat experiment	26
A.10	Tumors reported in male rat experiment	29
A.11	Combination tumors reported in female rat experiment	33
A.12	Combination tumors reported in male rat experiment	34
A.13	Tumors reported significant in female rat experiment	35
A.14	Combination tumors reported significant in female rat experiment	36
A.15	Tumors reported secondary organs (rat study)	37
A.16	Organs reported autolytic in female rat experiment	39
A.17	Organs reported autolytic in male rat experiment	40
A.18	Organs reported unexamined in female rat experiment	41
A.19	Organs reported unexamined in male rat experiment	42
A.20	Weight changes by group (rats)	43

List of Figures

2.1	Survival curves for female rats	7
2.2	Survival curves for male rats	8
2.3	Survival curves for control groups (female rat experiment)	9
2.4	Survival curves for control groups (male rat experiment)	10
2.5	Autolysis rates for male rats	13

Background

In this submission the sponsor included reports of one animal carcinogenicity study, in rats, to assess the carcinogenic potential of Rapivab when administered by gavage, once daily at appropriate drug levels for about 104 weeks. Results of this review have been discussed with the reviewing pharmacologist, Kuei-Meng Wu, PhD.

In this review, the phrase “dose response relationship” refers to the linear component of the effect of treatment, and not necessarily to a strictly increasing or decreasing mortality or tumor incidence rate as dose increases.

Chapter 1

Summary of findings

Both a study of rats and a study of mice were conducted. However, since data were only submitted for the rat study, and this review is a review of the rat study alone.

The rat study was negative. There are no indications of any tumorigenic effect for any of the endpoints tested. However, the Zymbal's gland was not routinely examined in either male or female animals, so the study should be considered inconclusive rather than negative for tumors associated with this organ.

There is no evidence of a dose related effect on survival. Among male rats, there does appear to be a dose related reduction in weight gain, suggesting that the high dose level was indeed sufficiently close to the MTD that some animals started experiencing some toxicity effects. However, there is no such evidence for the female rats. Accordingly, the survival and weight gain data cannot allow us to say with any certainty that the high dose level posed an adequate challenge to the female animals.

In the male experiment, the autolysis rates were quite high for some intestinal organs (ileum, jejunum, cecum). The fact that the autolysis rates seems to increase with dose means that even though there is no evidence of a positive dose effect for intestinal tumors, we should still treat the experiment as inconclusive rather than negative for these endpoints. Aside from this, there were no problems with either autolysis or unexamined organs in either the female or male rat experiments.

There were some data quality issues with the data submitted for the female rat experiment. Specifically, the sponsor made several coding errors in the data. Those that were most striking have been queried and corrected, but the number of errors raises doubts about the quality of the rest of the data: a minor coding error would not necessarily attract the attention of a review, and so might not be queried with the sponsor.

Chapter 2

Rat Study

2.1 Experimental design

This study comprised two separate experiments; one on male rats and one on female rats. In each case, 260 animals were used, allocated to four groups of 65: The vehicle control group, the low dose group, the mid dose group, and the high dose group. These groups received, by gavage, daily doses of 0mg/kg, 150mg/kg, 1000mg/kg and 3000mg/kg of Rapivab, in a dose of 20mL/kg of the vehicle (distilled water).

For each sex, there was an additional group of untreated animals; data from these animals played no part in the statistical review.

All animals surviving to 104 weeks were sacrificed. All animals, regardless of cause of death, underwent a complete necroscopy.

2.2 Sponsor's analysis

2.2.1 Survival analysis

The sponsor found no indication of a dose effect on mortality for the female rats. For the male rats, the vehicle control animals were found to experience significantly higher mortality rates than the untreated control ($p = 0.0362$), but no difference was noted between the groups receiving the vehicle (with or without Rapivab).

2.2.2 Tumor analysis

In neither sex were any tumor types or combinations found to increase with dose.

2.3 Data analysis

2.3.1 Survival analysis

The Kaplan-Meier survival plots are shown as figures 2.1 and 2.2. The numbers and proportions of animals surviving to various times are presented in table A.1. The results of log-rank tests of heterogeneity of survival and of dose response across the groups are presented in table A.2, and the results of log-rank survival tests comparing the treated groups with the vehicle control group are presented in table A.3.

Commentary Survival rates were generally good, with between 29 (45%) and 36 (55%) animals in each group surviving until the scheduled sacrifice. There was no indication of a dose response in mortality in either sex.

Figure 2.1

Kaplan-Meier survival plot
Animal carcinogenicity study
NDA 206426
Rats - Female

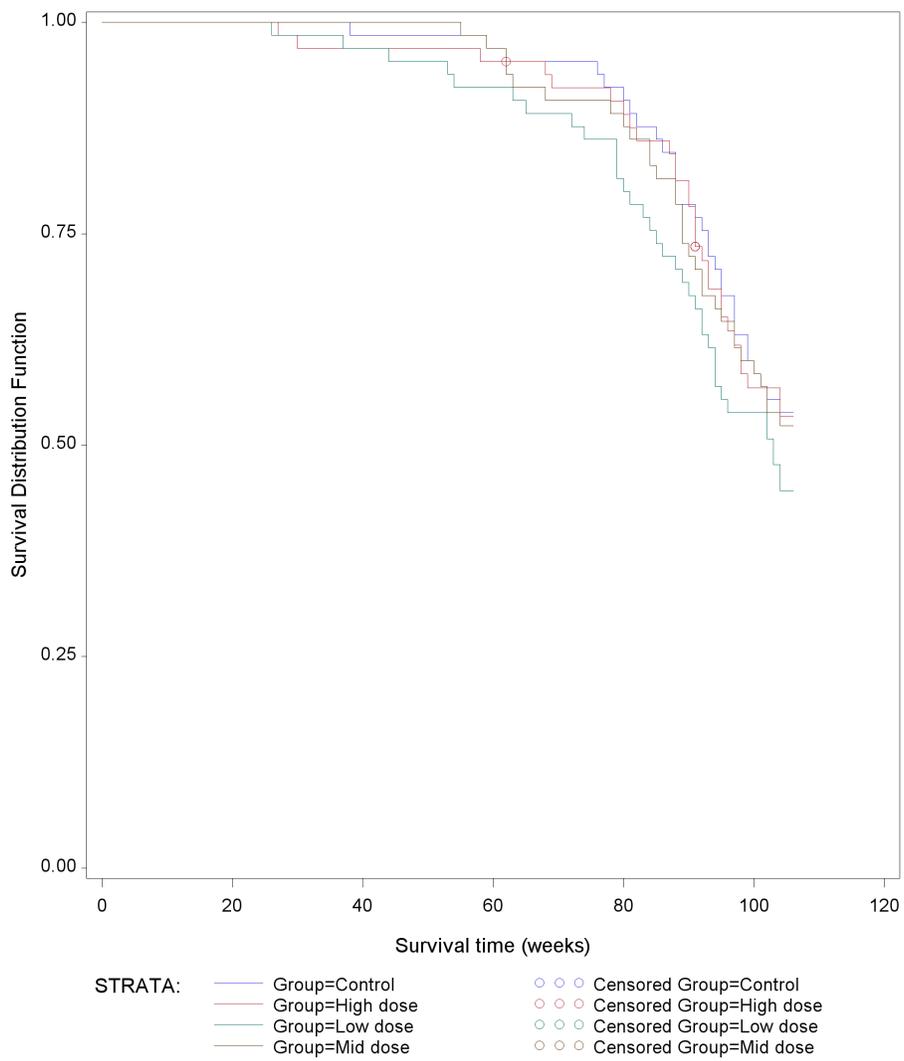
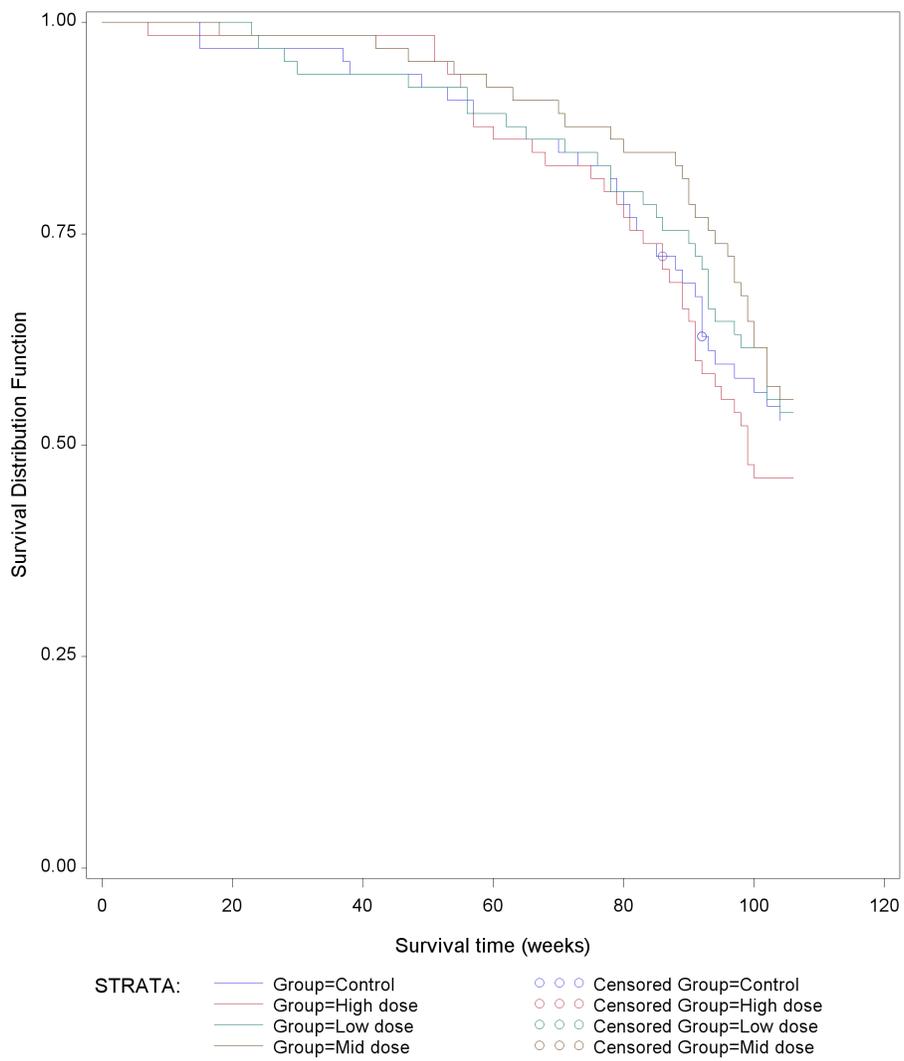


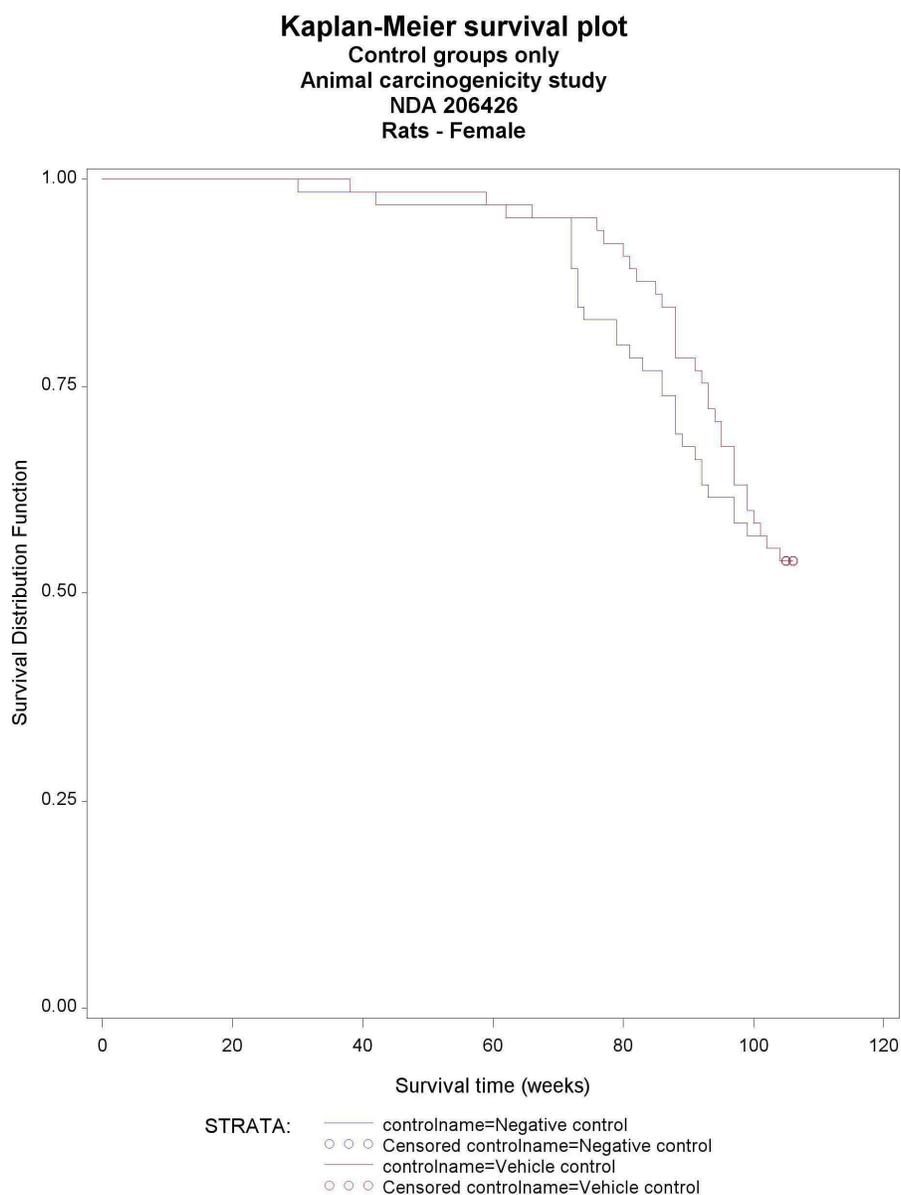
Figure 2.2

Kaplan-Meier survival plot
Animal carcinogenicity study
NDA 206426
Rats - Male



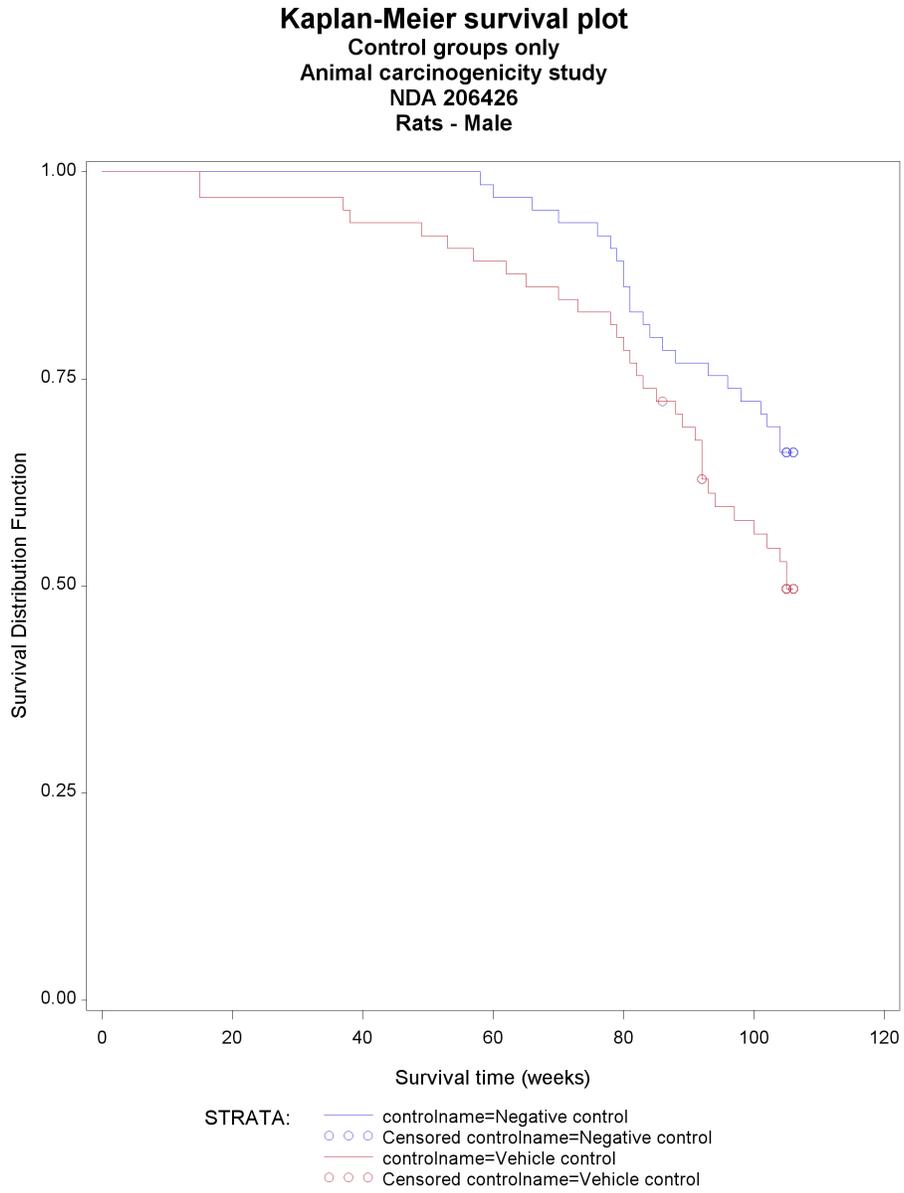
Comparison of control groups Kaplan-Meier plots of the control groups are shown as figures 2.3 and 2.4. The results of log-rank tests of survival between the control groups are presented in table A.4.

Figure 2.3: Survival curves for control groups (female rat experiment)



The Kaplan-Meier plots are interesting, suggesting that the for female animals, the vehicle control group had better survival rates than the negative control group, but that for male rats, it was the negative control group who had better survival than the vehicle controls. In the case of the female animals, however, this result is not even close to being statistically significant. For males, the difference is a close to statistical significance ($p = 0.0532$).

Figure 2.4: Survival curves for control groups (male rat experiment)



2.3.2 Tumor analysis

Endpoints

Analyses have been conducted using the sponsor’s submitted dataset, and the sponsor’s chosen nomenclature. In this dataset, organs or tissue types are described as being either tumorous, examined but found unusable due to autolysis, or unexamined. An organ that has been examined but was not found to be tumorous is not mentioned in the dataset.

From these data, we can infer the numbers of animals for which each organ or tissue type was examined, but only in those cases where at least one anomalous finding (i.e., a tumor was found, or a sample that was planned to be analyzed could not be, either because no sample was taken or because the sample was unusable due to autolysis) was reported. Organs which can thus be deduced to have been successfully analyzed in the majority of animals are, for the purposes of this review, considered *primary*. The lists of primary organs in the experiments on female and male mice respectively are presented in tables A.5 and A.6.

Organ or tissue types which were examined in only a few animals are considered *secondary*.

Organs identified as secondary in the female rat experiment are presented in table A.7. No secondary organs were identified in the male rat experiment.

Each tumor type found in a primary organ of at least one animal is considered a primary endpoint. In addition, in consultation with Kuei-Meng Wu, PhD, a list of combination endpoints has been drawn up. This list is presented in table A.8.

Statistical procedure

The tumor data were analyzed for dose response relationships and pairwise comparisons of tumor incidence in each of the treated groups versus the vehicle control group. Both the dose response relationship tests and pairwise comparisons were performed using the poly- k method described in the paper of Bailer and Portier[?] and developed in the paper of Bieler and Williams[?]. In this method, given a tumor type T , an animal h that lives the full study period (w_m) or dies before the terminal sacrifice with at least one tumor of type T gets a score of $s_h = 1$. An animal that dies at week w_h before the end of the study without such a tumor gets a score of

$$s_h = \left(\frac{w_h}{w_m} \right)^k < 1.$$

The adjusted group size is defined as $\sum_h s_h$. As an interpretation, an animal with score $s_h = 1$ can be considered as a whole animal while an animal with score $s_h < 1$ can be considered as a partial animal. The adjusted group size $\sum s_h$ is equal to N (the original group size) if all animals live up to the end of the study or if each animal develops at least one tumor of type T , otherwise the adjusted group size is less than N . These adjusted group sizes are then used for the dose response relationship (or the pairwise) tests using the Cochran-Armitage test. The test is repeated for each tumor type T .

One critical point to consider in the application of the poly- k test is the choice of the appropriate value of k , which depends on the relationship between tumor onset time and increased dose. For long term 104 week standard rat and mouse studies, a value of $k = 3$ is suggested in the literature, and so has been used in this review. For the calculation of p -values, the exact permutation method was used.

For the adjustment of multiple testing of dose response relationship, the FDA guidance for the carcinogenicity study design and data analysis suggests the use of significance levels $\alpha = 0.005$ for common tumors and $\alpha = 0.025$ for rare tumors for a submission with two species, and a significance level $\alpha = 0.01$ for common tumors and $\alpha = 0.05$ for rare tumors for a submission with one species study in order to keep the false-positive rate at the nominal level of approximately 10%. A rare tumor is defined as one in which the published spontaneous tumor rate is less than 1%. For multiple pairwise comparisons of treated group with control, the FDA guidance suggests the use of test levels $\alpha = 0.01$ for common tumors and $\alpha = 0.05$ for rare tumors, for both submissions with one or two species, in order to keep the false-positive rate at the nominal level of approximately 10%.

It should be noted that the FDA guidance for multiple testing for dose response relationship is based on a publication by Lin and Rahman [?]. In this work the authors investigated the use of this rule for Peto analysis. However, in a later work Rahman and Lin [?] showed that this rule for multiple testing for dose response relationship is also suitable for poly- k tests.

Since this is a study involving two species, it follows that for the comparisons of Rapivab with vehicle control, we use the thresholds for significance presented in table 2.1.

Table 2.1: Critical p -values used to determine statistical significance

Type of test	Rare tumor	Common tumor
Trend	0.025	0.005
Pairwise test between placebo and high dose	0.10	0.05

The results of the statistical analyses of tumor incidence in primary endpoints are presented in tables A.9 (female rats) and A.10 (male rats). The results of analyses of customized endpoints (see table A.8) are presented in tables A.11 and A.12.

Noteworthy results

Individual tumor types in female rats for which tests yielding p -values below 0.05 were conducted are presented in table A.13, which is excerpted from table A.9. Combination tumor types for which tests yielding p -values below 0.05 were conducted are presented in table A.14, which is excerpted from table A.11. No statistical tests were conducted in the male rat experiment which resulted in p -values below 0.05.

Incidence rates for tumors found in secondary organs have not been analyzed statistically. Count data for such tumors are presented in table A.15.

Pituitary tumors in female rats The only tests that yielded p -values below 0.05 were for pituitary tumors in female rats. In this case, the test of interest was the pairwise comparison between the vehicle control and low dose groups. However, given that neither the trend test nor either of the other pairwise tests yielded worrying p -values, there is no reason to consider this a positive finding.

2.3.3 Analysis of unexamined and autolytic organs

Unexamined animals

Many female rats were initially reported as being completely unexamined. However, after corresponding with the sponsor, this was determined to be a coding error. After corrections were made, no animals (of either sex) were reported as being completely unexamined.

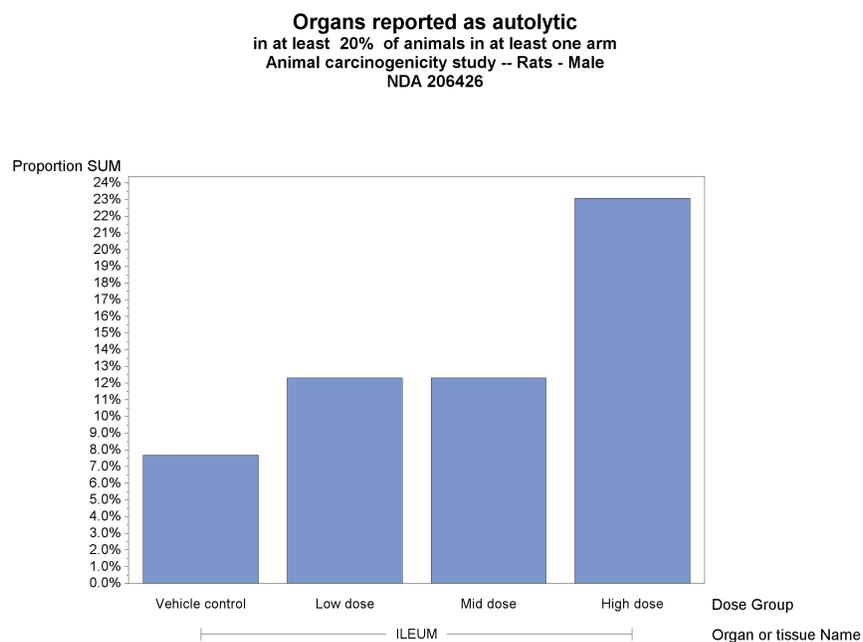
Organs reported autolytic

The numbers of organs found in female rats to be autolytic to the extent that analysis of collected tissue was not possible are presented in table A.16. No organs in male rats were found to be autolyzed to this extent.

There were no problems with excessive autolysis rates in the female rat experiment.

In the male rat experiment, the autolysis rates for the various intestinal organs (primarily the jejunum, ileum, and cecum) were slightly higher than would typically be expected and, rather more worryingly, were associated with dose. See Figure 2.5. Consequently, the experiment should be considered inconclusive rather than negative for tumors of these endpoints.

Figure 2.5



Organs reported as unexamined

The numbers of animals with organs reported as being unexamined are presented in tables A.18 and A.19.

Among female rats, all animals without systemic tumors were originally listed as having “HEMATO NEOPLASIA” unexamined. This was presumed to be another coding error. Aside from this, there were no problems with high rates of unexamined organs. However, it should be noted that in both the female and male rat experiments, most animals did not have their Zymbal’s glands examined, so tumors of this organ should be considered outside the scope of this study.

Chapter 3

Assessment of the validity of a negative study

3.1 Issues of concern when selecting the dose levels

The selection of an appropriate dose level for the high dose group is made difficult by the need to satisfy two competing imperatives: on the one hand, if the dose level is insufficiently high, then genuine carcinogenicity effects may not be apparent, but on the other hand, if the dose level is too high, then there is a risk of non-carcinogenic toxic effects killing the animals before they have a chance to demonstrate a carcinogenicity effect.

Haseman [?] suggested that a satisfactory balance between these two imperatives has been found when the following two conditions are both satisfied:

1. Were enough animals exposed, for a sustained amount of time, to the risk of late developing tumors?
2. Were dose levels high enough to pose a reasonable tumor challenge to the animals?

There is no consensus among experts regarding the number of animals and length of time at risk, although most carcinogenicity studies are designed to run for two years with fifty animals per treatment group. The following are some rules of thumb regarding these two issues as suggested by experts in this field:

Haseman [?] has done an investigation on the first issue. He gathered data from 21 studies using Fischer 344 rats and B6C3F1 mice conducted at the National Toxicology Program (NTP). It was found that, on the average, approximately 50% of the animals in the high dose group survived the two year study period. Also, in a personal communication with Dr. Karl Lin of Division of Biometrics-6, Haseman suggested that, as a rule of thumb, a 50% survival of 50 initial animals or 20 to 30 animals still alive in the high dose group, between weeks 80–90, would be considered as a sufficient number and adequate exposure. In addition Chu, Cueto and Ward [?], suggested that “to be considered adequate, an experiment that has not shown a chemical to be carcinogenic should have groups of animals with greater than 50% survival at one year.”

It appears, from these three sources that the proportions of survival at 52 weeks, 80–90 weeks, and two years are of interest in determining the adequacy of exposure and number of animals at risk.

Regarding the question of adequate dose levels, it is generally accepted that the high dose should be close to the maximum tolerated dose (MTD). In the paper of Chu, Cueto and Ward [?], the following criteria are mentioned for dose adequacy. A high dose is considered as close to MTD if any of the criteria is met:

1. A dose is considered adequate if there is a detectable loss in weight gain of up to 10% in a dosed group relative to the controls.

2. The administered dose is also considered an MTD if dosed animals exhibit clinical signs or severe histopathologic toxic effects attributed to the chemical.
3. In addition, doses are considered adequate if the dosed animals show a slight increased mortality compared to the controls.

3.2 Assessment of the validity of the rat study

In both sexes, the survival rates were very good, and there is no concern that excess toxicity has reduced the animals' survival rates below an acceptable level. However, neither the female or male rats exhibited any dose related reduction in survival. The male rats did exhibit a dose related reduction in weight gain (Table A.20), but no such effect was observed in the female rat experiment. There is therefore some concern that the female animals did not receive a sufficiently high dose. The assessment of this matter is beyond the scope of this statistical review.

Appendix A

Tables from rat study

A.1 Survival analysis

Table A.1

Survival rates at key times
NDA 206426
Animal carcinogenicity study
Rats

<i>Species and Sex</i>	<i>Dose Group</i>	<i>Dose (mg per kg)</i>	<i>Number at start</i>	<i>Number alive after 52 weeks</i>	<i>Percentage alive after 52 weeks</i>	<i>Number alive after 78 weeks</i>	<i>Percentage alive after 78 weeks</i>	<i>Number alive after 90 weeks</i>	<i>Percentage alive after 90 weeks</i>	<i>Number sacrificed</i>	<i>Percentage sacrificed</i>	<i>Maximum survival (weeks)</i>
Rats - Female	Vehicle control	0	65	64	98%	60	92%	51	78%	35	54%	106
	Negative control	0	65	63	97%	54	83%	44	68%	35	54%	106
	Low dose	150	65	62	95%	56	86%	45	69%	29	45%	106
	Mid dose	1000	65	65	100%	59	91%	48	74%	34	52%	106
	High dose	3000	65	63	97%	59	91%	52	80%	32	49%	106
Rats - Male	Vehicle control	0	65	60	92%	54	83%	44	68%	30	46%	106
	Negative control	0	65	65	100%	60	92%	50	77%	43	66%	106
	Low dose	150	65	60	92%	54	83%	49	75%	35	54%	106
	Mid dose	1000	65	62	95%	57	88%	53	82%	36	55%	106
	High dose	3000	65	62	95%	52	80%	43	66%	30	46%	106

Table A.2

Log-rank tests of survival
NDA 206426
Animal carcinogenicity study
Rats

Sex	Test of homogeneity: chi squared statistic	Test of homogeneity: degrees of freedom	Number of groups	Test of homogeneity: p-value	Test of trend (two tailed): p-value	Test of trend (one tailed): p-value
Female	2.0552	4	5	0.7256	0.6353	0.6824
Male	6.6997	4	5	0.1526	0.1070	0.0535

Table A.3

Pairwise comparisons (log-rank) of survival between treated groups and controls
NDA 206426
Animal carcinogenicity study
Rats

<i>Species and Sex</i>	<i>Quantity</i>	<i>Low dose</i>	<i>Mid dose</i>	<i>High dose</i>
Rats - Female	Chi squared test statistic	1.6042	0.0793	0.0135
	p-value of comparison with control	0.2053	0.7782	0.9074
Rats - Male	Chi squared test statistic	0.2782	0.7911	0.2248
	p-value of comparison with control	0.5979	0.3738	0.6354

Table A.4

**Log-rank tests of heterogeneity of survival between control groups
NDA 206426
Animal carcinogenicity study**

<i>Species and Sex</i>	<i>Chi²</i>	<i>DF</i>	<i>P-value</i>
Rats - Female	0.1285	1	0.7200
Rats - Male	3.7381	1	0.0532

A.2 Tumor analysis

Table A.5

**Primary organs in study of female rats
NDA 206426
Animal carcinogenicity study**

<u>Organ or tissue name</u>
ADRENAL, CORTEX
ADRENAL, MEDULLA
BONE, FEMUR
BONE, STERNUM
BRAIN
CERVIX
COLON
CORD, THORACOLUMB
DUODENUM
EYE
HEART
HEMATO NEOPLASIA
ILEUM
JEJUNUM
KIDNEY
LIVER
LN, MAND BULAR
MAMMARY
MARROW, FEMUR
MARROW, STERNUM
MUSCLE, SKELETAL
NERVE, OPTIC
NERVE, SCIATIC
OVARY
PANCREAS
PARATHYROID
PITUITARY
THYMUS
THYRO D
UTERUS
<u>VAGINA</u>

Table A.6

**Primary organs in study of male rats
NDA 206426
Animal carcinogenicity study**

<i>Organ or tissue name</i>
ADRENAL, CORTEX
ADRENAL, MEDULLA
AUDITORY SEB GL
BRAIN
CAVITY, ABDOM
CAVITY, THORACIC
CECUM
COAGULATING GL
COLON
CORD, THORACOLUMB
DUODENUM
EPID DYMIS
ESOPHAGUS
EYE
HEART
HEMATO NEOPLASIA
ILEUM
JEJUNUM
KIDNEY
LN, MESENTERIC
LUNG
MAMMARY
MARROW, FEMUR
MEDIAST NUM
NERVE, OPTIC
PANCREAS
PARATHYROID
PITUITARY
RECTUM
SALIV GL, MANDIB
SEM NAL VESICLE
SKIN
SKIN, OTHER
STOMACH, NONGL
TESTIS
THYMUS
THYRO D
TRACHEA

Table A.7

**Secondary organs in study of female rats
NDA 206426
Animal carcinogenicity study**

<u>Organ or tissue name</u>
BONE, OTHER
CAVITY, ABDOM
CAVITY, ORAL
CAVITY, THORACIC
HEAD, CORONAL
LN, OTHER
NASAL TURBINATE
PINNA
SK N, OTHER
SUBCUTANEOUS TIS
ZYMBAL'S GLAND

Table A.8

**Customized and combination endpoints analyzed
NDA 206426
Animal carcinogenicity study**

<i>Composite endpoint</i>
Adenomas and carcinomas of the pituitary
All astrocytomas
All leiomyosarcomas
All papillomas
All schwannomas including neurofibrosarcomas
All squamous cell tumors (including keratoacanthoma)
Carcinomas and adenomas of the adrenal cortex
Endometrial tumors of the uterus
Follicular cell adenomas and carcinomas of the thyroid
Hemangiomas and hemangiosarcomas
Histiocytic sarcomas
Leiomyosarcomas and Leiomyomas
Sarcomas of skin and skin (other)

Table A.9

**Table of reported tumors in Rat Study
NDA 206426
Animal carcinogenicity study
Female rats**

<i>Organ or tissue name</i>	<i>Tumor name</i>	<i>Quantity</i>	<i>Vehicle control Size = 65</i>	<i>Low dose Size = 65</i>	<i>Mid dose Size = 65</i>	<i>High dose Size = 65</i>
ADRENAL, CORTEX	ADENOMA	P-value of test of trend or comparison	.7871	.7276	.4857	1
		Number of animals reported with tumor	1	1	2	0
	CARCINOMA	P-value of test of trend or comparison	.3901	1	.7477	.7379
		Number of animals reported with tumor	1	0	1	1
ADRENAL, MEDULLA	PHEOCHROMOCYTOMA	P-value of test of trend or comparison	.0604			.2335
		Number of animals reported with tumor	0	0	0	2
BRAIN	ASTROCYTOMA	P-value of test of trend or comparison	.0628			.2383
		Number of animals reported with tumor	0	0	0	2
CERVIX	GRANULAR CELL TUMOR	P-value of test of trend or comparison	.7379	.4757		
		Number of animals reported with tumor	0	1	0	0
	LEIOMYOMA	P-value of test of trend or comparison	1	1	1	1
		Number of animals reported with tumor	1	0	0	0
	LEIOMYOSARCOMA	P-value of test of trend or comparison	.2476			.4857
		Number of animals reported with tumor	0	0	0	1
	SCHWANNOMA	P-value of test of trend or comparison	.4358	.2938	1	.4727
		Number of animals reported with tumor	2	4	0	3
EYE	MELANOMA	P-value of test of trend or comparison	.7379	.4757		
		Number of animals reported with tumor	0	1	0	0
HEART	SCHWANNOMA	P-value of test of trend or comparison	1	1	1	1
		Number of animals reported with tumor	1	0	0	0
HEMATO NEOPLASIA	LYMPHOMA	P-value of test of trend or comparison	.7536	1	.7477	1
		Number of animals reported with tumor	1	0	1	0
	SARCOMA, HISTIOCYTIC	P-value of test of trend or comparison	.1187	.4757	.4953	.2383
		Number of animals reported with tumor	0	1	1	2
KIDNEY	ADENOMA, TUBULAR CELL	P-value of test of trend or comparison	.2476			.4857
		Number of animals reported with tumor	0	0	0	1
LIVER	ADENOMA, HEPATOCELLULAR	P-value of test of trend or comparison	.5915	1	1	.8677

Table A.9

**Table of reported tumors in Rat Study
NDA 206426
Animal carcinogenicity study
Female rats**

<i>Organ or tissue name</i>	<i>Tumor name</i>	<i>Quantity</i>	<i>Vehicle control Size = 65</i>	<i>Low dose Size = 65</i>	<i>Mid dose Size = 65</i>	<i>High dose Size = 65</i>	
MAMMARY	ADENOMA	Number of animals reported with tumor	2	0	0	1	
		P-value of test of trend or comparison	.2316	.9285	.0838	.4788	
	CARCINOMA	Number of animals reported with tumor	3	1	8	4	
		P-value of test of trend or comparison	.8655	.6761	.8773	.9052	
	FIBROADENOMA	Number of animals reported with tumor	14	12	9	9	
		P-value of test of trend or comparison	.5347	.7928	.3914	.7032	
	FIBROMA	Number of animals reported with tumor	20	16	22	18	
		P-value of test of trend or comparison	1	1	1	1	
	FIBROSARCOMA	Number of animals reported with tumor	1	0	0	0	
		P-value of test of trend or comparison	.4976		.4857		
	OVARY	CYSTADENOMA, PAPILLARY	Number of animals reported with tumor	0	0	1	0
			P-value of test of trend or comparison	1	1	1	1
INTERSTITIAL CELL ADENOMA		Number of animals reported with tumor	1	0	0	0	
		P-value of test of trend or comparison	.8768	1	.8714	1	
THECOMA		Number of animals reported with tumor	2	0	1	0	
		P-value of test of trend or comparison	1	1	1	1	
TUBULOSTROMAL ADENOMA	Number of animals reported with tumor	1	0	0	0		
	P-value of test of trend or comparison	.2476			.4857		
PANCREAS	ISLET CELL ADENOMA	Number of animals reported with tumor	0	0	0	1	
		P-value of test of trend or comparison	.7175	1	1	.9338	
PITUITARY	ADENOMA	Number of animals reported with tumor	3	0	0	1	
		P-value of test of trend or comparison	.9960	.0146	.2316	.9259	
THYMUS	HEMANGIOMA	Number of animals reported with tumor	48	55	52	40	
		P-value of test of trend or comparison	.2560			.4943	
THYROID	"C" CELL ADENOMA	Number of animals reported with tumor	0	0	0	1	
		P-value of test of trend or comparison	.2904	.9650	1	.6829	
		Number of animals reported with tumor	8	3	0	7	

Table A.9

**Table of reported tumors in Rat Study
NDA 206426
Animal carcinogenicity study
Female rats**

<i>Organ or tissue name</i>	<i>Tumor name</i>	<i>Quantity</i>	<i>Vehicle control Size = 65</i>	<i>Low dose Size = 65</i>	<i>Mid dose Size = 65</i>	<i>High dose Size = 65</i>
	"C" CELL CARCINOMA	P-value of test of trend or comparison	.7379	.4757		
		Number of animals reported with tumor	0	1	0	0
	FOLLICULAR CELL ADENOMA	P-value of test of trend or comparison	.2476			.4857
		Number of animals reported with tumor	0	0	0	1
UTERUS	ADENOCARCINOMA	P-value of test of trend or comparison	.3043	.4757		.4857
		Number of animals reported with tumor	0	1	0	1
	ENDOMETRIAL ADENOMA	P-value of test of trend or comparison	.2476			.4857
		Number of animals reported with tumor	0	0	0	1
	ENDOMETRIAL STROMAL POLYP	P-value of test of trend or comparison	.8258	.9819	.1850	.9840
		Number of animals reported with tumor	5	1	9	1
VAGINA	GRANULAR CELL TUMOR	P-value of test of trend or comparison	.7175	1	1	.9338
		Number of animals reported with tumor	3	0	0	1
	SCHWANNOMA	P-value of test of trend or comparison	.2900	.4757	.4953	.4857
		Number of animals reported with tumor	0	1	1	1

Table A.10

**Table of reported tumors in Rat Study
NDA 206426
Animal carcinogenicity study
Male rats**

<i>Organ or tissue name</i>	<i>Tumor name</i>	<i>Quantity</i>	<i>Vehicle control Size = 65</i>	<i>Low dose Size = 65</i>	<i>Mid dose Size = 65</i>	<i>High dose Size = 65</i>
ADRENAL, CORTEX	ADENOMA	P-value of test of trend or comparison	.3127	.5155	.2784	.5000
		Number of animals reported with tumor	0	1	2	1
ADRENAL, MEDULLA	PHEOCHROMOCYTOMA	P-value of test of trend or comparison	.0748	.7003	.0987	.1668
		Number of animals reported with tumor	3	3	9	7
AUDITORY SEB GL	CARCINOMA	P-value of test of trend or comparison	.2386			.5000
		Number of animals reported with tumor	0	0	0	1
BRAIN	ASTROCYTOMA	P-value of test of trend or comparison	.3611	.8863	1	.6835
		Number of animals reported with tumor	2	1	0	2
	GRANULAR CELL TUMOR	P-value of test of trend or comparison	.9440	.7678	1	1
		Number of animals reported with tumor	1	1	0	0
CAVITY, ABDOM	HEMANGIOMA	P-value of test of trend or comparison	.5076		.5300	
		Number of animals reported with tumor	0	0	1	0
	SCHWANNOMA, MALIGNANT	P-value of test of trend or comparison	.6321	.5204	.2784	
		Number of animals reported with tumor	0	1	2	0
CORD, THORACOLUMB	ASTROCYTOMA	P-value of test of trend or comparison	1	1	1	1
		Number of animals reported with tumor	1	0	0	0
EPIDIDYMIS	MALIGNANT MESOTHELIOMA	P-value of test of trend or comparison	.5076		.5300	
		Number of animals reported with tumor	0	0	1	0
HEART	ENDOCARDIAL SCHWANNOMA	P-value of test of trend or comparison	.8966	.3242	.5375	1
		Number of animals reported with tumor	1	3	2	0
HEMATO NEOPLASIA	LEUKEMIA, LARGE GRANULAR	P-value of test of trend or comparison	1	1	1	1
		Number of animals reported with tumor	2	0	0	0
	LYMPHOMA	P-value of test of trend or comparison	.8121	1	.7278	1
		Number of animals reported with tumor	2	0	2	0
	MALIGNANT FIBROUS HISTIOC	P-value of test of trend or comparison	.5101		.5347	
		Number of animals reported with tumor	0	0	1	0
	SARCOMA, HISTIOCYTIC	P-value of test of trend or comparison	.0521	.5155	.2784	.1250

Table A.10

Table of reported tumors in Rat Study
NDA 206426
Animal carcinogenicity study
Male rats

<i>Organ or tissue name</i>	<i>Tumor name</i>	<i>Quantity</i>	<i>Vehicle control Size = 65</i>	<i>Low dose Size = 65</i>	<i>Mid dose Size = 65</i>	<i>High dose Size = 65</i>
		Number of animals reported with tumor	0	1	2	3
JEJUNUM	LEIOMYOMA	P-value of test of trend or comparison	.2278			.4767
		Number of animals reported with tumor	0	0	0	1
KIDNEY	LIPOMA	P-value of test of trend or comparison	.5076		.5300	
		Number of animals reported with tumor	0	0	1	0
LN, MESENTERIC	HEMANGIOSARCOMA	P-value of test of trend or comparison	.7565	.5104		
		Number of animals reported with tumor	0	1	0	0
LUNG	ADENOMA, BRONCHIOLAR-ALVE	P-value of test of trend or comparison	.7602	.5104		
		Number of animals reported with tumor	0	1	0	0
MAMMARY	FIBROADENOMA	P-value of test of trend or comparison	1	1	1	1
		Number of animals reported with tumor	1	0	0	0
MARROW, FEMUR	HEMANGIOMA	P-value of test of trend or comparison	.5076		.5300	
		Number of animals reported with tumor	0	0	1	0
MEDIASTINUM	SARCOMA	P-value of test of trend or comparison	.5076		.5300	
		Number of animals reported with tumor	0	0	1	0
PANCREAS	ADENOMA, ISLET CELL	P-value of test of trend or comparison	.6343	.5155	.2784	
		Number of animals reported with tumor	0	1	2	0
PITUITARY	ADENOMA	P-value of test of trend or comparison	.3258	.3130	.2489	.2989
		Number of animals reported with tumor	18	21	24	22
	CARCINOMA	P-value of test of trend or comparison	1	1	1	1
		Number of animals reported with tumor	1	0	0	0
RECTUM	LEIOMYOSARCOMA	P-value of test of trend or comparison	1	1	1	1
		Number of animals reported with tumor	1	0	0	0
SALIV GL, MANDIB	SCHWANNOMA, MALIGNANT	P-value of test of trend or comparison	.1850		.5300	.5000
		Number of animals reported with tumor	0	0	1	1
SKIN	OSTEOSARCOMA	P-value of test of trend or comparison	1	1	1	1
		Number of animals reported with tumor	1	0	0	0

Table A.10

**Table of reported tumors in Rat Study
NDA 206426
Animal carcinogenicity study
Male rats**

<i>Organ or tissue name</i>	<i>Tumor name</i>	<i>Quantity</i>	<i>Vehicle control Size = 65</i>	<i>Low dose Size = 65</i>	<i>Mid dose Size = 65</i>	<i>High dose Size = 65</i>
SKIN, OTHER	SARCOMA	P-value of test of trend or comparison	.8369	.7678	.7816	1
		Number of animals reported with tumor	1	1	1	0
	BASAL CELL CARCINOMA	P-value of test of trend or comparison	.2386			.5000
		Number of animals reported with tumor	0	0	0	1
	FIBROMA	P-value of test of trend or comparison	.5202	.7678	1	.7527
		Number of animals reported with tumor	1	1	0	1
	KERATOACANTHOMA	P-value of test of trend or comparison	.0636	1	.5454	.3084
		Number of animals reported with tumor	1	0	2	3
	SARCOMA	P-value of test of trend or comparison	.5076		.5300	
		Number of animals reported with tumor	0	0	1	0
SCHWANNOMA	P-value of test of trend or comparison	.1850		.5300	.5000	
	Number of animals reported with tumor	0	0	1	1	
STOMACH, NONGL	SEBACEOUS GLAND ADENOMA	P-value of test of trend or comparison	.9422	.7627	1	1
		Number of animals reported with tumor	1	1	0	0
	SQUAMOUS CELL PAPILLOMA	P-value of test of trend or comparison	.8503	.5155	.5375	1
		Number of animals reported with tumor	1	2	2	0
SQUAMOUS CELL PAPILLOMA	P-value of test of trend or comparison	.2386			.5000	
	Number of animals reported with tumor	0	0	0	1	
TESTIS	INTERSTITIAL CELL TUMOR	P-value of test of trend or comparison	.4023	.9748	.9784	.7736
		Number of animals reported with tumor	4	1	1	3
MESOTHELIOMA	P-value of test of trend or comparison	.5076		.5300		
		Number of animals reported with tumor	0	0	1	0
THYMUS	THYMOMA	P-value of test of trend or comparison	.5285	.7471	1	.7471
		Number of animals reported with tumor	1	1	0	1
THYROID	"C" CELL ADENOMA	P-value of test of trend or comparison	.5056	.6992	.0641	.6835
		Number of animals reported with tumor	2	2	8	2
	"C" CELL CARCINOMA	P-value of test of trend or comparison	.9673	.9438	.9523	1

Table A.10

**Table of reported tumors in Rat Study
NDA 206426
Animal carcinogenicity study
Male rats**

<i>Organ or tissue name</i>	<i>Tumor name</i>	<i>Quantity</i>	<i>Vehicle control Size = 65</i>	<i>Low dose Size = 65</i>	<i>Mid dose Size = 65</i>	<i>High dose Size = 65</i>
		Number of animals reported with tumor	3	1	1	0
	FOLLICULAR CELL ADENOMA	P-value of test of trend or comparison	.9410	.7577	1	1
		Number of animals reported with tumor	1	1	0	0
	FOLLICULAR CELL CARCINOMA	P-value of test of trend or comparison	.2398			.5000
		Number of animals reported with tumor	0	0	0	1

Table A.11

**Table of reported tumors in Rat Study
NDA 206426
Animal carcinogenicity study
Female rats
Composite endpoints**

<i>Composite endpoint</i>	<i>Quantity</i>	<i>Vehicle control Size = 65</i>	<i>Low dose Size = 65</i>	<i>Mid dose Size = 65</i>	<i>High dose Size = 65</i>
Adenomas and carcinomas of the pituitary	P-value of test of trend or comparison	.9960	.0146	.2316	.9259
	Number of animals reported with tumor	48	55	52	40
All astrocytomas	P-value of test of trend or comparison	.0628			.2383
	Number of animals reported with tumor	0	0	0	2
All leiomyosarcomas	P-value of test of trend or comparison	.4348	1	1	.7379
	Number of animals reported with tumor	1	0	0	1
All papillomas	P-value of test of trend or comparison	.1860		.4906	.4857
	Number of animals reported with tumor	0	0	1	1
All schwannomas including neurofibrosarcomas	P-value of test of trend or comparison	.4471	.3045	.9387	.4678
	Number of animals reported with tumor	3	5	1	4
All squamous cell tumors (including keratoacanthoma)	P-value of test of trend or comparison	.7379	.4757		
	Number of animals reported with tumor	0	1	0	0
Carcinomas and adenomas of the adrenal cortex	P-value of test of trend or comparison	.6428	.8597	.4911	.8677
	Number of animals reported with tumor	2	1	3	1
Endometrial tumors of the uterus	P-value of test of trend or comparison	.6839	.9819	.1850	.9343
	Number of animals reported with tumor	5	1	9	2
Follicular cell adenomas and carcinomas of the thyroid	P-value of test of trend or comparison	.2476			.4857
	Number of animals reported with tumor	0	0	0	1
Hemangiomas and hemangiosarcomas	P-value of test of trend or comparison	.2476			.4857
	Number of animals reported with tumor	0	0	0	1
Histiocytic sarcomas	P-value of test of trend or comparison	.1187	.4757	.4953	.2383
	Number of animals reported with tumor	0	1	1	2
Leiomyosarcomas and Leiomyomas	P-value of test of trend or comparison	.5915	1	1	.8677
	Number of animals reported with tumor	2	0	0	1

Table A.12

**Table of reported tumors in Rat Study
NDA 206426
Animal carcinogenicity study
Male rats
Composite endpoints**

<i>Composite endpoint</i>	<i>Quantity</i>	<i>Vehicle control Size = 65</i>	<i>Low dose Size = 65</i>	<i>Mid dose Size = 65</i>	<i>High dose Size = 65</i>
Adenomas and carcinomas of the pituitary	P-value of test of trend or comparison	.3721	.3882	.3192	.3737
	Number of animals reported with tumor	19	21	24	22
All astrocytomas	P-value of test of trend or comparison	.5006	.9437	1	.8059
	Number of animals reported with tumor	3	1	0	2
All leiomyosarcomas	P-value of test of trend or comparison	1	1	1	1
	Number of animals reported with tumor	1	0	0	0
All papillomas	P-value of test of trend or comparison	.5894	.5155	.5375	.7474
	Number of animals reported with tumor	1	2	2	1
All schwannomas including neurofibrosarcomas	P-value of test of trend or comparison	.5403	.2071	.0730	.4920
	Number of animals reported with tumor	1	4	6	2
All squamous cell tumors (including keratoacanthoma)	P-value of test of trend or comparison	.1645	.7067	.3874	.3286
	Number of animals reported with tumor	2	2	4	4
Carcinomas and adenomas of the adrenal cortex	P-value of test of trend or comparison	.3127	.5155	.2784	.5000
	Number of animals reported with tumor	0	1	2	1
Follicular cell adenomas and carcinomas of the thyroid	P-value of test of trend or comparison	.5184	.7577	1	.7474
	Number of animals reported with tumor	1	1	0	1
Hemangiomas and hemangiosarcomas	P-value of test of trend or comparison	.6298	.5155	.5300	
	Number of animals reported with tumor	0	1	1	0
Histiocytic sarcomas	P-value of test of trend or comparison	.0521	.5155	.2784	.1250
	Number of animals reported with tumor	0	1	2	3
Leiomyosarcomas and Leiomyomas	P-value of test of trend or comparison	.4212	1	1	.7527
	Number of animals reported with tumor	1	0	0	1
Sarcomas of skin and skin (other)	P-value of test of trend or comparison	.7902	.7678	.5454	1
	Number of animals reported with tumor	1	1	2	0

Table A.13

**Table of tumors reported significant ($\alpha < 0.05$) in at least one arm - Rat Study
NDA 206426
Animal carcinogenicity study
Female rats**

<i>Organ or tissue name</i>	<i>Tumor name</i>	<i>Quantity</i>	<i>Vehicle control</i>	<i>Low dose</i>	<i>Mid dose</i>	<i>High dose</i>
PITUITARY	ADENOMA	P-value of test of trend or comparison	.9960	.0146	.2316	.9259
		Number of animals reported with tumor	48	55	52	40
		Poly-3 adjusted incidence rate	79%	94%	85%	69%
		95% CI for poly-3 adjusted incidence rate (%)	(66.3,89.2)	(83.5,98.9)	(73.8,94.1)	(55.5,81.6)
		Poly-3 adjusted number of animals at risk	61.0	58.8	60.9	57.7

Table A.14

Table of tumors reported significant ($\alpha < 0.05$) in at least one arm - Rat Study
NDA 206426
Animal carcinogenicity study
Female rats
Composite endpoints

<i>Composite endpoint</i>	<i>Quantity</i>	<i>Vehicle control</i>	<i>Low dose</i>	<i>Mid dose</i>	<i>High dose</i>
Adenomas and carcinomas of the pituitary	P-value of test of trend or comparison	.9960	.0146	.2316	.9259
	Number of animals reported with tumor	48	55	52	40
	Poly-3 adjusted incidence rate	79%	94%	85%	69%
	95% CI for poly-3 adjusted incidence rate (%)	(66.3,89.2)	(83.5,98.9)	(73.8,94.1)	(55.5,81.6)
	Poly-3 adjusted number of animals at risk	61.0	58.8	60.9	57.7

Tumor counts for organs reported widely analyzed
NDA 206426
Animal carcinogenicity study
Rats

<i>Species and Sex</i>	<i>Organ or tissue name</i>	<i>Tumor name</i>	<i>Quantity</i>	<i>Vehicle control</i>	<i>Low dose</i>	<i>Mid dose</i>	<i>High dose</i>
Rats - Female	CAVITY, THORACIC	MESOTHELIOMA	Number of tumors found	0	0	1	0
			Number of animals examined	1	0	1	3
	HEAD, CORONAL	PAPILLOMA	Number of tumors found	0	0	1	0
			Number of animals examined	0	1	1	3
	SKIN, OTHER	MELANOMA, AMELANOTIC	Number of tumors found	0	0	0	1
			Number of animals examined	11	18	16	24
		PAPILLOMA	Number of tumors found	0	0	0	1
			Number of animals examined	11	18	16	24
	SUBCUTANEOUS TIS	LEIOMYOSARCOMA	Number of tumors found	1	0	0	0
			Number of animals examined	1	2	0	1
		NEUROFIBROSARCOMA	Number of tumors found	0	1	0	0
			Number of animals examined	1	2	0	1
	ZYMBAL'S GLAND	SQUAMOUS CELL CARCINOMA	Number of tumors found	0	1	0	0
			Number of animals examined	0	1	0	0

Table A.15

A.3 Unexamined and autolytic organs

Table A.16

**Organs reported as autolytic
NDA 206426
Animal carcinogenicity study
Female Rats**

<i>Organ or tissue name</i>	<i>Vehicle control(count)</i>	<i>Vehicle control(%)</i>	<i>Low dose(count)</i>	<i>Low dose(%)</i>	<i>Mid dose(count)</i>	<i>Mid dose(%)</i>	<i>High dose(count)</i>	<i>High dose(%)</i>	<i>Total(count)</i>	<i>Total(%)</i>
CAVITY, ORAL	1	1.5%	1	0.4%
DUODENUM	.	.	1	1.5%	1	1.5%	.	.	2	0.8%
ILEUM	1	1.5%	1	1.5%	1	1.5%	1	1.5%	4	1.5%
JEJUNUM	1	1.5%	.	.	2	3.1%	1	1.5%	4	1.5%
THYMUS	11	17%	14	22%	7	11%	10	15%	42	16%

Table A.17

**Organs reported as autolytic
NDA 206426
Animal carcinogenicity study
Male Rats**

<i>Organ or tissue name</i>	<i>Vehicle control(count)</i>	<i>Vehicle control(%)</i>	<i>Low dose(count)</i>	<i>Low dose(%)</i>	<i>Mid dose(count)</i>	<i>Mid dose(%)</i>	<i>High dose(count)</i>	<i>High dose(%)</i>	<i>Total(count)</i>	<i>Total(%)</i>
ADRENAL, MEDULLA	1	1.5%	1	0.4%
CECUM	9	14%	9	14%	2	3.1%	10	15%	30	12%
COLON	1	1.5%	2	3.1%	.	.	2	3.1%	5	1.9%
DUODENUM	3	4.6%	3	4.6%	3	4.6%	3	4.6%	12	4.6%
EYE	1	1.5%	1	0.4%
ILEUM	5	7.7%	8	12%	8	12%	15	23%	36	14%
JEJUNUM	4	6.2%	8	12%	6	9.2%	10	15%	28	11%
LN, MESENTERIC	1	1.5%	.	.	1	0.4%
MAMMARY	.	.	1	1.5%	1	1.5%	.	.	2	0.8%
PARATHYROID	1	1.5%	3	4.6%	2	3.1%	.	.	6	2.3%
PITUITARY	.	.	3	4.6%	2	3.1%	1	1.5%	6	2.3%
RECTUM	2	3.1%	2	3.1%	4	1.5%
SEMINAL VESICLE	1	1.5%	.	.	1	1.5%	6	9.2%	8	3.1%
THYMUS	.	.	2	3.1%	2	3.1%	.	.	4	1.5%
THYROID	1	1.5%	1	0.4%

Table A.18

**Organs reported as unexamined
NDA 206426
Animal carcinogenicity study
Female Rats**

<i>Organ or tissue name</i>	<i>Vehicle control(count)</i>	<i>Vehicle control(%)</i>	<i>Low dose(count)</i>	<i>Low dose(%)</i>	<i>Mid dose(count)</i>	<i>Mid dose(%)</i>	<i>High dose(count)</i>	<i>High dose(%)</i>	<i>Total(count)</i>	<i>Total(%)</i>
BONE, FEMUR	.	.	1	1.5%	1	0.4%
BONE, OTHER	65	100%	64	98%	65	100%	62	95%	256	98%
BONE, STERNUM	1	1.5%	1	0.4%
BRAIN	1	1.5%	.	.	1	0.4%
CAVITY, ABDOM	64	98%	63	97%	64	98%	64	98%	255	98%
CAVITY, ORAL	65	100%	65	100%	65	100%	64	98%	259	100%
CAVITY, THORACIC	64	98%	65	100%	64	98%	62	95%	255	98%
COLON	1	1.5%	.	.	1	0.4%
CORD,THORACOLUMB	1	1.5%	1	0.4%
HEAD, CORONAL	65	100%	64	98%	64	98%	62	95%	255	98%
JEJUNUM	1	1.5%	1	0.4%
LN, MANDIBULAR	2	3.1%	.	.	2	3.1%	2	3.1%	6	2.3%
LN, OTHER	65	100%	65	100%	64	98%	64	98%	258	99%
MAMMARY	2	3.1%	.	.	2	0.8%
MARROW, FEMUR	.	.	1	1.5%	1	0.4%
MARROW, STERNUM	1	1.5%	1	0.4%
MUSCLE, SKELETAL	.	.	1	1.5%	1	0.4%
NASAL TURBINATE	64	98%	64	98%	64	98%	64	98%	256	98%
NERVE, OPTIC	.	.	1	1.5%	2	3.1%	1	1.5%	4	1.5%
NERVE, SCIATIC	.	.	1	1.5%	1	0.4%
PARATHYROID	4	6.2%	2	3.1%	5	7.7%	4	6.2%	15	5.8%
PINNA	65	100%	64	98%	65	100%	65	100%	259	100%
PITUITARY	1	1.5%	1	1.5%	.	.	1	1.5%	3	1.2%
SKIN, OTHER	54	83%	47	72%	49	75%	41	63%	191	73%
SUBCUTANEOUS TIS	64	98%	63	97%	65	100%	64	98%	256	98%
THYMUS	.	.	2	3.1%	1	1.5%	.	.	3	1.2%
ZYMBAL'S GLAND	65	100%	64	98%	65	100%	65	100%	259	100%

Table A.19

Organs reported as unexamined
NDA 206426
Animal carcinogenicity study
Male Rats

<i>Organ or tissue name</i>	<i>Vehicle control(count)</i>	<i>Vehicle control(%)</i>	<i>Low dose(count)</i>	<i>Low dose(%)</i>	<i>Mid dose(count)</i>	<i>Mid dose(%)</i>	<i>High dose(count)</i>	<i>High dose(%)</i>	<i>Total(count)</i>	<i>Total(%)</i>
ADRENAL, MEDULLA	1	1.5%	1	0.4%
CAVITY, ABDOM	1	1.5%	1	1.5%	2	0.8%
CAVITY, THORACIC	.	.	1	1.5%	.	.	1	1.5%	2	0.8%
COAGULATING GL	1	1.5%	1	0.4%
DUODENUM	1	1.5%	1	0.4%
ESOPHAGUS	.	.	1	1.5%	1	0.4%
HEART	.	.	1	1.5%	1	0.4%
LN, MESENTERIC	1	1.5%	1	1.5%	1	1.5%	1	1.5%	4	1.5%
LUNG	.	.	1	1.5%	1	0.4%
MAMMARY	6	9.2%	11	17%	7	11%	4	6.2%	28	11%
NERVE, OPTIC	2	3.1%	3	4.6%	2	3.1%	1	1.5%	8	3.1%
PARATHYROID	4	6.2%	1	1.5%	5	7.7%	2	3.1%	12	4.6%
THYMUS	6	9.2%	7	11%	12	18%	7	11%	32	12%
THYROID	.	.	1	1.5%	1	0.4%
TRACHEA	.	.	1	1.5%	1	0.4%

A.4 Weight changes

Table A.20: Weight changes by group (rats)

Sex	Vehicle control	Rapivab					
	Δ_{CP}	Δ_L	$\frac{\Delta_L}{\Delta_{CP}} - 1$	Δ_M	$\frac{\Delta_M}{\Delta_{CP}} - 1$	Δ_H	$\frac{\Delta_H}{\Delta_{CP}} - 1$
Female	160	157	-2%	163	2%	160	0%
Male	311	320	3%	296	-5%	293	-6%

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

MATTHEW T JACKSON
09/20/2014

KARL K LIN
09/23/2014

The following two notes are added to this review. The primary reviewer is no longer available to make the changes.

A. Years of publications --- 2nd and 3rd lines on page 2: Lin and Rahman (1998), and Rahman and Lin (2008).

B. The following bibliography section should be added to the end of the report.

1. A J Bailer & C J Portier, *Biometrics*, 44(2):417-431, 1998
2. G S Bieler & R L Williams, *Biometrics*, 49(3):793-801, 1993
3. KC Chu, C Cueto and J M Ward, *Journal of Toxicology and Environmental Health* (8(1-2) 251-280, 1981.
4. J K Haseman, *Fundamental and Applied Toxicology*, 3(4):334-339, 1983
5. K K Lin & M A Rahman, *Journal of Biopharmaceutical Statistics*, 8(1):1-22.
6. M A Rahman & K K Lin, *Journal of Biopharmaceutical Statistics*, 18(5):949-958, 2008.

STATISTICAL REVIEW AND EVALUATION

NDA#: 206426 SDN 001

DRUG NAME: Peramivir

INDICATION: Treatment of Acute Uncomplicated
Influenza

TYPE OF REVIEW: Clinical

APPLICANT: BioCryst Pharmaceuticals

DATES: Dec 23, 2013

REVIEW PRIORITY: Standard

BIOMETRICS DIVISION: Division of Biometrics IV

STATISTICAL REVIEWER: Thomas Hammerstrom, (HFD-725)

TEAM LEADER: Greg Soon, PhD, (HFD-725)

MEDICAL DIVISION: DAVP

CLINICAL TEAM: Peter Miele, MD (HFD-530), Linda Lewis,
M.D. (HFD-530)

PROJECT MANAGER: Elizabeth Thompson, (HFD-530)

STATISTICAL REVIEW AND EVALUATION

NDA#:

206426

1. Executive Summary
2. Introduction
 - 2.1 Overview
 - 2.2 Data Sources
 - 2.2.1 Objectives in Trials
 - 2.2.2 Summary of Study Design
 - 2.2.3 Patient Accounting and Baseline Characteristics
 - 2.2.4 Summary of Methods of Assessment
 - 2.2.4.1 Schedule of Measurements and Assessment of Treatment Effects
 - 2.2.5 Summary of Statistical Analysis
 - 2.2.6 Summary of Applicant's Results
 - 2.2.6.1 Trial 722
 - 2.2.6.2 Trials 211 and 311
 - 2.2.6.3 Trial 212
 - 2.2.6.4 Trial 815
 - 2.2.6.5 Uncontrolled Trials 816 and 918
 - 2.2.7 Summary of Applicant's Conclusions
 3. Statistical Evaluation
 - 3.1 Discussion of Missing and Incorrect Times in Diary Entry Records
 - 3.2 FDA Analysis of Time to Symptom Alleviation
 - 3.2.1 Kaplan-Meier Plots of Time to Symptom Alleviation
 - 3.2.2 Cox Regressions of Time to Symptom Alleviation
 - 3.2.3 Estimation of Hazard Ratio of Peramivir to Placebo from Trial 815 and Tamiflu NDA
 - 3.2.4 Comparison of Hazard Ratio of Peramivir to Placebo for Types A and B
 - 3.2.5 Confidence Intervals for Medians from Kaplan-Meier Curves
 - 3.2.6 Pooling Evidence from Trials Excluding Tamiflu-Controlled 815
 - 3.2.7 Dose Response Modelling
 - 3.3 Time to Resolution of Fever
 4. Results in Special Populations
 - 4.1 Gender, Race, and Age
 - 4.2 Other Covariate
 5. Summary and Conclusions

1. Executive Summary

The applicant has conducted seven trials to test the efficacy of peramivir (PVR) in the treatment of acute uncomplicated influenza. Two of these trials, 0722T0621 and 0815T0631 (henceforth, trials 722 and 815), used the 600 mg IV dose proposed in the application. These trials were conducted by the Japanese sponsor of the drug, Shionogi. These trials were large, randomized, double blind, controlled trials. Trial 722 was placebo controlled and 815 was active controlled with Tamiflu. Trial 722 will be considered the pivotal trial. A flu season with an unforeseen and widespread Tamiflu resistant influenza strain means that the majority of the control subjects in trial 815 were not treated with a provably effective active drug. Thus, establishment of non-inferiority in trial 815 cannot be interpreted as non-inferiority to an active drug. This prevents trial 815 from being a pivotal trial.

The primary objective of the two large trials, 722 (pivotal) and 815 (intended pivotal with a problematic active control), was to establish the efficacy of a single dose of peramivir at 600mg IV for the treatment of acute uncomplicated influenza. In these and all other trials, the method of primary efficacy analysis is ITTI, intent-to-treat on the infected subjects.

Three other trials were conducted by the US applicant, BioCryst. These trials, 211, 311, and 212, were randomized, double blind, placebo controlled trials using PVR at 300 or 600mg IM. The applicant has provided PK data supporting the assertion that IM and IV dosings are bioequivalent. The applicant pooled two of these trials, 211 and 311, because the sample sizes were somewhat small. The applicant analyzed them as one trial across two influenza seasons. This was done despite some differences in the design of the trials.

The objective of trials 211, 311, and 212 was originally to determine whether a single dose of peramivir IM at one of 150 mg, 300 mg, or 600 mg was effective in the treatment of acute uncomplicated influenza. In the current submission, their purpose is to support the efficacy of 600 mg IV peramivir for the same indication. PK data supporting the bioequivalence of IM and IV doses is required to complete this argument. The bioequivalence argument is reviewed in the pharmacokinetics review and is taken for granted here.

The one large placebo controlled trial, 722, showed a sufficient superiority of peramivir to placebo to provide statistical evidence equivalent to two studies statistically significant at the conventional one sided .025 level.

None of the other studies achieved formal statistical significance at one sided level .025. Nonetheless, they all provide evidence supporting the conclusion from trial 722. The median times to healing were clinically meaningfully shorter for peramivir than for placebo in all trials: 30 hours shorter for both 300mg and 600mg in trial 722, 8 hours shorter for 150mg and 14 hours shorter for 300mg in the pooled results of trials 211 and 311, 21 hours shorter for 600mg in trial 212 (provided one looked only at the Tamiflu susceptible subjects in trial 212: type A H3N2 and H1N1 Wild).

The dose response pattern seen in the difference in the medians was part of a statistically significant pattern in the log hazard ratios that could be found by comparing the 150, 300, and 600mg responses in all four placebo controlled trials.

The three BioCryst trials also suggest that peramivir does not work in type B influenza. In all three trials, the log hazard ratios of peramivir to placebo were positive for type A and negative for type B, although all the confidence intervals straddled zero. There were almost no type B cases in trial 722 so all the evidence concerning efficacy in type B comes from these three trials.

With respect to baseline covariates, five covariates showed a suggestion of an interaction with treatment. Peramivir's performance relative to placebo declines with increasing age and does better in Asia than in the rest of the world. It also declines with later start of treatment and higher baseline symptom score. Asians tend to seek treatment earlier and with lower baseline symptom scores than Americans and Africans so this effect of country is a surrogate for those effects. Blacks also did worse on peramivir relative to placebo; if this is anything other than random, the reason for it is unclear.

Finally, there are reasons to doubt whether peramivir will work against Tamiflu resistant strains of type A influenza. In trial 212, positive results for peramivir were obtained in the sub-group with type A, H1N1 wild or H3N2. Results were negative

for the sub-group with type B or type A, H1N1 with H275Y substitution. In trial 815, there was no testing for the H275Y substitution but all subjects with type A H1N1 had $IC_{50} > 15$, all subjects with type A H3N2 had $IC_{50} < 5$. It will be shown (in section 3.2.5, table A) that the median healing times for all three arms were about 11 hours shorter among H3N2 subjects than among H1N1 or B subjects. Furthermore, the correlation between IC_{50} for Tamiflu and IC_{50} for peramivir was high.

In short summary, peramivir at 600mg is convincingly effective against Tamiflu susceptible strains of influenza A with about a 30 hour reduction in symptom duration. Peramivir may be effective down to doses as low as 150mg against those influenza strains but the reduction in symptom duration is only about 8 hours. Finally, it is unlikely that peramivir is very effective against influenza B or against strains of influenza A that are Tamiflu resistant.

2. Introduction

2.1 Overview

The applicant has conducted seven trials to test the efficacy of peramivir (PVR) in the treatment of acute uncomplicated influenza. Two of these trials, 0722 and 0815 (conducted by Shionogi) used the 600 mg IV dose proposed in the application. These trials were large, randomized, double blind, controlled trials. Trial 722 was placebo controlled and 815 was active controlled. Trial 722 may be considered the pivotal trial. A flu season with an unforeseen and widespread Tamiflu resistant influenza strain prevents trial 815 from being a pivotal trial.

Three other trials, 211, 311, and 212, were conducted by BioCryst and were randomized, double blind, placebo controlled trials using PVR at 300 or 600mg IM. The applicant has provided PK data supporting the assertion that IM and IV dosings are bioequivalent. The applicant pooled two of these trials, 211 and 311, because the sample sizes were somewhat small. The applicant analyzed them as one trial across two influenza seasons. This was done despite some differences in the design of the trials.

The remaining two trials, 0816T0632 and 0918T0633 (henceforth, trials 816 and 918) provide supplementary information about high risk or pediatric subjects. Trial 816 compared 600mg PRV IM to 300mg PRV IM; trial 918 was an uncontrolled pediatric trial.

2.2 Data Sources

2.2.1 Objectives in Trials

The primary objective of the two pivotal trials, 722 and 815, was to establish the efficacy of a single dose of peramivir at 600mg IV for the treatment of acute uncomplicated influenza. The primary analysis is conducted by the ITTI (intent to treat, infected) method.

The objective of trials 211, 311, and 212 was originally to determine whether a single dose of peramivir IM at one of 150 mg, 300 mg, or 600 mg was effective in the treatment of acute uncomplicated influenza. In the current submission, their purpose is to support the efficacy of 600 mg IV peramivir for the same indication. PK data supporting the bioequivalence of IM and IV doses is required to complete this argument. The bioequivalence argument is reviewed in the clinical pharmacology review and is taken for granted here.

The objective of trial 722 and 918, is to provide supplementary (b) (4) of peramivir against acute uncomplicated influenza in high risk and pediatric populations.

2.2.2 Summary of Study Design

Trial 722 was a randomized, double blind, placebo controlled trial, conducted in the 2007-08 flu season in Japan. 300 subjects at 75 centers were randomized 1:1:1 to a single dose of either PVR at 300 or 600 mg IV or placebo. Randomization was stratified by current smoking status and composite symptom score at baseline (sum over the 7 symptoms of scores each ranging from 0 to 3).

Trial 815 was a randomized, double blind, double dummy, three arm, active controlled trial. 1099 subjects at 146 centers were randomized 1:1:1 to either a single dose of PRV at 300 or 600 mg IV or twice oral daily doses of 15mg Tamiflu for 5 days

during the flu season of 2008-2009. Randomization was by dynamic allocation with the objective of balancing the trial on 4 factors: current smoking status, composite symptom score at baseline (\leq or $>$ 14), country, and influenza type.

Trial 212 was a randomized, double blind, placebo controlled trial. 405 subjects at 69 centers were randomized 1:1 to a single dose of either 600 mg IM or placebo during the flu season of 2008-2009. Randomization was stratified by current smoking behavior and RAT result for influenza A or B.

Trials 211 and 311 were randomized, double blind, placebo controlled trials. Trial 211 was a multinational trial conducted during the year 2007 (Jan-Sept) in which 344 subjects at 151 centers were randomized 1:1:1 to a single dose of either PVR at 150 or 300 mg IM or placebo. Randomization was stratified by current smoking behavior.

Trial 311 was a US trial conducted in the year 2008 (Jan-Apr) in which 82 subjects at 37 centers were randomized 2:1 to a single dose of either a single dose of either PRV at 300 mg IM or placebo during the year 2008. Randomization in trial 311 was stratified by current smoking behavior and RAT result for influenza A or B. Subjects in both trials were randomized after having an RAT test for influenza based on an anterior nasal swab.

Trial 816 was a randomized, double blind, dose ranging study in which 42 subjects at 37 centers were randomized to a single dose of either 300 or 600 mg of PRV IV, once daily for 1 to 5 days. On Day 2 and later, study drug was continued at the discretion of the investigator based on the body temperature of $\geq 37.5^{\circ}\text{C}$ or clinical manifestations. Within the safety analysis set, 88.1% of the subjects received either 1 or 2 doses of study drug. Study drug was administered by IV drip infusion over 15 to 60 minutes. Randomization was by dynamic allocation.

Trial 918 was a non-randomized, open label pediatric study in 177 subjects aged 28 days to 16 years were assigned to PRV IV at 10mg/kg (with a max of 600 mg) once daily for 1 to 5 days.

2.2.3 Patient Accounting and Baseline Characteristics

Trial 722 was conducted in the 2007-08 flu season in Japan. There were 297 subjects in the trial; 49% were female; they were aged between 20 and 62 with a mean of 35 years. 34% were smokers. 12% sought treatment with 12 hours of symptom onset; another 42% within 12 to 24 hours. 77% had baseline symptom score <14, adding up 0 to 3 for none to severe over seven symptoms.

Their influenza typing is given in table 2.2.3 A

TABLE 2.2.3 A
INFLUENZA SUBTYPE, TRIAL 722

A/H1N1	216	72.73
A/H3N2	70	23.57
A/Indeterminate	8	2.69
B	3	1.01

Trial 815 was conducted in 2008-9 flu season in East Asia. Of the 1099 subjects, 742 (68%) of subjects were Japanese; 105 (10%) were Korean; 246 (23%) were Taiwanese. The subjects were 48% female, aged between 20 and 80 with a mean of 35 years. Only 2% were over 65. 31% were smokers. 8% sought treatment with 12 hours of symptom onset; another 35% within 12 to 24 hours. 63% had a baseline symptom score <14.

Their influenza typing is given in table 2.2.3 B

TABLE 2.2.3 B
INFLUENZA SUBTYPE, TRIAL 815

A/H1N1	599	55%
A/H3N2	329	30%
A/Indeterminate	54	4.9%
B	70	6.4%
Indeterminate	41	3.8%

Trial 211 was conducted over the 2006-7 (40% of the 319 subjects) and 2007-8 flu seasons in the northern hemisphere (1% of subjects) and the 2007 flu season in the southern hemisphere (59% of subjects). Subjects were enrolled from different countries according to table 2.2.3 C

TABLE 2.2.3 C
ENROLLMENT BY COUNTRY, TRIAL 211

USA	99	31%
Canada	25	7.8%
Britain	2	0.63%
Hong_Kong	5	1.6%
Australia	84	26%
New_Zealand	35	11%
South_Africa	69	22%

Subjects were 70% white, 10% Asian or Pacific Islander, and 15% Black. They were 53% female, aged between 18 and 92 years with a mean of 36 years; only 3% were over age 65. 22% were smokers. Only 3% sought treatment with 12 hours of symptom onset; another 27% sought treatment within 12 to 24 hours. (Compare these figures to the East Asian trials: 12% and 42% or 8% and 35%.) 40% had baseline symptom score <14 (compared to 77% and 63% in the East Asian trials.)

Their influenza typing is given in table 2.2.3 D

TABLE 2.2.3 D
INFLUENZA SUBTYPE, TRIAL 211

A/H1N1	83	26%
A/H3N2	160	50%
A/Indeterminate	7	2.2%
B	65	20%
A+B	3	0.9%
Indeterminate	1	0.3%

Trial 311 enrolled 82 patients in the US in the 2007-8 flu season. Subjects were 68% white, 7% Black, 6% Asian, and 18% had unspecified race. They were 51% female, age between 18 and 81 with a mean of 33 years. Only 2 were over age 65. 22% were smokers. Only 6% sought treatment with 12 hours of symptom onset; another 26% sought treatment within 12 to 24 hours. (Compare these figures to the East Asian trials: 12% and 42% or 8% and 35%.) 26% had baseline symptom score <14 (compared to 77% and 63% in the East Asian trials.)

Trial 212 enrolled 174 subjects in the 2008 southern hemisphere flu season and another 160 in the 2008-9 northern hemisphere flu season. Subjects were enrolled from different countries according to table 2.2.3 E

TABLE 2.2.3 E
ENROLLMENT BY COUNTRY, TRIAL 212

USA	160	48%
Australia	20	6%
New_Zealand	26	8%
South_Africa	128	38%

The subjects were 61% white, 28% Black, and 11% Asian. They were 49% female, aged between 18 and 71 with a mean of 34 years. Only 4 (1.2%) were over age 65. 20% were smokers. Only 7% sought treatment with 12 hours of symptom onset; another 44% sought treatment within 12 to 24 hours. Unlike trials 211 and 311, these rates are not too inferior to the rates in the East Asian trials: 12% and 42% or 8% and 35%. 37% had baseline symptom score <14 (compared to 77% and 63% in the East Asian trials.)

Their influenza typing is given in table 2.2.3 F.

TABLE 2.2.3 F
INFLUENZA SUBTYPE, TRIAL 212

A-H1N1, H275Y	230	69%
A-H1N1, Wild Type	15	4.5%
A-H3N2	36	11%
B	52	16%
Indeterminate	1	0.3%

It is worth summarizing the differences between the Shionogi trials in East Asia and the BioCryst trials in the rest of the world. The East Asian trial subjects mostly sought treatment earlier and had less severe symptoms at baseline. East Asians were slightly heavier smokers (about 33% vs about 20% in the BioCryst trials). The sex and age distributions were similar in all trials.

Two of the trials (Shionogi 815 and BioCryst 212) had a substantial amount of Tamiflu resistance. 69% of subjects in 212 had the resistance mutation H275Y. This mutation results in reduced sensitivity to Tamiflu and peramivir compared to wild type. For external documentation of this assertion see

1) [Abed, Y., N. Goyette, et al. \(2004\)](#). "A reverse genetics study of resistance to neuraminidase inhibitors in an influenza A/H1N1 virus." *Antivir Ther* **9**(4): 577-581

2) [Okomo-Adhiambo, M., G. J. Demmler-Harrison, et al. \(2010\)](#). "Detection of E119V and E119I mutations in influenza A (H3N2) viruses isolated from an immunocompromised patient: challenges in diagnosis of oseltamivir resistance." **360**(25): 2605-2615. *Antimicrob Agents Chemother*

3) [Whitley, R. J., C. A. Boucher, et al. \(2013\)](#). "Global assessment of resistance to neuraminidase inhibitors, 2008-2011: the Influenza Resistance Information Study (IRIS)." **283**(8): 1016-1024. *Clin Infect Dis*

This mutation was not tested for in trial 815 but examination of the IC50 (concentration with 50% inhibition of influenza virus) for Tamiflu and peramivir (see section 3.2.3 below) shows resistance was widespread in this trial as well. In addition, the literature shows that for the period of trial 815 (November 2008 - May 2009) the predominant circulating influenza strain contained the H275Y mutation (Kohno, Yen et al. 2011).

The disposition of patients was as tabulated in table 2.2.3 G. As might be expected in 14 day trials, 94% to 98% of subjects completed the study. (In this table subjects were counted as censored at day 14 if they did not heal by the end of the study and had a measurement after 318 hours after the start of treatment. These subjects can be considered as completing the study but not being observed to heal.) These results are obtained from the applicant's computer files and differ by one or two in some cases from their written report.

TABLE 2.2.3 G
FINAL DISPOSITIONS OF SUBJECTS, ALL TRIALS

TRIAL 722	IV_300		IV_600		PLACEBO	
SCREENED	300					
RANDOMIZED, ITTI	99		98		100	
COMPLETE	94	94.9%	93	94.9%	96	96.0%
CENSOR_DAY_14	3	3.0%	1	1.0%	3	3.0%
I/E_VIOLATE	0		1	1.0%	0	
AE	0		0		1	1.0%
LTFU	2	2.0%	3	3.1%	0	
TRIAL 815	IV_300		IV_600		TAMIFLU	
SCREENED	1099					
RANDOMIZED, ITTI	364		364		365	
COMPLETE	330	90.7%	335	92.0%	331	90.7%
CENSOR_DAY_14	12	3.3%	14	3.8%	16	4.4%
I/E_VIOLATE	1	0.3%	2	0.5%	0	
AE	9	2.5%	10	2.7%	9	2.5%
LTFU	12	3.3%	3	0.8%	9	2.5%
AE=adverse event, LTFU=loss to follow-up before hour 318, I/E VIOLATE=IE or protocol violation						

TABLE 2.2.3 G (continued)
 FINAL DISPOSITIONS OF SUBJECTS, ALL TRIALS

TRIAL 211	IM_150		IM_300		PLACEBO	
SCREENED	344					
ENROLLED	343					
RANDOMIZED, ITTI	104		106		109	
COMPLETE	97	93.3%	101	95.3%	98	89.9%
CENSOR_DAY_14	6	5.8%	2	1.9%	9	8.3%
AE	0		1	0.9%	0	
LTFU	1	1.0%	2	1.9%	2	1.8%
TRIAL 311	IM_300		PLACEBO			
SCREENED	83					
RANDOMIZED, ITTI	57		25			
COMPLETE	53	93.0%	24	96.0%		
CENSOR_DAY_14	1	1.8%	0			
LTFU	3	5.3%	1	4.0%		
TRIAL 212	IM_600		PLACEBO			
ENROLLED	405					
RANDOMIZED, ITTI	160		174			
COMPLETE	158	98.8%	172	98.9%		
CENSOR_DAY_14	0		1	0.6%		
LTFU	2	1.3%	1	0.6%		

AE=adverse event, LTFU=loss to follow-up before hour 318,
 I/E VIOLATE=IE or protocol violation

These trials all made extensive use of the rapid antigen test (RAT) and thus had much lower rates of influenza like illness (ILI) than did previous NDAs for influenza drugs.

2.2.4 Summary of Methods of Assessment

2.2.4.1 Schedule of Measurements and Assessment of Treatment Effects

The primary endpoint in all five trials is based on a diary, filled out twice daily and assessing the severity of seven symptoms (cough, stuffy nose, sore throat, myalgia, fatigue, feverishness, and headache), each on a four point scale (none, mild, moderate, severe). Body temperature was also measured twice daily. There is a slight difference in the wording of the diary question compared to the Tamiflu trials which also had a twice daily, 7 symptom, 4 point diary.

The peramivir diaries asked 'How bad is the symptom now?'; the Tamiflu diaries asked 'How bad has the symptom been during the last 12 hours?'

Symptoms can moderate and then worsen before final alleviation, as the diaries demonstrate. The peramivir wording allows the possibility that moderate to severe symptoms occurring 6 hours after one diary entry and 6 hours before another diary entry might be overlooked. The Tamiflu wording depends on reliability of memory of still somewhat ill subjects. Because the trials were blinded, neither choice of wording should produce biased results. The difference in the wording will increase the difficulties in making cross-NDA comparisons between Tamiflu and peramivir.

Trial 722 also assessed ability to perform daily activities. The latter endpoint was assessed once daily using a 10 point, visual analogue scale.

Nasal and throat swabs for viral load measurements were taken at start, middle, and late visits in all 5 trials. The exact days were screening and days 3, 5, and 9 for trial 722; at screening and on days 3 and 8 (and, if possible, day 2) for trial 815; at screening and on days 3, 4, and 9 in trial 212; on days 1, 5, and 9 in trial 211; on days 1, 3, 5, and 9 in trial 311.

Trials 722, 815, and 211 also measured time to resumption of normal activities on a 10 point visual analog scale, with 10 corresponding to normal activity. Trial 311 measured vaguely similar e
ed days

(b) (4)

2.2.5 Summary of Statistical Analysis

The primary analysis population is all subjects with confirmed influenza and at least one dose of drug. The second restriction only matters in one arm of one trial because there only is one dose of drug except for the Tamiflu arm in trial 815.

The primary endpoint was the time until all seven symptoms were mild or none for two consecutive diary entries at least 21.5 hours apart. This is referred to as time to healing or time to symptom alleviation. For computational purposes, the time to healing is defined as the first of those two consecutive diary entries with no moderate, severe, or missing symptom scores. This is slightly different from the computational rule in Tamiflu NDA, which used the last diary entry with any moderate, severe, or missing symptom score just prior to two consecutive diary entries with all symptoms mild or none. The difference reflects the different wording of the questions: 'How do you feel now?' vs 'What is the worst you felt since the last diary entry?' The latter wording, used by Tamiflu, suggests healing occurred by the time of the last entry with moderate or severe symptoms. Because the trials are double blind, either wording and either computational rule should yield unbiased results. The Tamiflu computational rule would lead to healing times approximately 12 hours shorter than the peramivir rule so one must be cautious about comparing results across NDAs.

The times to healing were analyzed by the following procedures. In trial 722, the test and confidence intervals are obtained from the hazard ratio computed by Cox proportional hazards regression, stratified by smoking status and composite symptom score at baseline. Both strata were binary with composite symptom score being either \leq or $>$ 14. The multiplicity problem created by the presence of two peramivir arms was dealt with by making the primary comparison placebo to the pooled 300mg and 600mg arms.

Trial 211 also based its conclusion on a Cox regression stratified by smoking status. (Composite symptom score was not used in the randomization and thus was not appropriate for the analysis.) A Bonferroni adjustment was used for the two peramivir arms. Trial 311 also specified a Cox regression with Bonferroni adjustment in its protocol. Due to the early termination with

sample size 10% of that planned, no testing at all was done in this trial. Instead the results were pooled with trial 211 and the methodology of 211's protocol was used for the pooled data.

Trial 815 also based its primary analysis on a Cox regression. Because this trial had an active control, Tamiflu, a non-inferiority comparison was proposed. A conclusion of efficacy for either the 300 mg or the 600 mg arm would be made if the 97.5% upper bound for the hazard ratio of that arm to Tamiflu was <1.17 (i.e. Tamiflu subjects healed no more than 1.17 times as fast as peramivir subjects. The derivation of the limit of 1.17 came from pooling results reported in the Tamiflu NDA and further discussion of the non-inferiority comparison will be given in section 3.2.3 below. One may briefly note that using 97.5% upper bound makes no adjustment for multiplicity. The appropriate confidence limit would have been 98.75%.

The protocol specified analysis in trial 212 differed from those in the other trials in two ways. First, the test for efficacy was based on the Wilcoxon-Gehan test, stratified by current smoking status, rather than Cox regression. Second, the primary comparison was restricted to type A influenza patients rather than all influenza patients. This trial had only one peramivir arm so no multiplicity adjustment would be needed.

2.2.6 Summary of Applicant's Results

2.2.6.1 Trial 722

Efficacy results for trial 722 are summarized in table 2.2.6 A. This table gives the median times to healing (symptom alleviation), which were similar for the two peramivir doses and 20-22 hours shorter than for placebo. It also gives the 95% confidence limits for the median time to healing in each arm. (Confidence limits on the difference in medians are not available for reasons given below.) Finally, the table also includes the hazard ratios of time to healing for each peramivir arm relative to placebo along with 95% confidence limits for these ratios. The values of .66-.68 for these ratios mean that the probability that a still symptomatic placebo subject will heal any given hour is about two-thirds of the probability that a still symptomatic peramivir subject will heal in that hour. Finally, the table gives the p-values for testing that the hazard ratios are <1.

TABLE 2.2.6 A
EFFICACY SUMMARY, TRIAL 722

	Peramivir 300mg IV	Peramivir 600mg IV	Placebo
Median Healing	59.1 hrs	59.9 hrs	81.8 hrs
95% Limits	50.9-72.4	54.4-68.1	68.0-
101.5			
Hazard Ratio	.681	.666	
95% Limits	.511-.909	.499-.89	
P-value	.0046	.0046	

(A technical note on why there are no confidence intervals for the difference in medians: The computation of Kaplan-Meier curves with time on the horizontal axis and percent healed on the vertical axis also yields standard errors for the percent healed at any time on each arm. These standard error bars extend up and down from the Kaplan-Meier curve. By combining the standard errors, one can compute a vertical confidence interval for the difference in, or ratio of, percent healed at any time between any two arms. The hazard ratio and its confidence intervals reflect an average of this ratio of percent healed over all times.

A sufficiently wide multiple of these standard errors yields simultaneous confidence bands which can also yield confidence limits horizontally for the time that a given percent, say 50%, are healed on any arm. These horizontal limits cannot be combined between since there is no horizontal standard error. Thus there is no confidence interval for the difference in medians.

It is possible to get a confidence interval for the difference in the medians not by calculating standard errors but by the method of the bootstrap. The applicant did use this method for all trials pooled together.)

Among secondary endpoints, time to resolution of fever was reduced from 42.4 hours with placebo to 29.3-30.2 hours in the two peramivir arms.

2.2.6.2 Trials 211 and 311

Because trial 311 did not achieve the sample size planned in the protocols (800 subjects planned, 82 achieved), the applicant analyzed the results from these two trials as one pooled database. Efficacy results for these trials are summarized in table 2.2.6 B. The first half of this table gives the results for both trials pooled together; the second half gives the results for trial 211 alone.

The table gives the median times to healing (symptom alleviation), which were similar for the two peramivir doses and 20 hours shorter than for placebo. It also gives the 95% confidence limits for the median time to healing in each arm. The table also includes the hazard ratio of time to healing for both peramivir doses relative to placebo along with the 95% confidence limits. Trial 311 included only 300 mg peramivir so the pooled half of the table only gives the hazard ratio for 300 mg; the trial 211 alone half of the table gives hazard ratio for both doses. The values of .82, .84, .86 for these ratios mean that the probability that a still symptomatic placebo subject will heal any given hour is about four/fifths of the probability that a still symptomatic peramivir subject will heal in that hour. The p-value for testing difference in the times to healing is based on the Wilcoxon-Gehan statistic, stratified by smoking status, influenza season, and influenza type. (The p-value for the Cox regression is not reported even though that was the protocol specified test. The Wilcoxon-Gehan test was the primary analysis in the protocol for trial 212. The applicant also reported an (b) (4) which is s nappropriate, because that analysis yielded a (b) (4) That result can be ignored.)

TABLE 2.2.6 B
EFFICACY SUMMARY, TRIALS 211,311 POOLED
Peramivir 150mg IM Peramivir 300mg IM Placebo

Pooled			
Median Healing	114.1 hrs	113.2 hrs	134.8
hrs			
95% Limits	95.2-145.5	88.4-130.4	113.5-
			163.8
Hazard Ratio		.838	
95% Limits		.648-1.085	
P-value		.161	
Trial 211 alone			
Median Healing	114.1 hrs	117.4 hrs	136.2
hrs			
95% Limits	95.2-145.5	78.0-135.9	114.3-
			165.8
Hazard Ratio	.859	.816	
P-value	.315	.18	

Among secondary endpoints, time to resolution of fever was reduced from 66.8 hours with placebo to 42.8 hours in the 300mg arm and to 51.7 hours in the 150mg arm. The 300mg peramivir arms, pooled, resulted in statistically significant reductions in time-weighted change from baseline viral titer compared to placebo and a lower proportion of peramivir-treated subjects who were shedding virus at Day 2 and 3.

2.2.6.3 Trial 212

Efficacy results for trial 212 are summarized in table 2.2.6 C. This table gives the results for the protocol specified analysis using only type A patients first and then the results for a secondary analysis using all influenza patients. Each half gives the median times to healing, which were about 16 hours shorter for peramivir than for placebo. It also gives the 95% confidence limits for the median time to healing in each arm. Finally, the table also includes the hazard ratios of time to healing for the peramivir arm relative to placebo along with 95% confidence limits for this ratio. The values of .72-1.188 for this ratio means that type A peramivir subjects healed between 28% faster and 19% slower than type A placebo subjects. Ignoring type, peramivir subjects healed between 21% faster and 25% slower than placebo subjects. Finally, the table gives the p-values for testing (with the Wilcoxon-Gehan test) that the healing rates are the same in both arms as opposed to being faster on peramivir.

TABLE 2.2.6 C
EFFICACY SUMMARY, TRIAL 212
Peramivir 300mg IMPlacebo

Protocol Analysis, Type A only		
N	132	147
Median Healing	91.1	106.9
95% Limits	77.7 - 109.7	90.4 - 127.4
Hazard Ratio	.927	
95% Limits	.723 - 1.188	
P-value	.222	
All Subjects		
N	159	172
Median Healing	92.6	107.1
95% Limits	81.7 - 114.3	90.5 - 122.5
Hazard Ratio	.995	
95% Limits	.791 - 1.251	
P-value	.31	

Among secondary endpoints, time to resolution of fever was reduced from 66.8 hours with placebo to 42.8 hours in the 300mg arm and to 51.7 hours in the 150mg arm.

2.2.6.4 Trial 815

Efficacy results for trial 815 are summarized in table 2.2.6 D. This table gives the median times to healing, which were about the same in all three arms. It also gives the 95% confidence limits for the median time to healing in each arm. Finally, the table also includes the hazard ratios of time to healing for the peramivir arm relative to Tamiflu along with 95% confidence limits for this ratio. The applicant asserted that a 95% upper bound for this ratio $< \text{[REDACTED]}^{(b)(4)}$ would be demonstrative of non-inferiority. The comparison for non-inferiority does not require a p-value.

TABLE 2.2.6 D
EFFICACY SUMMARY, TRIAL 815

	Peramivir 300mg IV	Peramivir 600mg IV	Tamiflu
--	--------------------	--------------------	---------

Pooled			
Median Healing	78.0 hrs	81.0 hrs	81.8 hrs
95% Limits	68.4-88.6	72.7-91.5	73.2-
			91.1
Hazard Ratio	.946	.970	
95% Limits	.793-1.129	.814 -1.157	

The applicant's $\text{[REDACTED]}^{(b)(4)}$ non-inferiority margin is based on an estimated upper confidence limit on the hazard ratio of Tamiflu to placebo of .84. The log hazard ratio has a normally distributed hazard ratio so the computation of the non-inferiority bound using the methodology of the FDA guidance would involve taking one half of the upper limit; the computation would be $\log \text{ hazard ratio of tamale to peramivir} = .5 * \log(.84) = .5 * (-.0757)$; hazard ratio of Tamiflu to peramivir = .917 or hazard ratio of peramivir to Tamiflu = $1/.917 = 1.09$. Thus, the applicant's proposed margin $\text{[REDACTED]}^{(b)(4)}$ indicated by the guidance; the applicant's methodology would not yield a conclusion of non-inferiority.

In point of fact, there are other more serious problems with the non-inferiority conclusion from this trial; specifically the likelihood that Tamiflu was not superior to placebo with the strain of influenza circulating at the time and place of this trial. See the FDA analysis in section 3.2.3 below for a more reliable analysis of this trial.

2.2.6.5 Uncontrolled Trials 816 and 918

Trial 816 enrolled 42 subjects in a randomized comparison of 300mg IV and 600mg IV peramivir among subjects with high risk factors such as poorly controlled diabetes, chronic respiratory disease, or the use of immune-suppressive drugs. In both arms, the drugs were administered for 1 to 5 days at the discretion of the investigator, who was supposed to rely on continuance of fever or other clinical symptoms. Results are given in table 2.2.6 E. Note that the confidence level reported is 90%, not 95%.

TABLE 2.2.6 E
EFFICACY SUMMARY, TRIAL 816

	Peramivir 300mg IV	Peramivir 600mg IV
Median Healing	114.4 hrs	42.3 hrs
90% Limits	40.2-235.3	30.0-82.7

Because of the small sample size, no testing was done to compare the arms. There is a suggestion of a dose effect but the uncontrolled nature of the administration of drug on days 2-5 and the wide confidence intervals both make it impossible to be confident about the reliability of this observation.

Trial 918 enrolled 117 pediatric subjects in a single arm study with all subjects receiving 10mg/kg of IV peramivir, up to a maximum of 600mg. There is no comparison possible to determine efficacy. There was reasonable consistency across the four age categories. (b) (4)



TABLE 2.2.6 E
MEDIAN TIME TO HEAL, BY AGE, TRIAL 918

	<2 yrs	2-5 yrs	6-11 yrs	12-15 yrs
N	12	20	46	37
Median Healing	31.0	26.4	25.6	29.1
95% Limits	21-51	18-69	21-32	21-36

2.2.7 Summary of Applicant's Conclusions

The applicant concluded that trial 722 showed that peramivir was effective in the treatment of acute uncomplicated influenza using a single intravenous doses of 300 mg or 600 mg.

Trial 815 showed that single IV doses of peramivir 300 mg and 600 mg were non-inferior to 75 mg of oseltamivir phosphate administered orally twice daily for 5 days.

Trials 211 and 311 pooled demonstrated clinical benefit with peramivir treatment over placebo for the treatment of acute uncomplicated influenza. Substantial reductions in the median time to alleviation of influenza symptoms and resolution of fever were observed with peramivir treatment (150 mg or 300 mg) compared to placebo; both endpoints were affected by peramivir treatment in a dose-proportional manner. Also the 300 mg peramivir arm was statistically significantly superior to placebo with respect to secondary viral titer endpoints.

Trial 212 showed an improvement in the time to alleviation of symptoms was observed for subjects treated with 600 mg peramivir (91.1 hours) compared to placebo (106.9 hours), although this difference was not statistically significant. No statistically significant differences were observed between treatment groups for the secondary or exploratory endpoints measured.

Trial 816 showed that 300mg and 600mg IV peramivir were associated with a rapid recovery from influenza symptoms in high-risk subjects with influenza infection, with evidence of enhanced efficacy with the 600-mg dose.

Trial 918 showed

(b) (4)

An integrated summary of the median time to healing, as a function of peramivir dose is given in table 2.2.7 A. The bootstrap method is used to get 95% confidence limits on the difference in the medians of each peramivir dose to placebo. Negative values of the difference correspond to reductions in time to healing on peramivir.

TABLE 2.2.7 A
 MEDIAN HEALING TIMES BY PERAMIVIR DOSE

	Placebo	150mg	300mg	600mg
Median Healing	107.4	114.1	84.1	79.4
95% Limits	96-116	95-146	69-102	69-92
Change from Placebo		5.4	-24.8	-28.3
95% Limits		-28,+40	-41,-4	-43,-13

The results show a dose response relationship with statistically significant differences at doses of 300mg and 600mg. (The FDA reviewer notes that comparisons are being made here which involve subjects who are not assigned by a single randomization and that placebo times are drawn from several influenza seasons with different circulating strains.)

The collectivity of evidence supports efficacy of 600mg IV peramivir in the treatment of acute uncomplicated influenza.

3. Statistical Evaluation

This NDA depends largely on a single trial, the Shionogi trial 722. The FDA's computation of the protocol-specified Cox regression comparing 600mg IV peramivir to placebo in that trial yields a p-value of .0013. A second Cox regression, comparing any peramivir (300mg or 600mg) to placebo yields a p-value of .00057, which is smaller than the value of $.025^2 = .000625$ conventionally required for approval. If anything, one would expect the 300mg dose of peramivir to be less effective than the 600mg dose. Thus, trial 722 by itself is a convincing demonstration of the efficacy of 600mg IV peramivir against acute uncomplicated influenza.

Nonetheless, it is still worthwhile to review the other four trials to examine the extent that they support or qualify the conclusions reached from trial 722. In particular, this review will show that peramivir appears to be ineffective against type B influenza and against some Tamiflu resistant strains of type A influenza. Also, it appears that the Japanese subjects enrolled in trial 722 received drug earlier than Americans in the BioCryst trials, that they had less severe disease, and that their benefit, relative to placebo was greater.

There are a number of issues that need to be addressed in this review before accepting the applicant's conclusions. First, there are difficulties reproducing the individual times to healing used in the applicant's analysis. A number of the times to healing reported in the applicant's subject level data sets are not compatible with the twice daily symptom scores reported in the applicant's diary entry level data sets.

The review will begin with a discussion of the extent to which the FDA was able to reproduce the applicant's times to healing statistics from their raw datasets.

This will be followed by the FDA statistical analysis which will include Kaplan-Meier plots of the times to symptom alleviation, Cox regressions comparing the hazard rates for healing of peramivir and placebo, an analysis of trial 815 and the possible non-inferiority of peramivir to Tamiflu, a comparison of results in influenza types A and B, comparison of the confidence intervals for the median times to healing in peramivir and placebo, and a modelling of the hazard ratios of

peramivir to placebo as a function of peramivir dose from 150mg to 600mg.

3.1 Discussion of Missing and Incorrect Times in Diary Entry Records

The applicant's primary dataset contains (after some simple preliminary manipulation) one record for each scheduled diary entry for each subject. There are 28 records per subject, from day 1 pre-dose to evening, day 14. Each record contains the scores from 0=mild to 3=severe for all of the seven symptoms (some values are missing for some times), the name and number of the visit, and the date and time of the visit. The latter variable is called ADTM and takes on values like 14AUG08:16:40. The FDA has used this dataset to compute the times to healing and a variable =1 or 0, according as the time of healing is observed or censored. The time of healing was defined as the first of two visits with all symptoms mild or none. There was an additional proviso that the later of those two visits with all symptoms mild or none had to be at least 21.5 hours after the last visit with at least one moderate or severe symptom. The ADTM variable was used to keep track of this waiting time as well as the time since first entry.

The rule used in the Tamiflu NDA was different. The Tamiflu questionnaire asked what was the worst severity of the symptom in the previous 12 hours. Therefore, healing was calculated to occur at the last visit where at least one symptom was moderate, severe or missing prior to two consecutive visits where all symptoms were observed mild or none because from that point on later diary entries indicate symptoms are alleviated. This rule would, in general, make the times to healing 12 hours shorter than those with the peramivir rule but would have little or no effect on the comparison between the arms. Some of the computations below were repeated using the Tamiflu rule and confirmed that no different conclusions were reached with respect differences between arms. Those repeat computations are not included here.

There are some problems with missing and incorrect data in the applicant's dataset containing diary entry results. These are documented here.

The number of records with observed and missing values of the ADTM variable by STUDYID is as follows

STUDYID	N_OBS	N_MISS
0722T0621	6742	33
BCX1812-212	8868	23
0815T0631	23197	0
BCX1812-211	7559	0
BCX1812-311	1797	0

These 56 records come from one subject in trial 212 and 33 subjects in trial 722. In trial 212, all 23 records are from one subjects, SUBJID= 432.038. This subject has no valid observations of symptoms although he was treated and in the ITTI population. This subject should be treated as having healing time censored at time 0.

In trial 722, there are three subjects: AA1.061-1 (treated at 11:35 on 9-Jan-08), CD1.155-2 (treated at 10:20 on 2-Feb-08) and CS1.180-5 (treated at noon on 13-Mar-08), who have no diary after day 1, pre-dose. They are also all in the ITTI population.

Also in trial 722, there are 30 subjects who have their pre-dose symptom scores recorded later than 6 PM on their first day. These subjects all have a record for day one, post-dose, symptom score with blank ADTM and blank symptom scores. In fact, it is reasonable to assume that the first measurement post-dose was actually made on the morning of day 2 and that the record for day 1, post-dose should not have been included in the dataset at all.

There are also a number of observations where the ADTM is not missing but is clearly incorrect because the ADTM time for one visit is earlier than the ADTM time for the preceding visit. This occurs for 4 subjects in trial 815, no subjects in trial 722, 6 subjects in trial 311, 29 subjects in trial 211, and 39 subjects in trial 212. This inconsistency can be observed to occur for one of three reasons.

1. There are 3 or 4 diary entries all on the same day with the next entry occurring two days later. This review assumes that the third and fourth entries on the same day are actually supposed to be on the next day, where no entries were recorded.

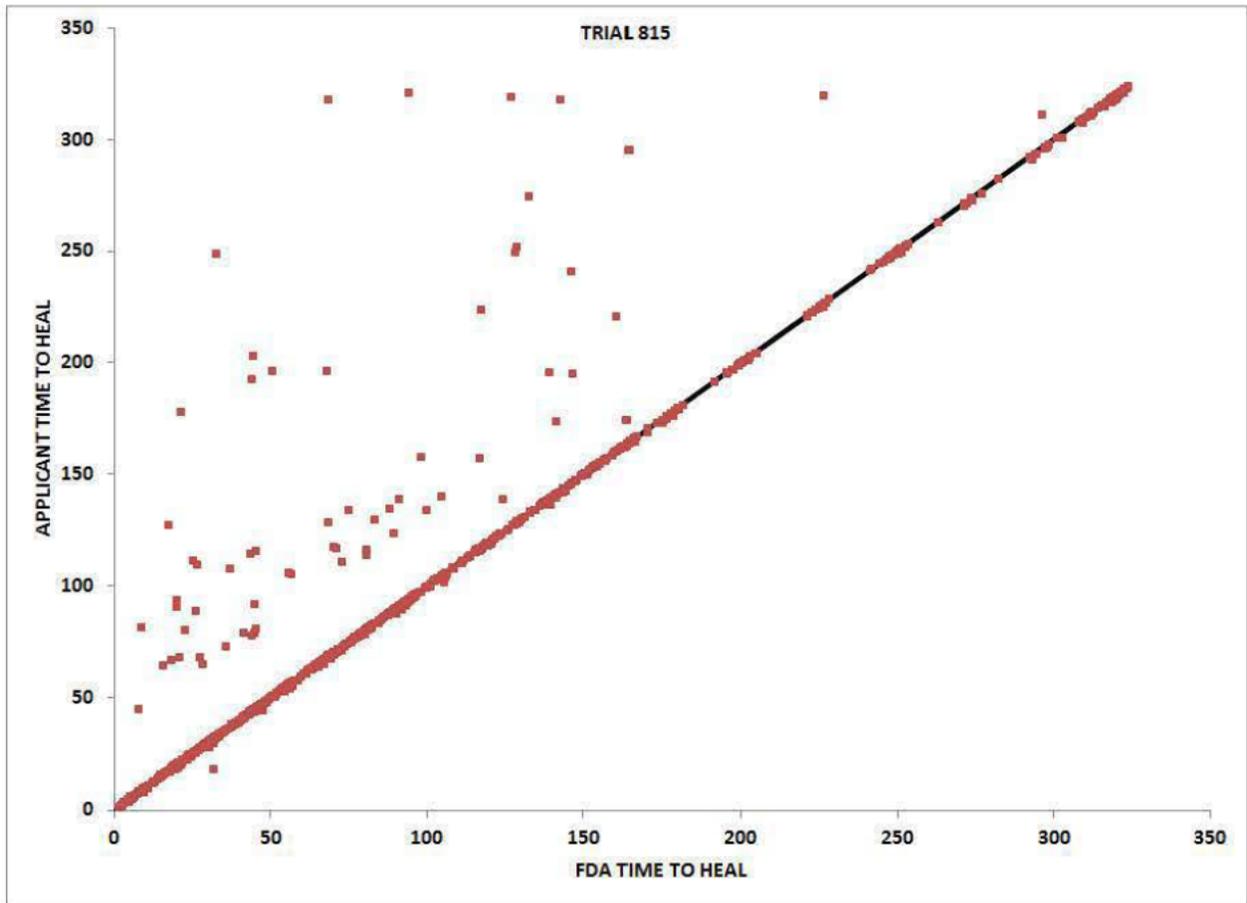
2. There are entries for morning and evening with both times later than noon; usually both are much later, around 6 - 9 pm. This review assumes a morning entry at 7 pm followed by an evening entry also at 7 pm or even at 6:30 pm should really be a morning entry at 7am followed by an evening entry at 7pm.

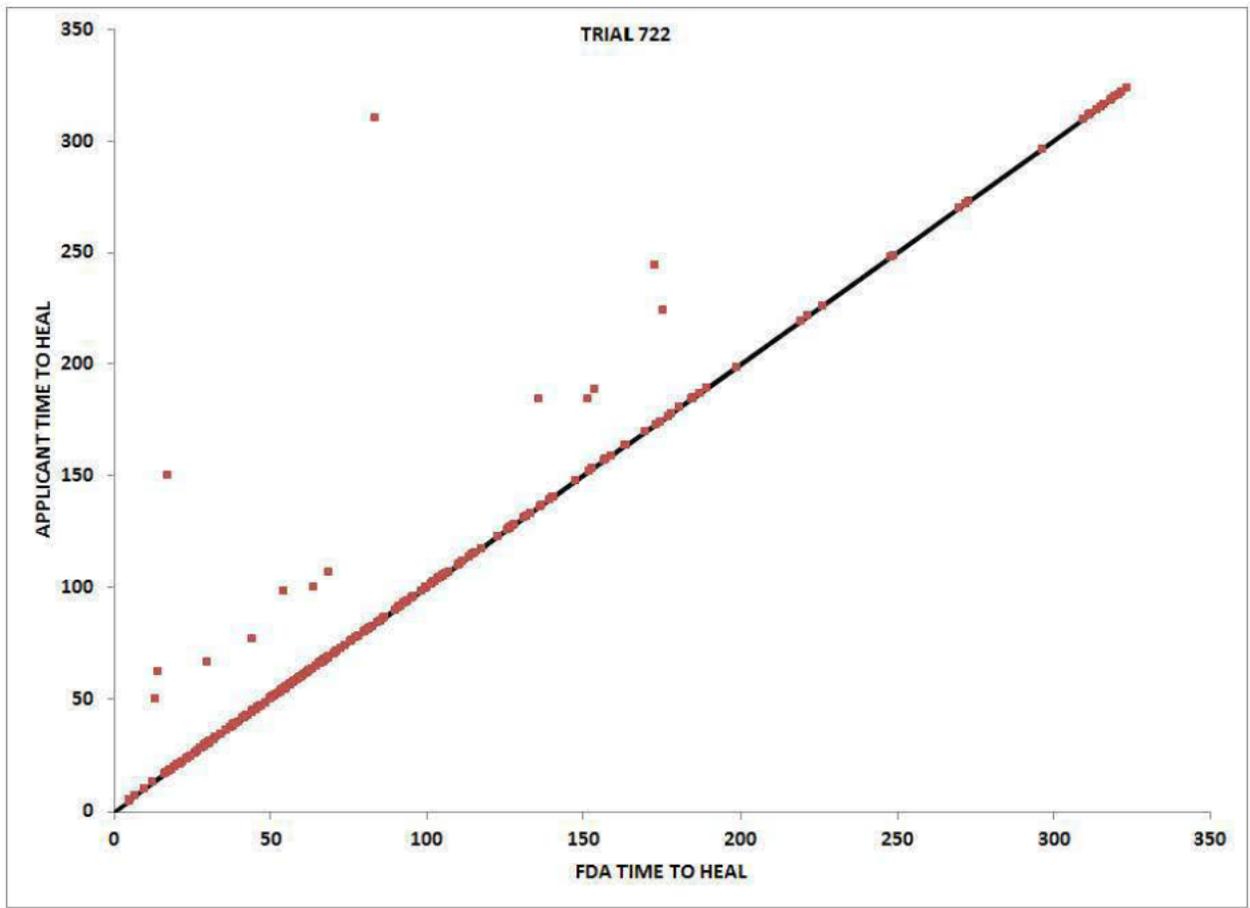
3. There are entries for morning and evening with both times earlier than noon. This review assumes a morning entry at 7 am followed by an evening entry also at 7 am should really be a morning entry at 7 am followed by an evening entry at 7pm.

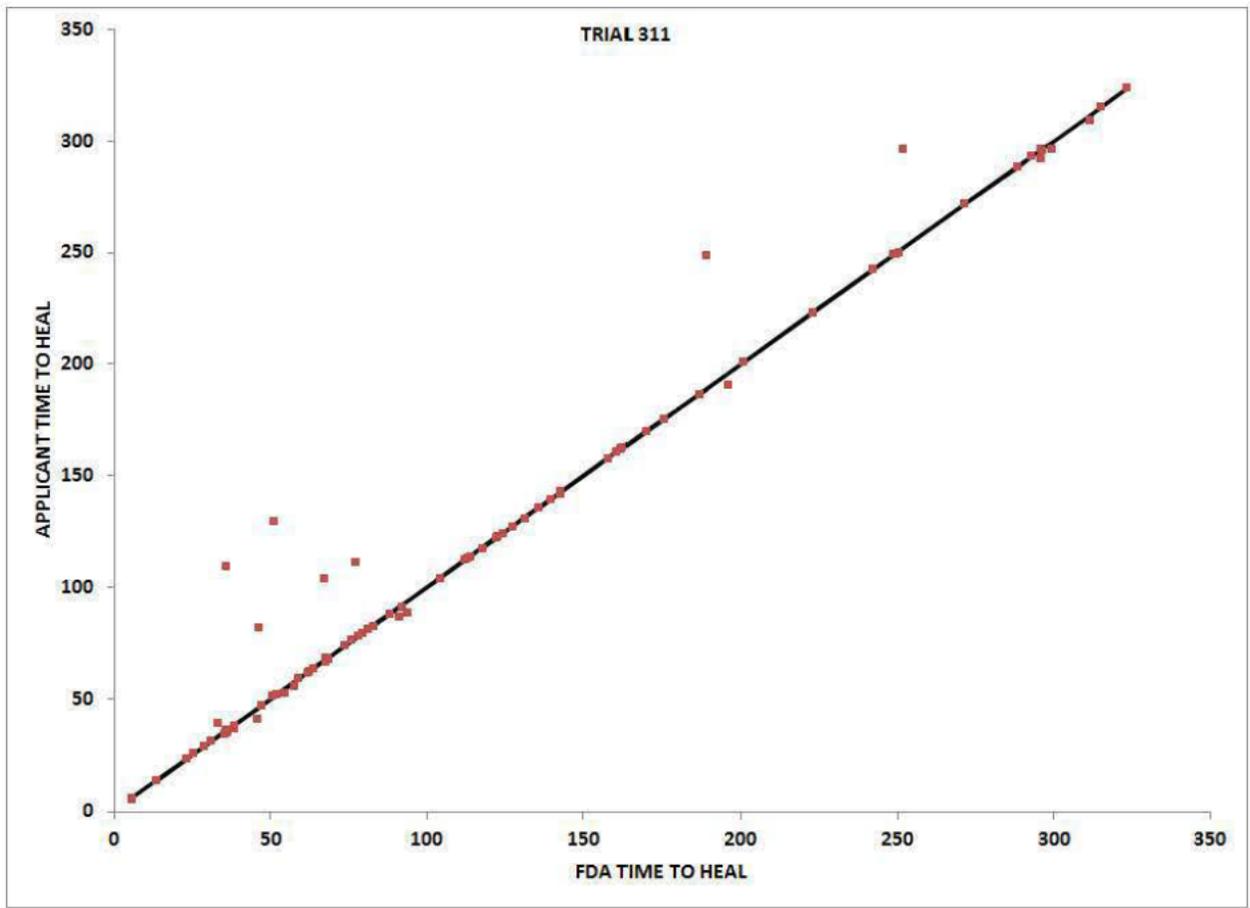
The differences between the FDA computations and the applicant's report with respect the number of censored and observed times to heal are given in the following table. Here 0 stands for censored times, 1 for observed times. The total number of disagreements between the FDA and the applicant is one out of 297 in trial 722, 11 out 1093 in trial 815, 5 out of 319 in trial 211 (if one corrects missing to 0), 29 out of 334 in trial 212 and none is trial 311.

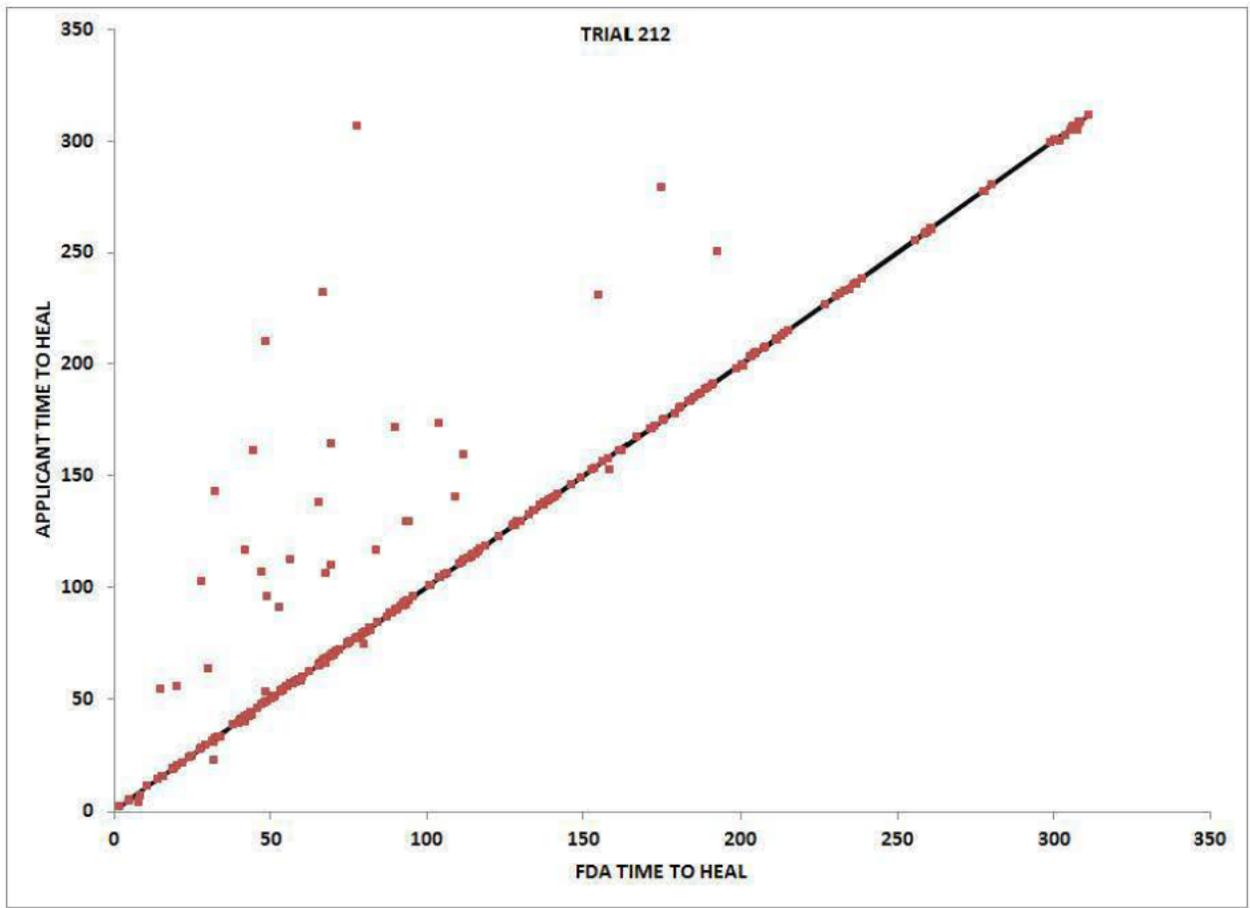
STUDYID	FDA_CENSOR	APPLICANT	COUNT
0722T0621	0	0	17
	1	0	1
	1	1	279
0815T0631	0	0	106
	1	0	11
	1	1	976
BCX1812-211			
	0	.	1
	0	0	50
	0	1	2
	1	0	3
	1	1	263
BCX1812-212			
	0	.	3
	0	0	31
	0	1	27
	1	0	2
	1	1	271
BCX1812-311			
	0	.	1
	0	0	12
	1	1	69

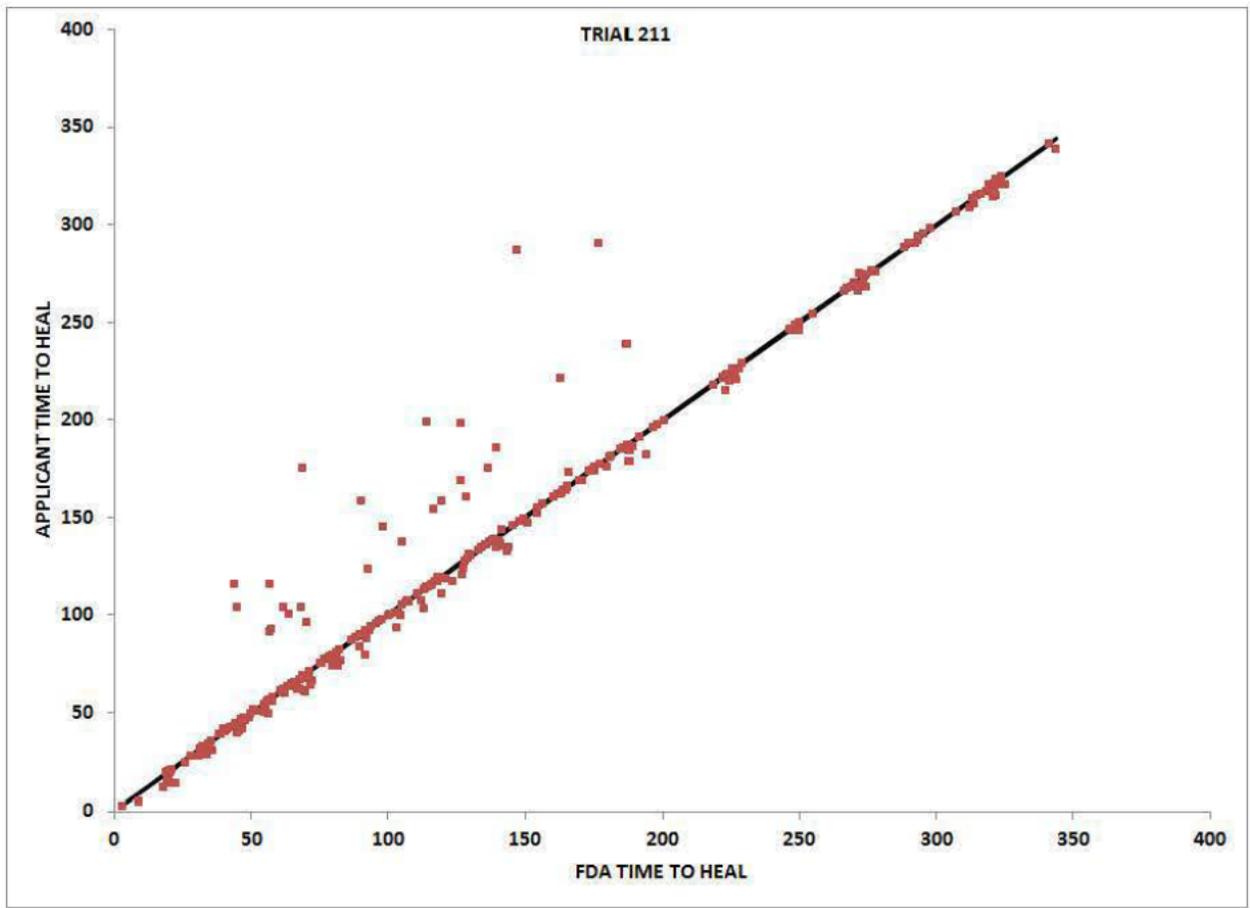
The following graphs show how the times to healing as computed by the FDA reviewer compare to those computed by the applicant for each of the five trials.











One can see that the FDA and applicant times to healing are, in the vast majority of cases, very close together. In a comparative handful of cases the applicant got longer times than the FDA reviewer.

A line listing of two of those cases is given below.

Subject =0815T0631.095.JXV01 FDA time=94.517 Applicant time=320.617

ADTM	HEAL	ELAPSE	SYMPTOM SCORES						
19JAN09:10:25	0	0.000	2	2	0	0	1	1	1
19JAN09:17:40	0	7.250	2	2	2	0	3	3	3
20JAN09:08:43	0	22.300	2	1	2	1	2	2	2
20JAN09:17:02	0	30.617	2	1	1	1	1	2	2
21JAN09:08:17	0	45.867	2	0	1	1	1	2	2
21JAN09:17:00	0	54.583	2	1	1	0	0	2	2
22JAN09:08:15	0	69.833	1	0	1	0	0	2	2
22JAN09:19:47	0	81.367	2	0	1	0	0	1	1
23JAN09:08:56	1	94.517	1	0	0	0	0	1	1
23JAN09:20:20	1	105.917	1	0	0	1	0	0	0
24JAN09:07:46	0	117.350	1	0	0	2	0	0	1
24JAN09:22:55	0	132.500	1	0	0	2	0	0	0
25JAN09:10:20	0	143.917	2	1	0	2	0	0	0
25JAN09:19:15	0	152.833	2	0	0	2	0	0	0
26JAN09:07:40	0	165.250	2	1	0	2	0	0	0
26JAN09:19:40	0	177.250	1	1	0	2	0	0	0
27JAN09:22:15	0	203.833	2	1	0	2	0	0	0
28JAN09:22:11	0	227.767	2	0	0	2	0	0	0
29JAN09:22:10	0	251.750	2	0	0	2	0	0	0
30JAN09:21:40	0	275.250	2	0	0	1	0	0	0
31JAN09:21:45	0	299.333	2	0	0	1	0	0	0
01FEB09:19:58	0	321.550	2	0	0	1	0	0	0

Subject=0815T0631.142.KMP08 FDA TIME=68.5 Applicant TIME=317.367

ADTM	HEAL	ELAPSE	SYMPTOM SCORES						
23APR09:12:30	0	0.000	2	2	1	2	3	2	3
24APR09:08:30	0	20.000	2	1	0	1	1	1	1
24APR09:17:40	1	29.167	1	1	0	1	1	1	1
25APR09:08:10	0	43.667	2	1	0	1	1	1	1
25APR09:18:30	0	54.000	2	2	0	1	1	1	0
26APR09:09:00	1	68.500	1	1	0	1	0	0	1
26APR09:19:20	1	78.833	1	1	0	1	0	0	1
27APR09:08:20	0	91.833	2	1	0	2	1	0	1
27APR09:18:40	0	102.167	2	1	0	2	1	0	0
28APR09:07:50	0	115.333	1	1	0	2	0	0	1
28APR09:19:20	0	126.833	1	1	0	2	1	0	1
29APR09:08:40	0	140.167	1	2	0	2	1	0	1
29APR09:20:00	0	151.500	2	1	0	1	0	0	1
30APR09:08:30	0	164.000	2	1	0	2	0	0	1
30APR09:20:20	0	175.833	2	1	0	1	0	0	1
01MAY09:19:30	0	199.000	2	0	0	2	0	0	1
02MAY09:19:20	0	222.833	2	1	0	1	0	0	1
03MAY09:18:50	0	246.333	2	0	0	1	0	0	1
04MAY09:20:50	0	272.333	2	1	1	1	0	1	1
05MAY09:21:20	0	296.833	2	1	0	0	0	0	1
06MAY09:20:20	1	319.833	1	0	0	0	0	0	0

One can see that the disagreement occurs when subjects heal and then relapse. The applicant has considered the subjects as ill until the last visit (or until a subsequent time after which all visits show suppressed symptoms); the FDA reviewer followed the protocol rule of counting healing at the first of two

consecutive visits with suppressed symptoms, regardless of what happened afterwards. The primary results that are given in the remainder of this review will use the FDA reviewer's computations. The applicant's times are used as a sensitivity analysis in which time to healing is allowed to be considered as delayed by relapses of symptoms.

There is one final point to be made with respect to the data on which the analyses below are based. One investigator, John Michael Wise, was unable to produce the original records when his site was visited by FDA inspectors. Since his data could not be verified, it has been excluded from the FDA analyses. The total number of subjects excluded for this reason were

Trial	Arm	Number_Excluded
212	IM_600_mg_q.d.	4
	IM_Placebo	1
211	IM_150_mg_q.d.	4
	IM_300_mg_q.d.	4
	IM_Placebo	5
311	IM_300_mg_q.d.	3

3.2 FDA Analysis of Time to Symptom Alleviation

The following material will describe the analysis of time to symptom alleviation as performed by the FDA statistical reviewer. The analysis will proceed in the following steps. First, there are Kaplan-Meier curves for time to symptom alleviation, plotted for each trial for all the arms in that trial. The trials covered will be 722 (600 and 300 mg IV vs placebo), 212 (600 and 300 mg IM vs placebo), and 211 plus 311 pooled (300 and 150 mg IM vs placebo). For brevity, the pooling of trials 211 and 311 will be referred to as trial 411.

Trial 815 had no placebo arm, comparing 300 and 600 mg IV peramivir to Tamiflu. There being reason to believe that this trial was conducted in the presence of a Tamiflu resistant strain of influenza, the analysis of trial 815 will be deferred to section 3.2.3.

Second, there will be plots, related to the Kaplan-Meier plots of the 95% confidence bands for the difference between the percent healed on each peramivir minus the percent healed on the control arm. Usually, these confidence bands are plotted with time on the x-axis. When the disease is influenza in otherwise healthy adults, essentially everyone heals and the placebo and peramivir Kaplan-Meier curves for time to healing will converge again at 100% healed. Furthermore, the times to healing vary quite a bit from one trial to the next, depending on the severity of the disease that season. Therefore, the 95% confidence bands for the difference will also be plotted with percent healed on peramivir on the x-axis. This is a more informative way to describe the region where peramivir patients are healing faster than placebo patients and one that doesn't vary with disease severity.

Third, results will be presented from Cox regressions of time to healing on treatment. The 95% confidence bands on the differences associated with the Kaplan-Meier curves do not yield confidence intervals for the difference between the median time to healing. Rather they yield a confidence interval for the percent healed on placebo at the time that 50% of the peramivir patients have healed (=median time for healing on peramivir). The Cox regressions generalize to all percents healed by fitting a model where percent not yet healed at time on control = percent not yet healed at time t on peramivir raised to the power H. I.e. $Q(t) = P(t)^H$ where Q and P are the percents not yet healed on control and peramivir at time t. H is called the hazard ratio

and if $H < 1$ then Q , percent not healed on control, is greater than P , percent not healed on peramivir. The Cox regressions will give point estimates and 95% confidence intervals for the hazard ratio H . Any trial where the confidence interval for H lies entirely less than one is a trial where peramivir is statistically significantly superior to placebo in time to healing.

This discussion of Cox regressions will also compare results using the corrected times to healing as computed by the FDA statistical reviewer (as discussed in the previous section) with the results using the uncorrected times reported by the applicant.

Fourth, this review will compute the hazard ratio for Tamiflu relative to placebo from the data in the Tamiflu NDA. This will permit one to calculate a confidence interval for the hazard ratio of peramivir to placebo from the data in the Tamiflu controlled trial, 815.

Fifth, there will be a comparison of results in influenza types A and B, which will suggest that there is little or no efficacy in type B influenza. Sixth, the review will present a comparison of the confidence intervals for the median times to healing in peramivir and placebo. The peramivir medians will be consistently shorter than the placebo medians. Direct computation of the confidence intervals for the difference in the medians is not possible; such confidence would require computer-intensive bootstrapping. Because confidence intervals on the hazard ratios already provide a test of statistical significance, this has been foregone.

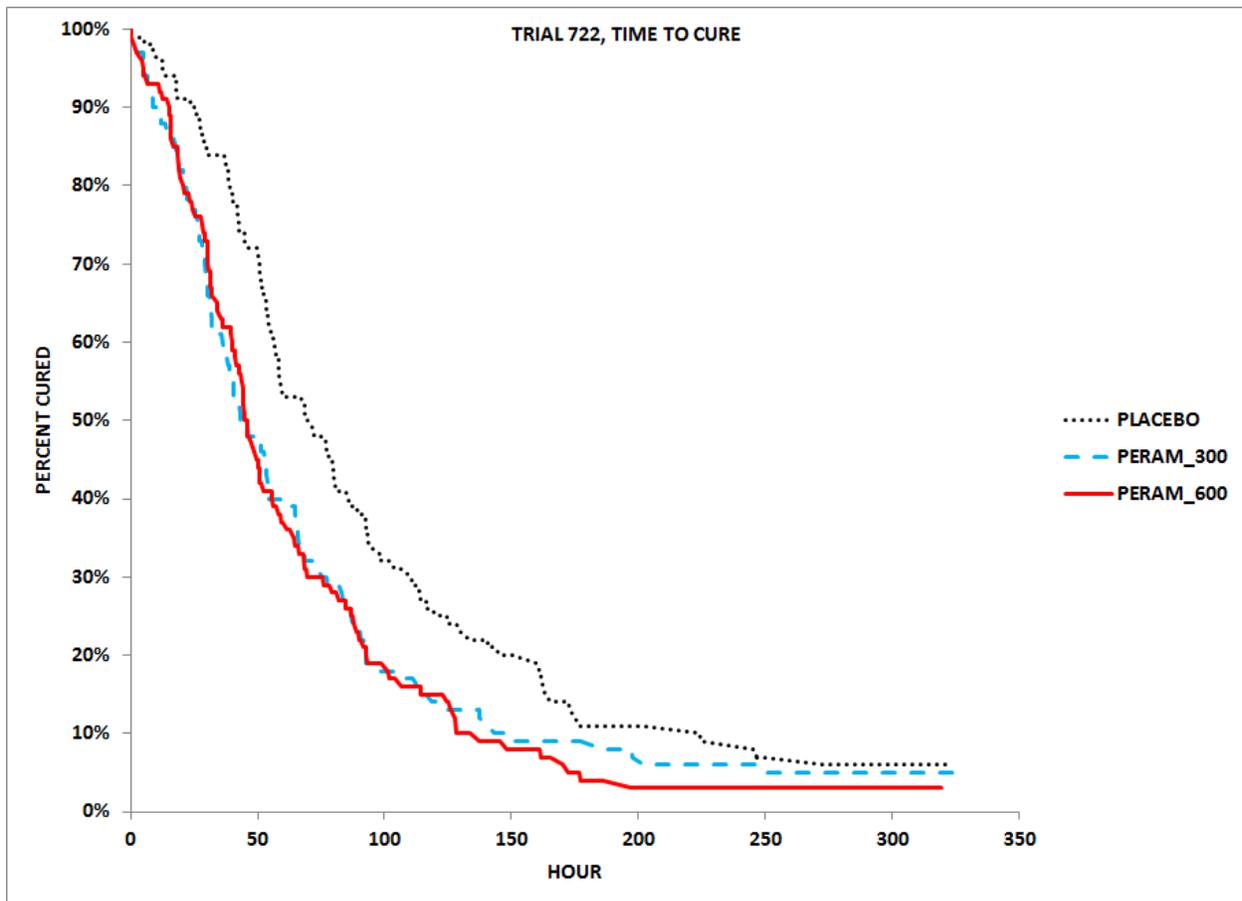
Finally, this review will explore the conclusions that can be drawn from the range of peramivir doses over the five trials presented. This will be a modelling of the log hazard ratios of peramivir to placebo as a function of peramivir dose from 150mg to 600mg. This will show that there is a statistically significant dose response curve with positive slope; i.e. higher hazard ratios (=faster healing times relative to matched placebo) for higher doses of peramivir. The slope would not be statistically different from zero if peramivir were ineffective.

It will be noticed that except for trial 722, the FDA reviewer will conduct a number of analyses on subsets of the data not specified in the protocols. Typically, this is considered statistically risky. This review does this because 1)the protocol specified analysis of trial 722 by itself is adequate to demonstrate approvability with respect to efficacy, 2)none of the subsets were selected on the basis of the performance of peramivir but rather on grounds of Tamiflu resistance, and 3)the subset analyses are mostly used to argue in favor of limits on the efficacy of peramivir rather than as post hoc support of efficacy.

3.2.1 Kaplan-Meier Plots of Time to Symptom Alleviation

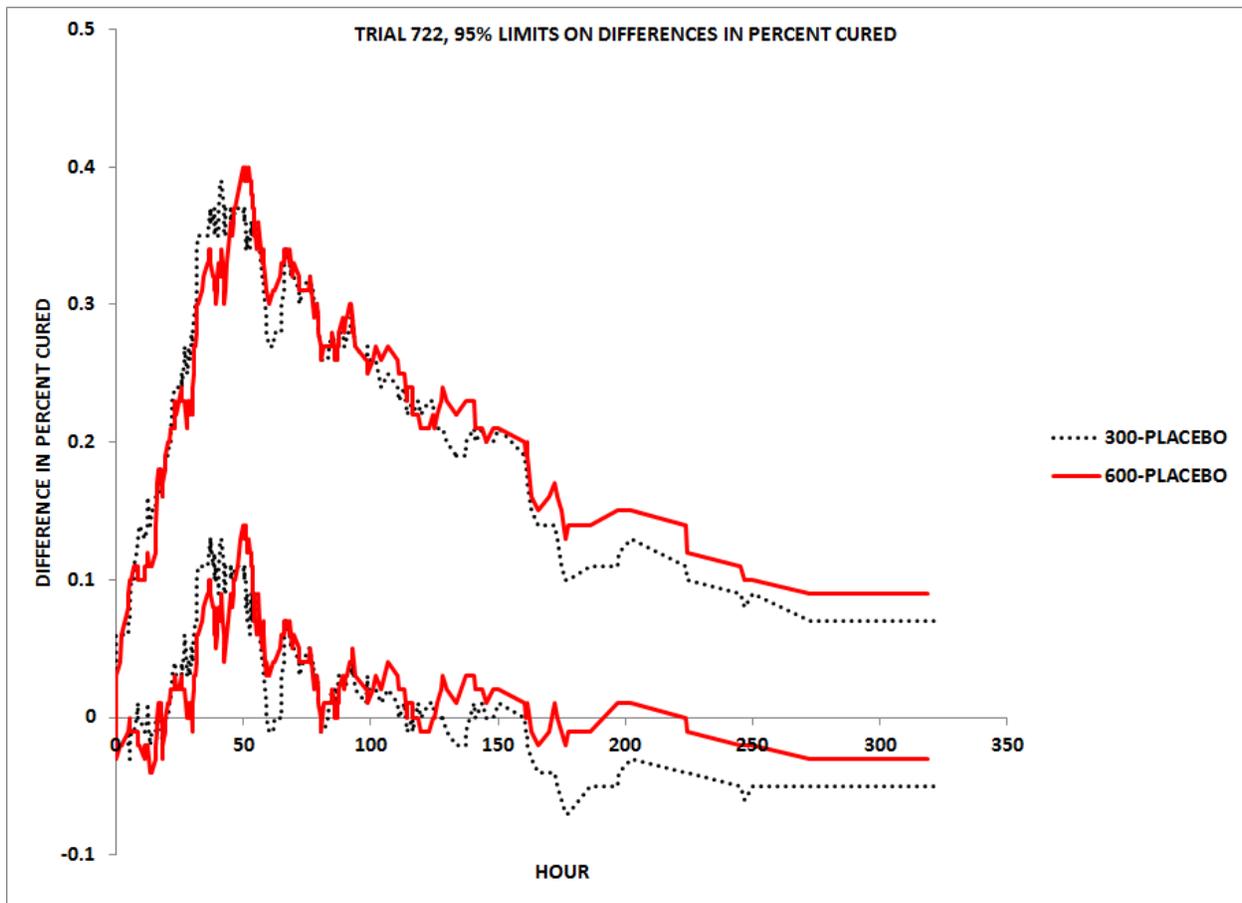
The following plots (figures 3.2.1 A-P) are the Kaplan-Meier curves for time to symptom alleviation, plotted for each trial for all the arms in that trial. The first trial is 722 (600 and 300 mg IV vs placebo).

Figure 3.2.1 A



One will notice that the placebo curve is visibly above the two peramivir curves (slower to heal than peramivir) throughout the middle range of times. The following two graphs (figures B and C) give the 95% confidence bands for the differences in percent healed on peramivir and placebo.

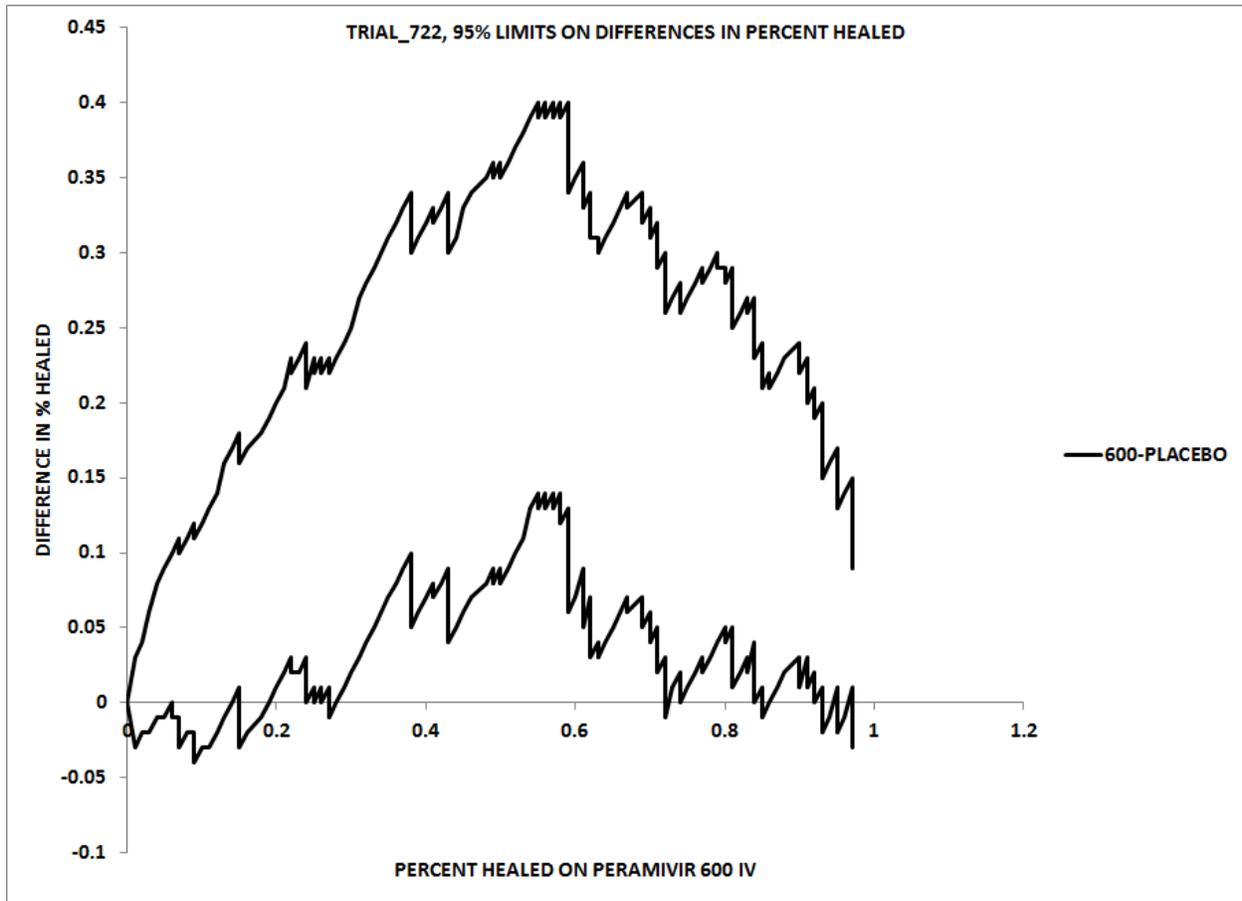
Figure 3.2.1 B



One can see that both peramivir doses are statistically significantly superior to placebo in percent healed from around 30 hours to around 150 hours. The estimated difference is not plotted here to keep the graph from getting too cluttered but one can see that the middle of the confidence band is around a 15%-20% superiority for peramivir in percent healed compared to placebo.

As mentioned above, everyone heals eventually so the Kaplan-Meier curves come back together at the end. Also hours to healing varies considerably from one flu season to the next. Thus, a more informative view of the confidence for the peramivir-placebo difference is the plot with percent healed on peramivir on the x-axis.

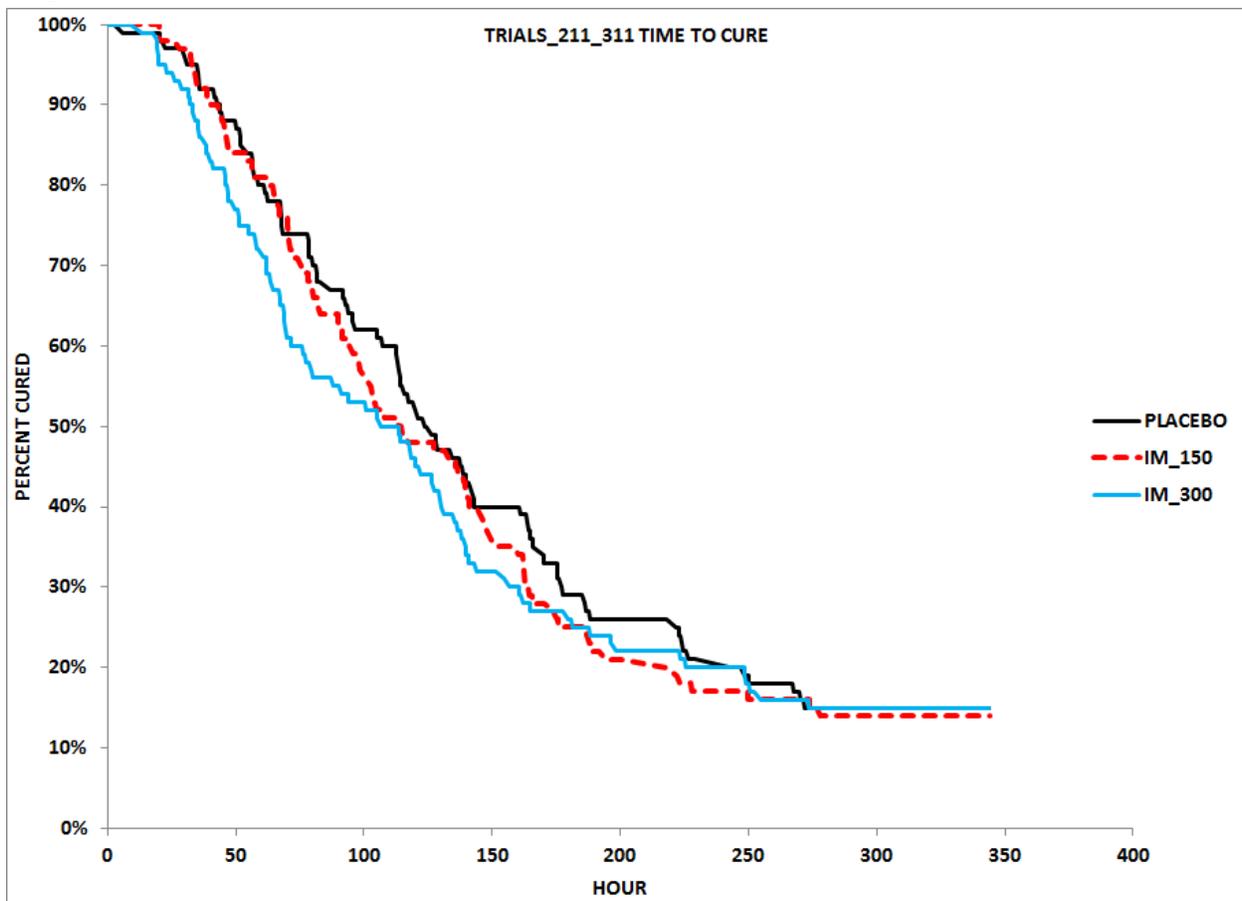
Figure 3.2.1 C



Here one can see that the 600 mg IV dose of peramivir is statistically significantly superior to placebo from the time 20% of peramivir subjects have healed until the time 80-85% of peramivir subjects have healed. One will have noticed that the results for 300 mg IV peramivir looked very similar to the results for the 600 mg peramivir. Therefore, the analogous graph for 300 mg -placebo difference is omitted here.

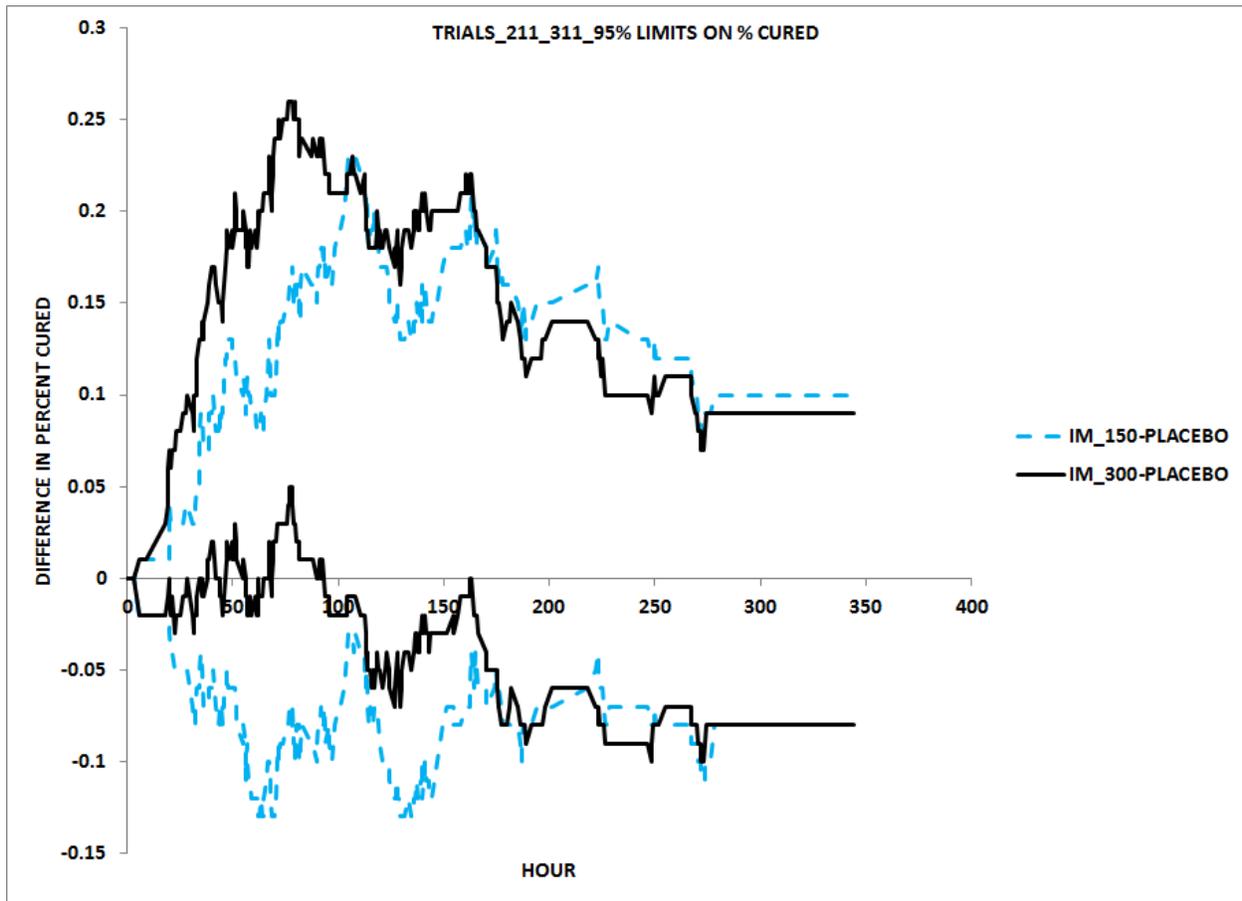
The next several graphs (figures D-F) describe the results for trials 211 and 311 pooled. These were two trials both much smaller than trials 722 and 815. Because they were small, occurred in successive flu seasons, and used similar peramivir regimens against control, the applicant has pooled them. The following results also pool these two trials. There are some statistical objections to this pooling: trial 211 randomized subjects to three arms; trial 311 to only two. This means in the pooled analysis, the 150mg dose from trial 211 would be compared to a mixture of randomized placebo subjects from trial 211 and effectively non-randomized placebo subjects from trial 311. Also the stratifying variables were different in the two trials.

Figure 3.2.1 D



In addition to the small sample size, these trials also used lower doses of peramivir than did the pivotal trials 722 and 815. For these two reasons, one may not expect even the pooled trial to show statistically significant superiority to placebo. As this graph shows, the point estimates for percent healed go in the right direction with placebo slowest healing and at least a suggestion that 150 mg IM peramivir produces slower healing than 300 mg IM peramivir.

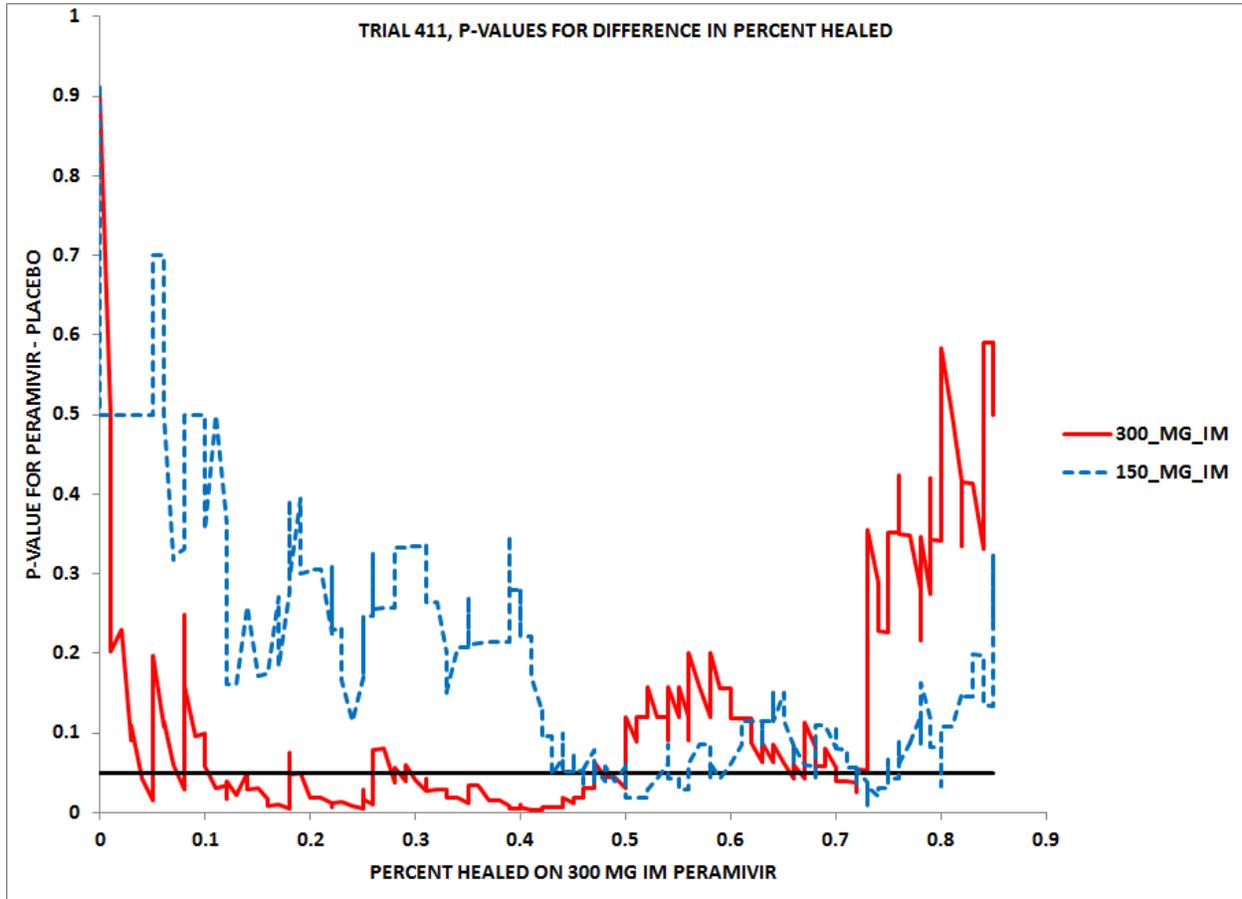
The following graph gives the 95% confidence bands for the difference between IM peramivir and placebo. Figure 3.2.1 E



One will not be surprised that neither the 150mg nor the 300 mg dose of peramivir shows statistically significant superiority over placebo. Credibly, these lower doses of peramivir are between 15-25% better, and 5-10% worse, than placebo.

The following graph gives the p-value for the difference in percent healed between each peramivir dose and placebo, plotted against the percent healed on the higher peramivir dose.

Figure 3.2.1 F

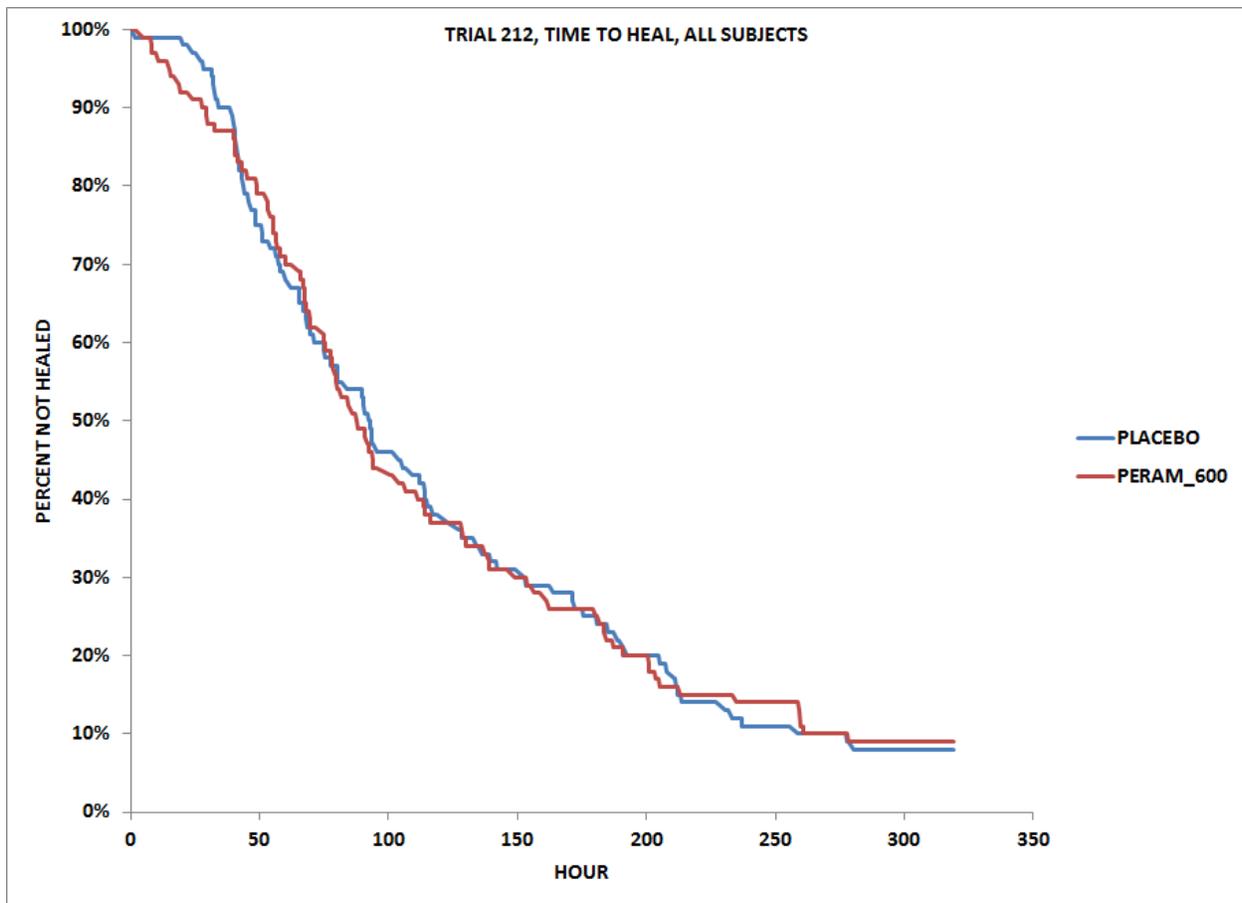


One can see that peramivir 300 mg is actually statistically significant, or nearly so (p-values usually $<.05$, always $<.10$) from the time when 10% of patients were healed on 300 mg peramivir until about 45% of patients were healed. This should count as moderately supportive evidence for the efficacy of peramivir.

The next several graphs (figures G-P) describe the results from trial 212. There are three issues that arise with this trial. The most important is that this trial, like trial 815, was conducted in the presence of a Tamiflu resistant strain of influenza A. As shown in table 2.2.3 F above, 69% of all subjects, 230 in all, had the Tamiflu resistant H275Y substitution. Second, for this trial only, the applicant planned to use only the type A influenza cases. This excludes another 16% of all subjects (52 in all). This will require a comparison of the results for type A only with the results for all types. Third, this trial, more so than the others, shows a greater discrepancy between the results using times to healing used by the applicant, which are contaminated by mistakes in the proof-reading of the times of diary entries, and the results using the times as corrected by the FDA reviewer.

The first two graphs (figures G-H) show Kaplan-Meier curves for the two arms for time to healing, first with all types of influenza and second with only type A influenza.

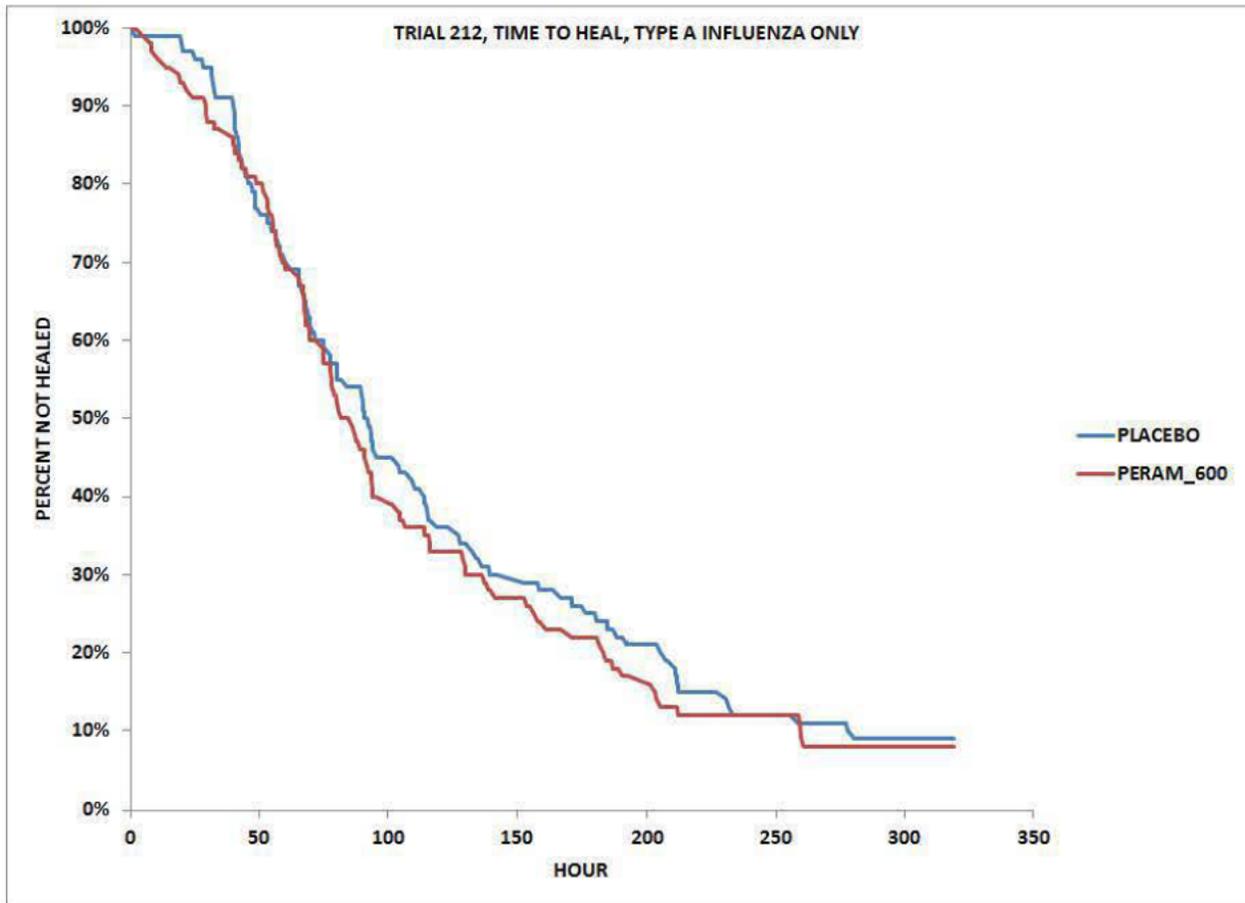
Figure 3.2.1 G



One can see that in this trial, in contrast to the others, there is no visible difference between peramivir and placebo when all types of influenza are considered. This supports the contention that peramivir is not effective against all types and sub-types of influenza. Over the course of the next several graphs, this review will show that type B and Tamiflu resistant strains of type A are likely the cases responsible for the lack of efficacy seen here.

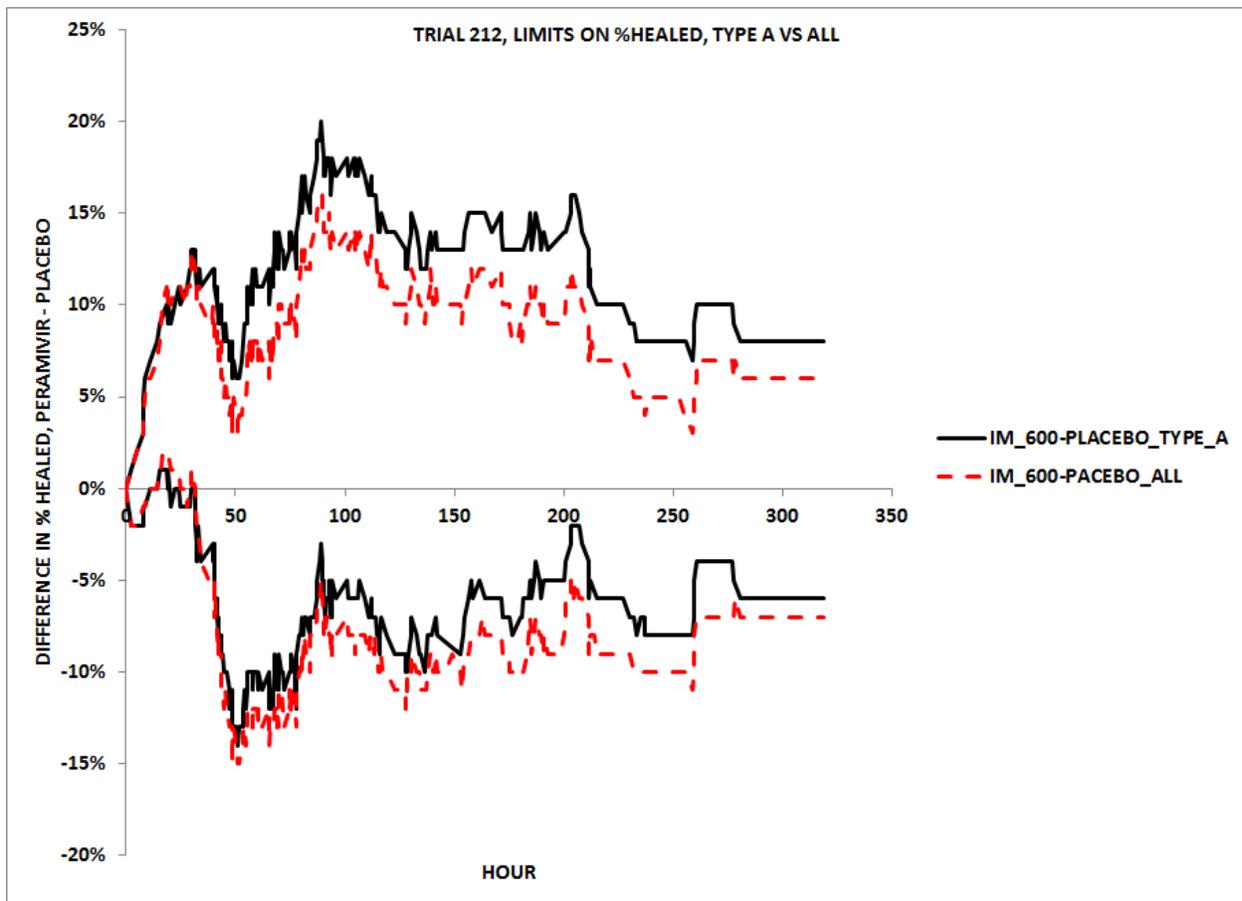
With type A influenza cases only, the following graph shows slightly more separation.

Figure 3.2.1 H



The difference between type A and all types is a little more apparent in the following graph (figure I), which gives the 95% confidence bands for the difference in percent healed, peramivir - placebo. One can see that the black solid bands, using type A data only, are shifted upward (toward better results with peramivir) relative to the dotted red bands, which use data from all types.

Figure 3.2.1 I



The applicant's results look slightly better when considering type A cases only but are close to the FDA results when all types are included. This is illustrated by the following Kaplan-Meier plots (figures J-K) for type A cases and all cases, using the applicant's timepoints.

Figure 3.2.1 J

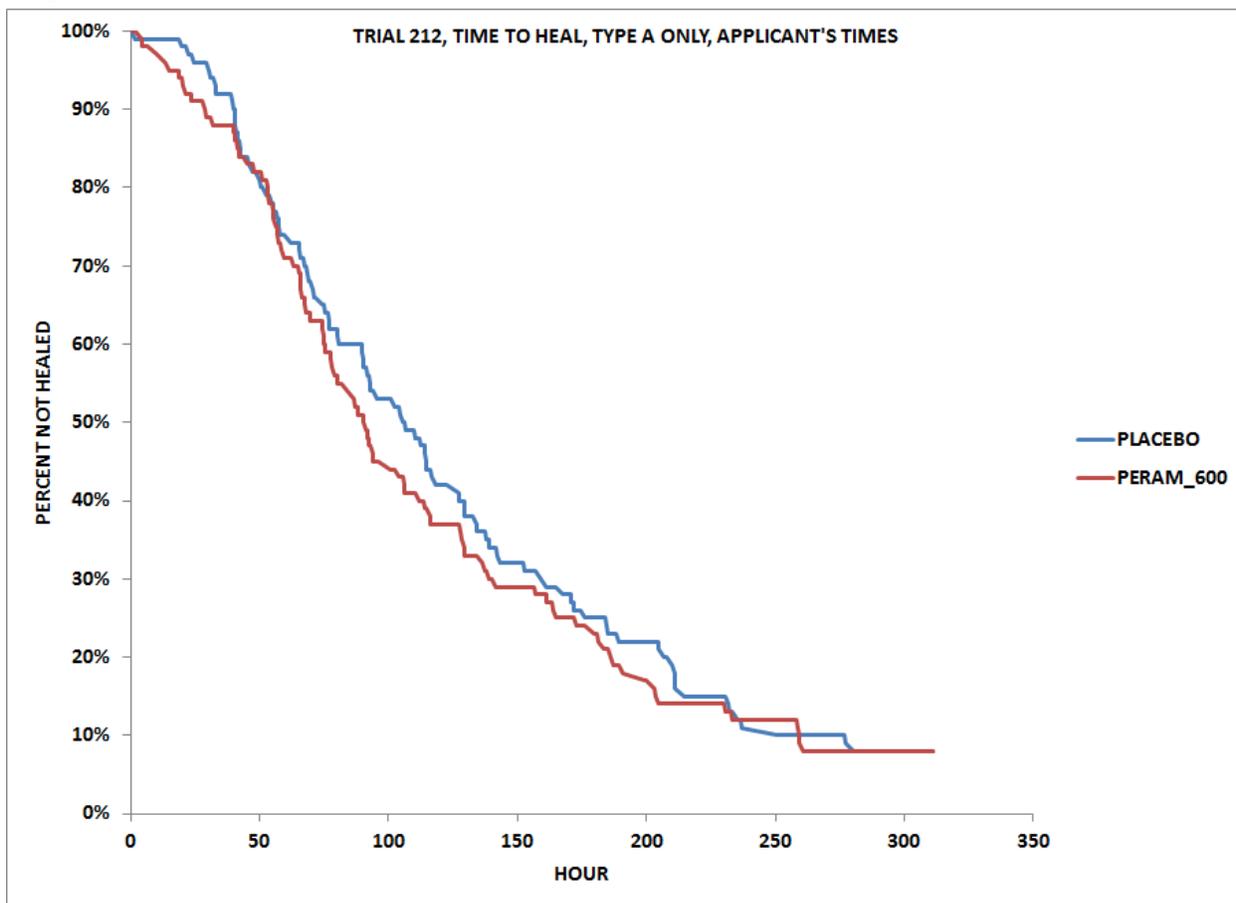
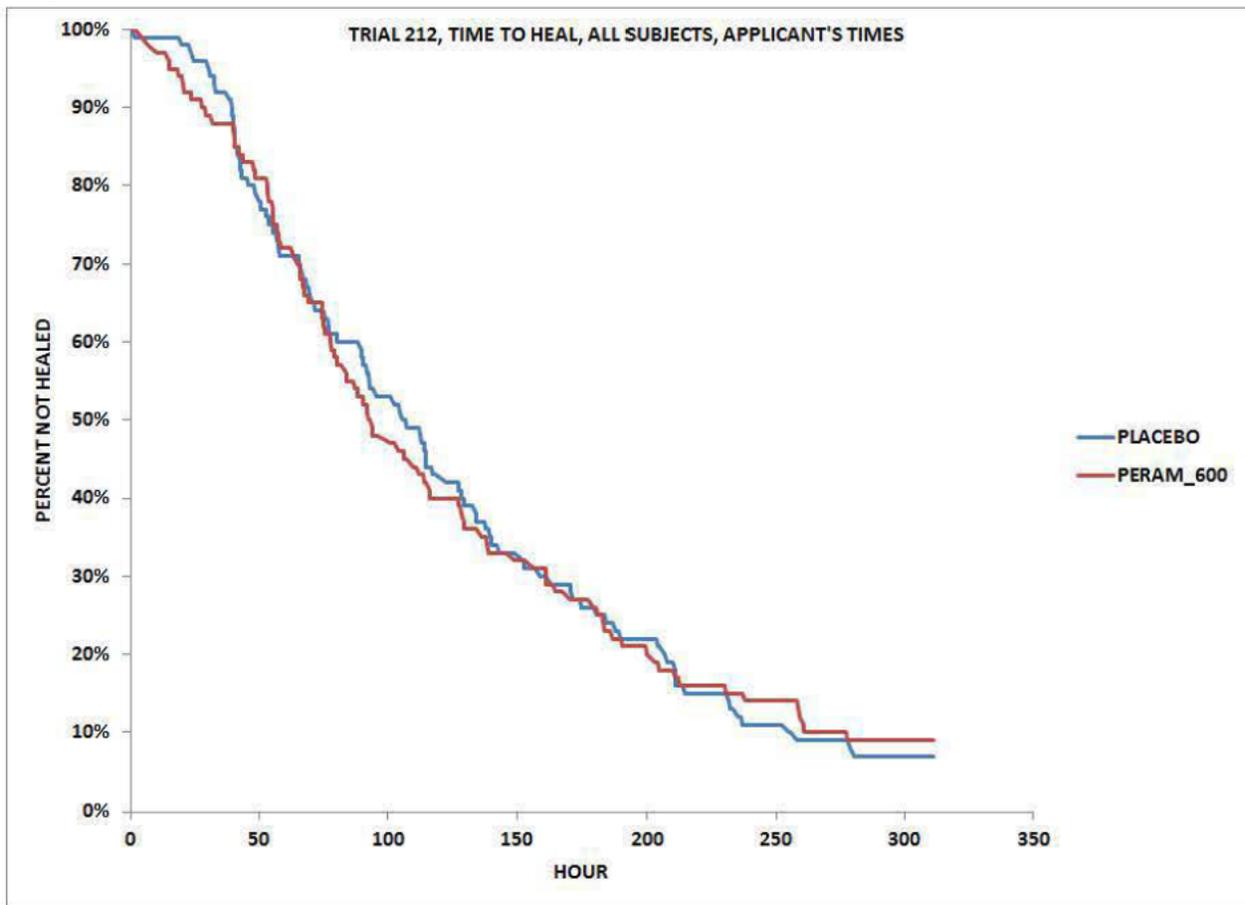
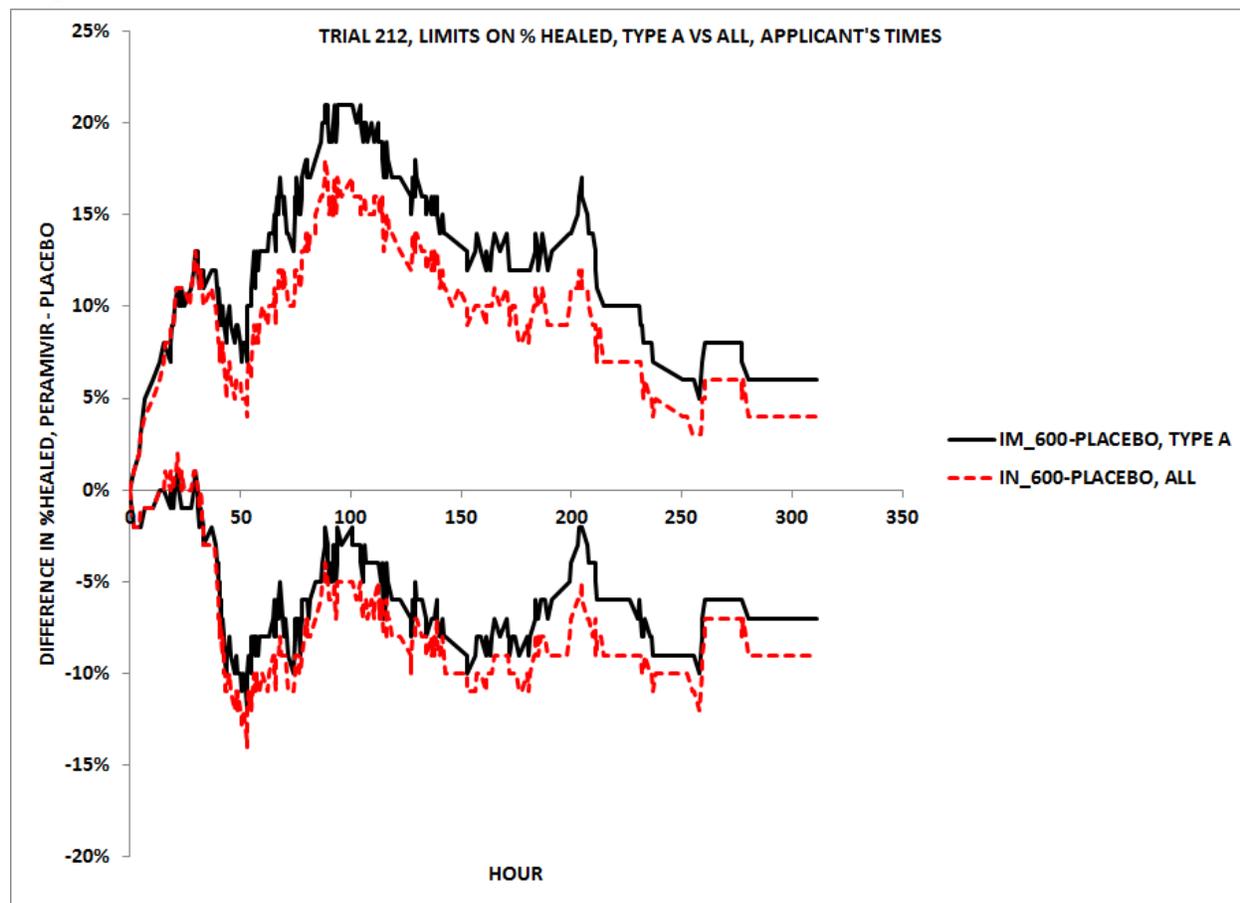


Figure 3.2.1 K



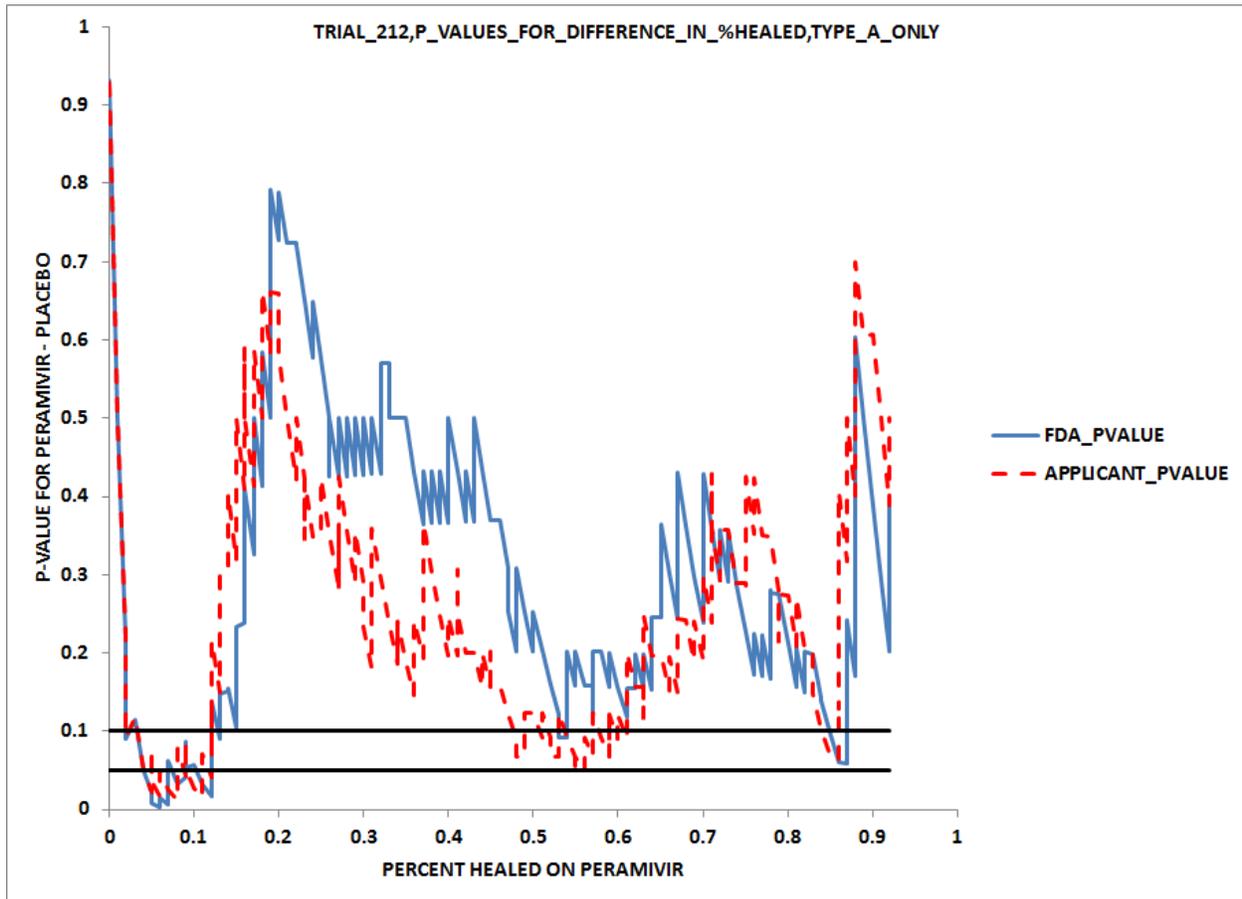
As was the case with the FDA's times to healing, one can see the difference between type A cases and all cases more clearly by looking at the 95% confidence bands for the difference in percent healed between peramivir and placebo, shown in the following graph.

Figure 3.2.1 L



As was the case for the FDA's results, the solid black curves, computed using only type A cases, are shifted visibly upward from the dotted red curves, computed using all cases. If one compares this graph with the corresponding one above for the FDA's corrected times, one will also notice that even the red curves for all cases are slightly favorable for peramivir in the applicant's analysis but not in the FDA's analysis.

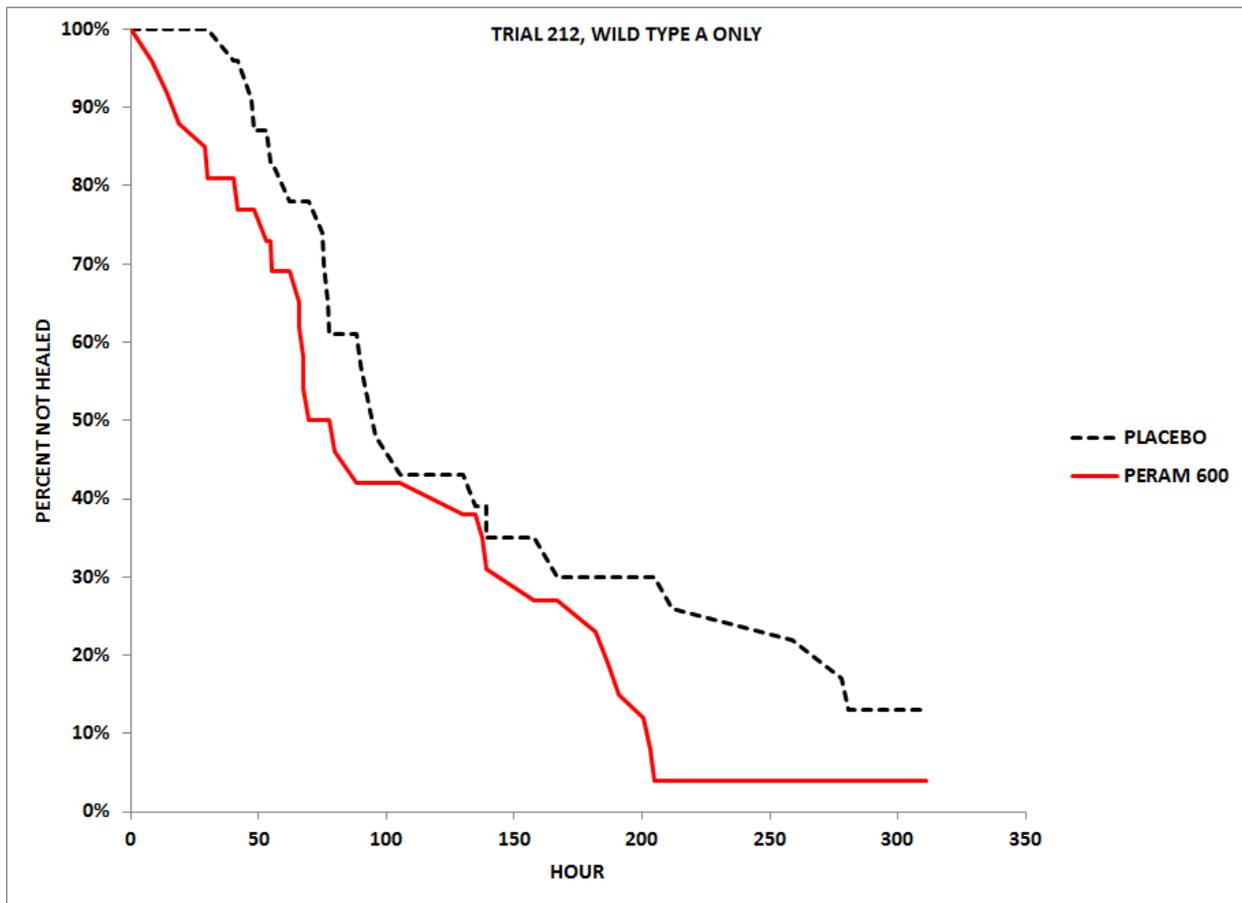
As a complementary way of assessing the data in trial 212, the following graph presents the p-values for testing the superiority of peramivir_600_mg_IM to placebo at each time. In this, it is more informative to plot these p-values against the percentage of patients healed rather than against time. The following graph shows this plot for type A cases. Figure 3.2.1 M



According to the applicant, the p-value was below .10 (but above .05 near the median time to healing (from about 45% healed to about 60% healed and was about .2 to .3 from about 30% healed to about 70% healed). The FDA calculates these p-values to be between .1 and .2 near the median and between .3 to .4 from 30% healed to 70% healed. P-values with all cases included would be much less impressive and are not plotted.

This anomalous result from trial 212 is mainly due to the widespread prevalence of the Tamiflu resistant substitution H275Y. There are a total of 15 subjects with type A H1N1 influenza without that substitution and another 36 with type A H3N2. The FDA reviewer computed Kaplan-Meier curves and Cox regression estimates on these 51 subjects (23 on placebo and 26 on peramivir after deleting two subjects treated by Wise). The results are given below in figures N-P.

Figure 3.2.1 N



(Technically, this isn't quite a Kaplan-Meier plot; each step in the Kaplan-Meier plot is graphed as a ramp here.) This plot shows noticeably wider separation between the arms than do any of the plots above using either all subjects or all type A subjects.

The 95% differences in the percent healed are given in the following plots, the first with hour on the x-axis (figure o) and the second with percent healed on peramivir on the x-axis (figure P).

Figure 3.2.1 O

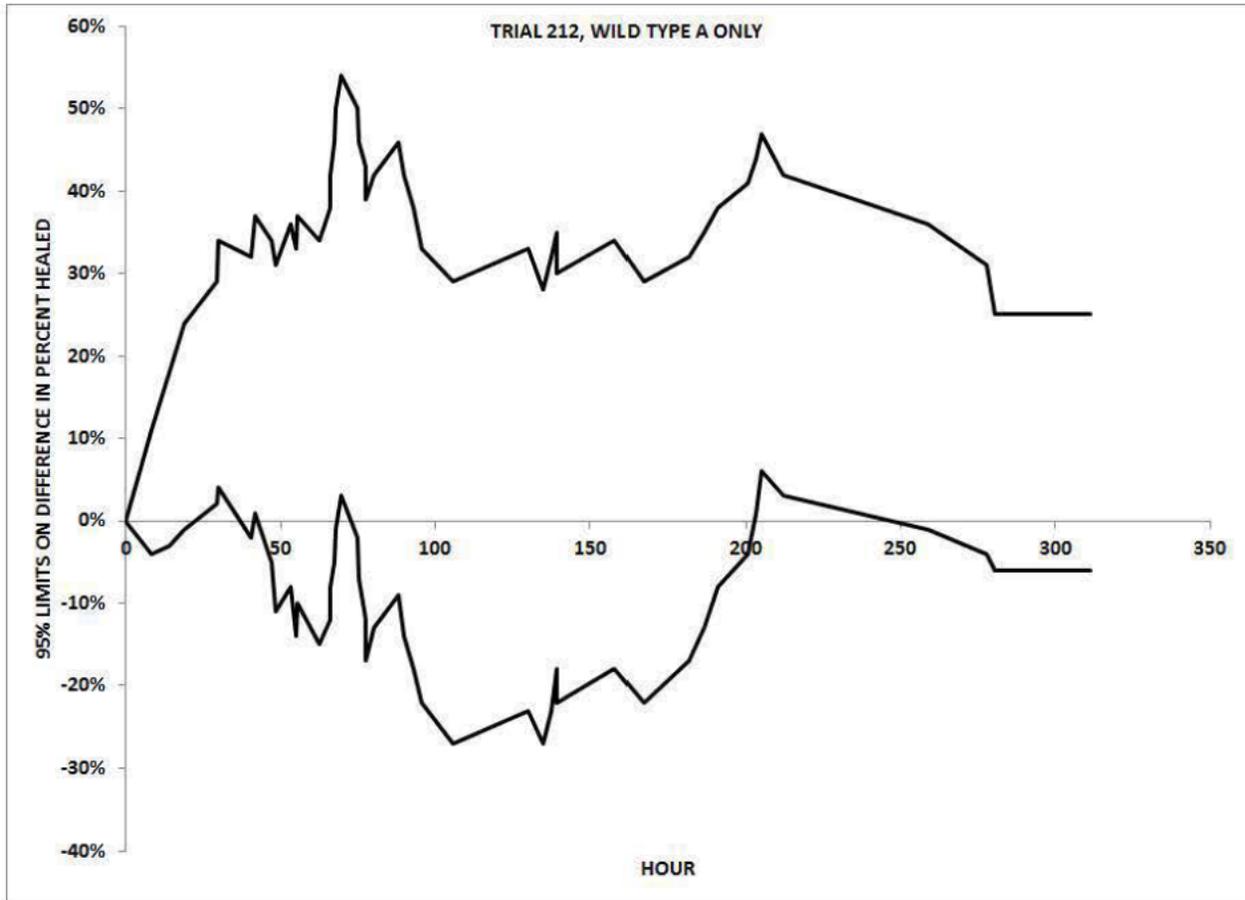
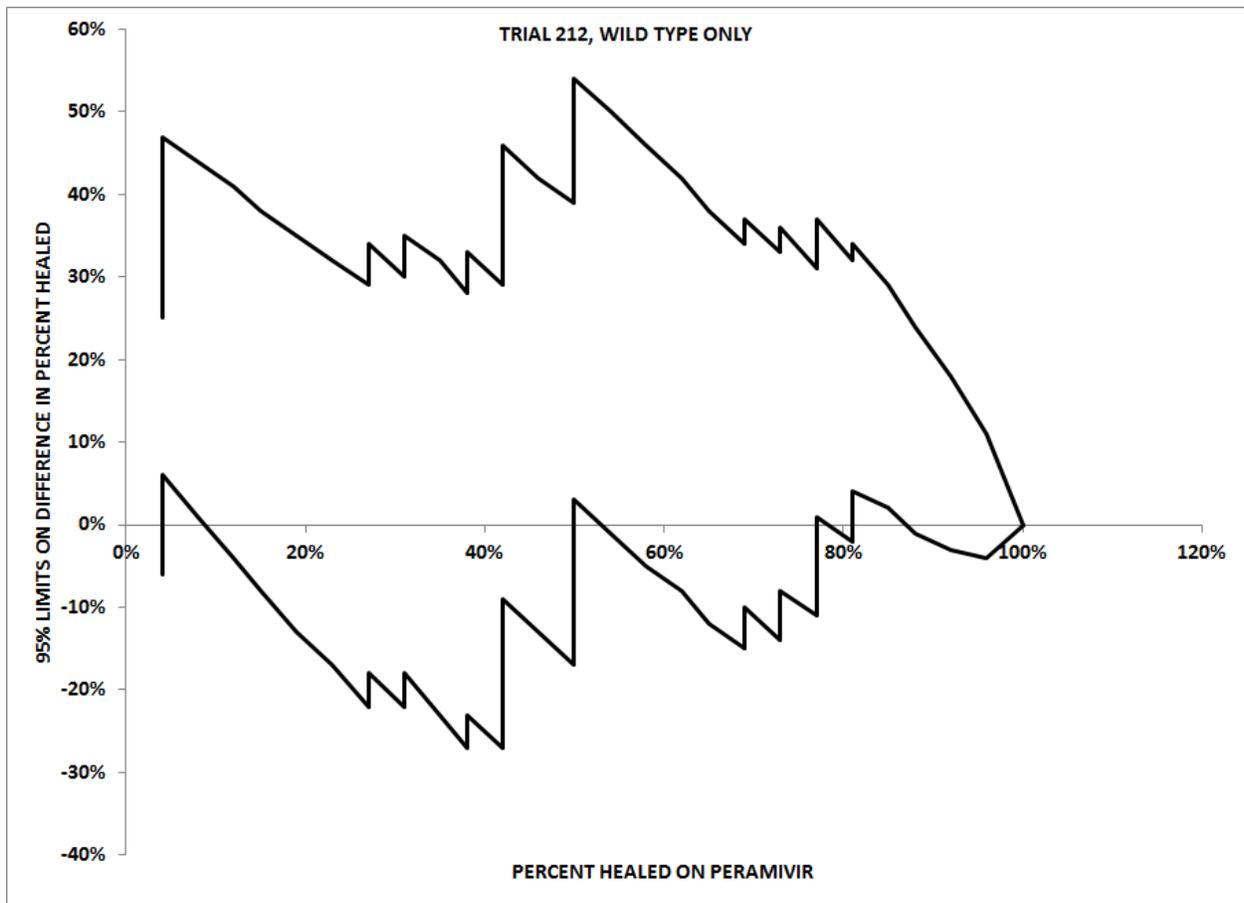


Figure 3.2.1 P



Considering the small sample size, these plots come close to achieving statistical significance.

3.2.2 Cox Regressions of Time to Symptom Alleviation

The results from Cox regressions of time to healing on treatment are described in the graph (figures 3.2.2 A) below. There are some points to make first. The Cox proportional hazard regression may be performed either unstratified or stratified. Since the trials in this NDA all used stratified randomization, the stratified analysis is more statistically reliable. To review, trial 722 is stratified by current smoking status and baseline sum of symptom scores (≤ 14), trial 311 by current smoking status, and type of influenza, trial 211 by current smoking status, and trial 212 by current smoking status and type of influenza.

Trial 815 was not exactly stratified but was balanced with respect to current smoking status, baseline total symptom score, type of influenza, and country by dynamic allocation. The correct analysis of this type of randomization can be approximated by stratifying on the same four variables. Unfortunately, that yields strata too small to get valid estimates. After exploring the strata sizes using only two or three of the baseline covariate, the FDA statistical reviewer elected to analyze this trial stratifying on current smoking status and baseline total symptom score.

An additional problem with trial 815 is that the hazard ratios are computed for peramivir to Tamiflu. The FDA reviewer obtained the hazard ratios for peramivir to placebo by multiplying the hazard ratio peramivir/Tamiflu by the hazard ratio Tamiflu/placebo from the Tamiflu NDA. (The standard errors needed to get confidence intervals were obtained by adding the standard errors for log hazard ratios from the same two sources.) Details of this computation will be discussed in section 3.2.3.

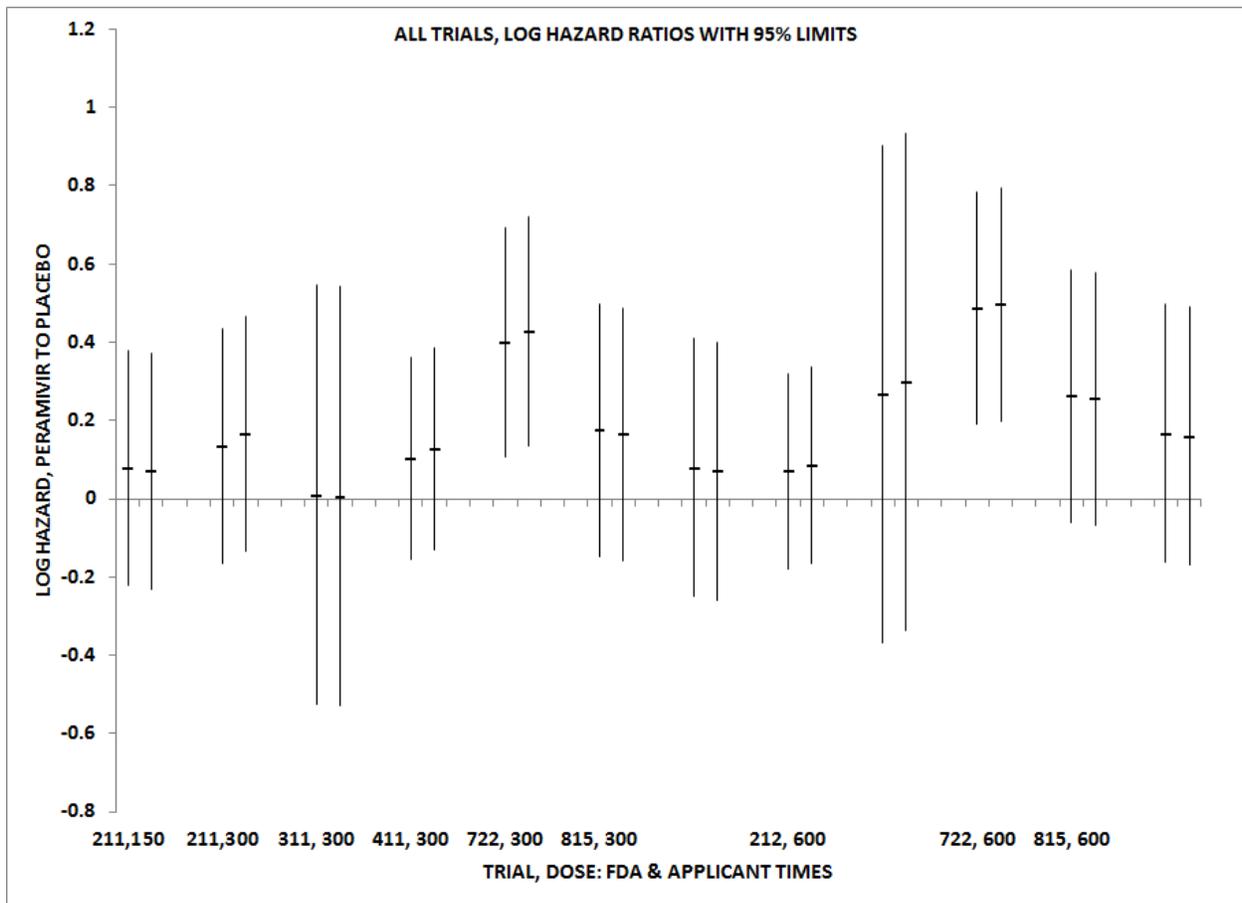
The graph (figure 3.2.2 A) gives the 95% confidence intervals for the log hazard ratio of each peramivir arm to the matching placebo arm. (This means the peramivir/Tamiflu*Tamiflu/placebo ratios in trial 815.) Because the results are given on the log scale, values greater than zero indicate peramivir better than placebo.

In this graph, the log hazard ratio relative to placebo and 95% confidence interval was estimated twice for each trial and peramivir arm. The first (left) estimate uses the FDA-protocol computation, using the first of two consecutive symptom free diary entries; the second (right) estimate uses applicant's computation, which took the first such visit after all visits with any moderate or severe symptoms. (I.e. the first estimates ignores relapses and the second counts healing only after all relapses). A stratified analysis was used for each endpoint.

In the graph (figure 3.2.2 A), summarizing all doses in all trials, the trials and arms are arranged so that the dose increases from 150mg on the left to 600mg on the right. Trial 411, dose 300, is just the pooled results of trials 211 and 311. Trials 815, doses 300 and 600, and trial 212, dose 600, are represented two pairs of lines, not just one. For trial 212, the left pair used all influenza types; the right pair used only the 49 subjects with type A H1N1 Wild type or type A H3N2. (The label 212, 600 on the horizontal axis is under the left pair; there is no label under the right pair.) One will notice that the H3N2 + Wild Type H1N1 gives a higher point estimate but, of course, a wider confidence interval.

For the two doses of trial 815 only subjects with type A, H3N2 influenza were used. The left pair (which has the label under it) of estimates were computed simply adding the point estimate and variance of log hazard ratio, Tamiflu to placebo, from the Tamiflu NDA to the point estimate and variances of the log hazard ratio, peramivir to Tamiflu from trial 815, H3N2 only. The right pair (just to the right of the left pair and without the label) added $.9 * \text{point estimate}$ and $1.21 * \text{variance of log hazard ratio, Tamiflu to placebo}$. The 10% decrease in point estimate and the 10% increase in standard error ($1.21 * \text{variance} = \text{square of } 1.1 * \text{standard error}$) are sensitivity adjustments for potential inter-trial variability. See section 3.2.3 below for further discussion.

Figure 3.2.2 A



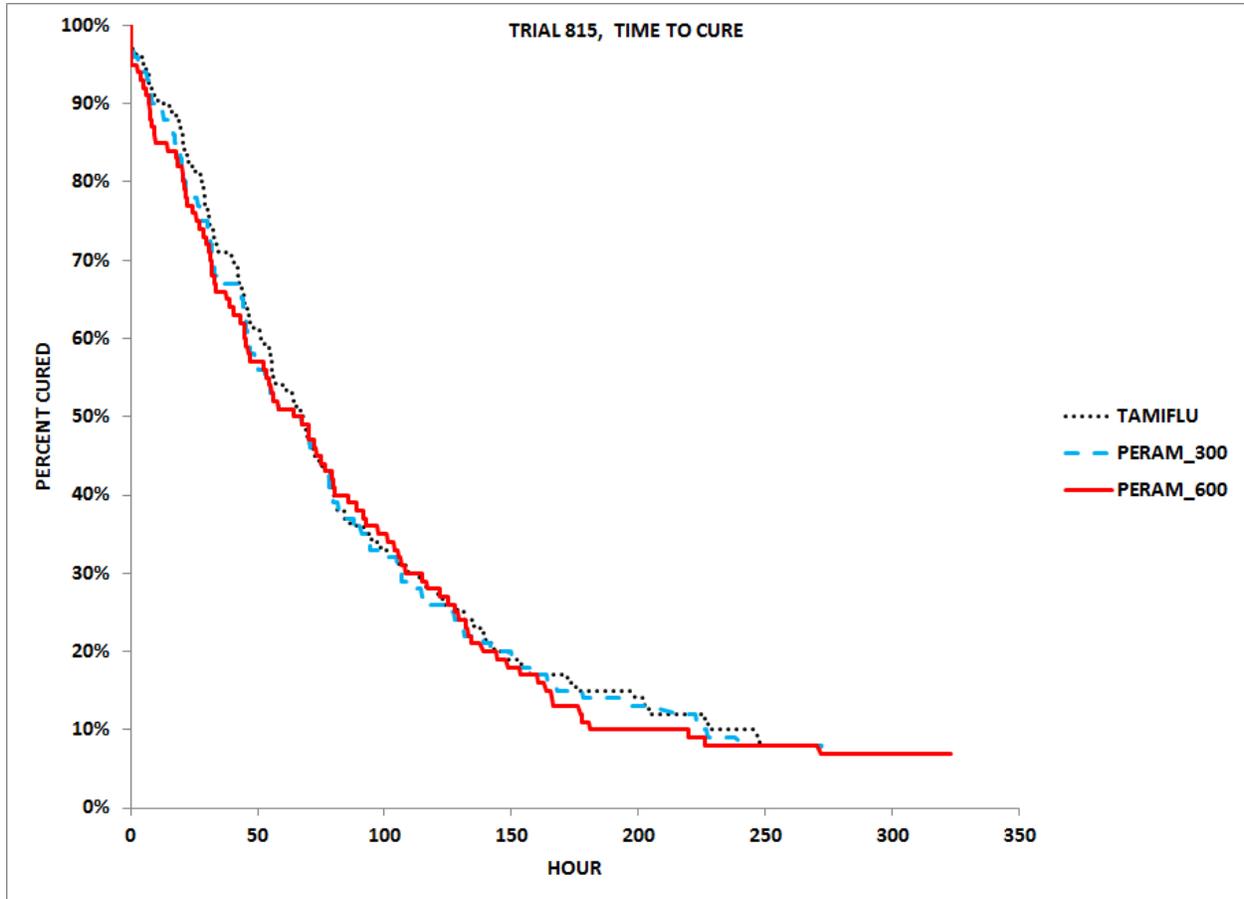
One will notice two things. First, trial 212, when all types are used, is noticeably anomalous and trial 311 is slightly anomalous. In all the other cases (dose 150 in trial 211, dose 300 in 722, 815 and 411=211+311 pooled, dose 600 in 722, 815, and type A H3N2+Wild H1N1 cases of trial 212) the point estimates are positive and at least the upper three fourths of the confidence interval lie above zero. The upper three fourths of the confidence interval above zero is about equivalent to a p-value of .16 or so.

Finally, the most important observation is that the large placebo controlled pivotal trial (722) with 600mg IV peramivir is quite statistically significantly superior to placebo. (Even the 95% lower bound is well above zero.) In fact, the 300mg IV dose is also statistically significantly superior in this trial.

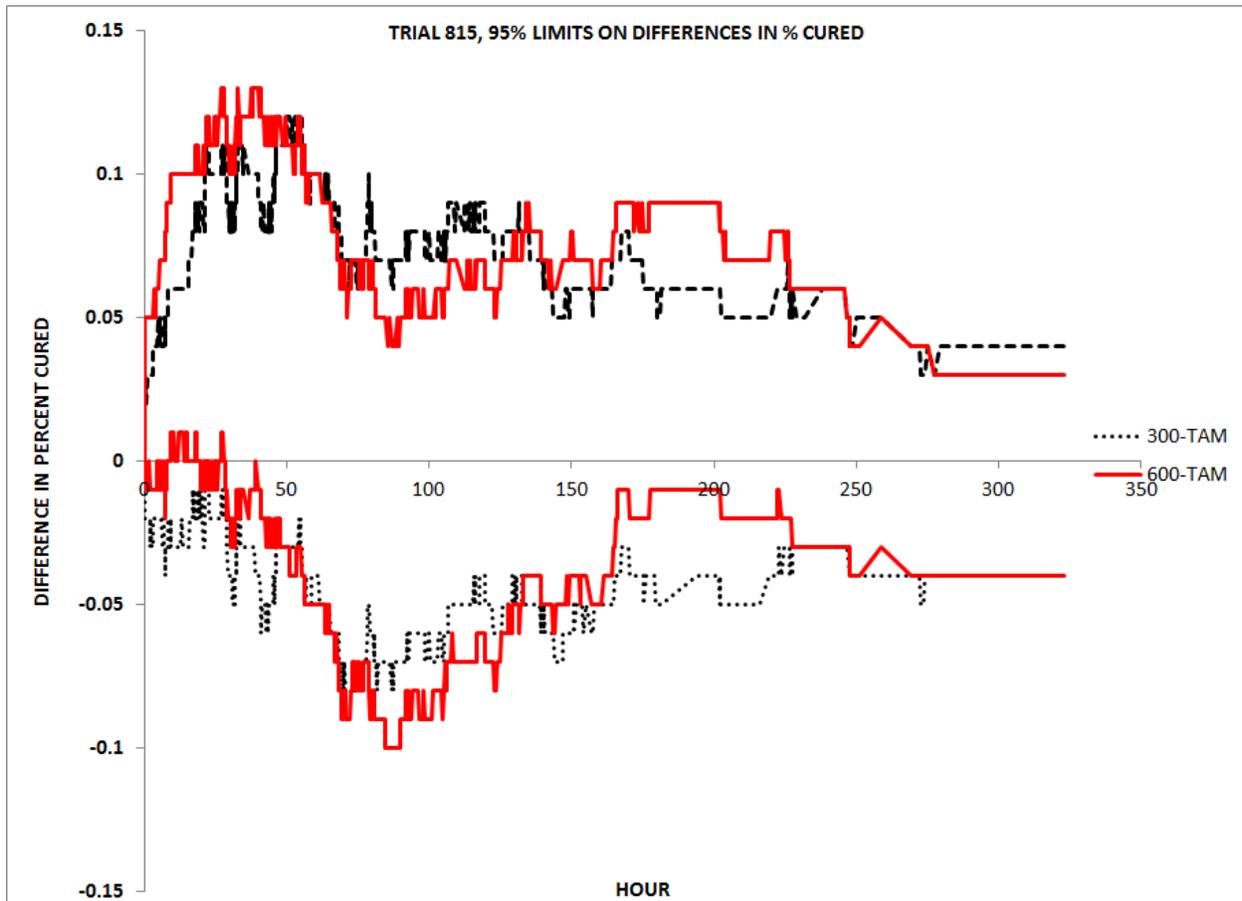
3.2.3 Estimation of Hazard Ratio of Peramivir to Placebo from Trial 815 and Tamiflu NDA

The next graphs (figures 3.2.3 A-B) describe trial 815. This is the only trial without a placebo control. The Kaplan-Meier plots of Tamiflu and the two peramivir doses in trial 815 are as follows.

Figure 3.2.3 A



The following graph gives the 95% confidence bands for the differences in percent healed between peramivir and Tamiflu. Figure 3.2.3 B



One can see that both doses of peramivir, with high confidence are somewhere between 10% (very briefly, 12%) better, and 5% (very briefly, 10%) worse, than Tamiflu.

Since Tamiflu is an effective drug, the fact that Tamiflu and both doses of IV peramivir nearly coincide would be encouraging, were it not for the articles cited above as evidencing the wide-spread presence of Tamiflu resistant virus in the population in which this study was conducted. As is, this data may only demonstrate that peramivir doesn't work in Tamiflu-resistant influenza.

This trial does not contain data showing presence or absence of the Tamiflu resistant H275Y substitution. However, one may use only the 329 H3N2 subtype subjects to get only Tamiflu susceptible influenza cases. There is data on the IC50's for peramivir and Tamiflu for each subject. The following table and graph shows some salient features of these data. Table 3.2.3 A gives counts of IC50 in equally spaced intervals (on the log scale).

TABLE 3.2.3 A
HISTOGRAM OF IC50 PERAMIVIR, TRIAL 815

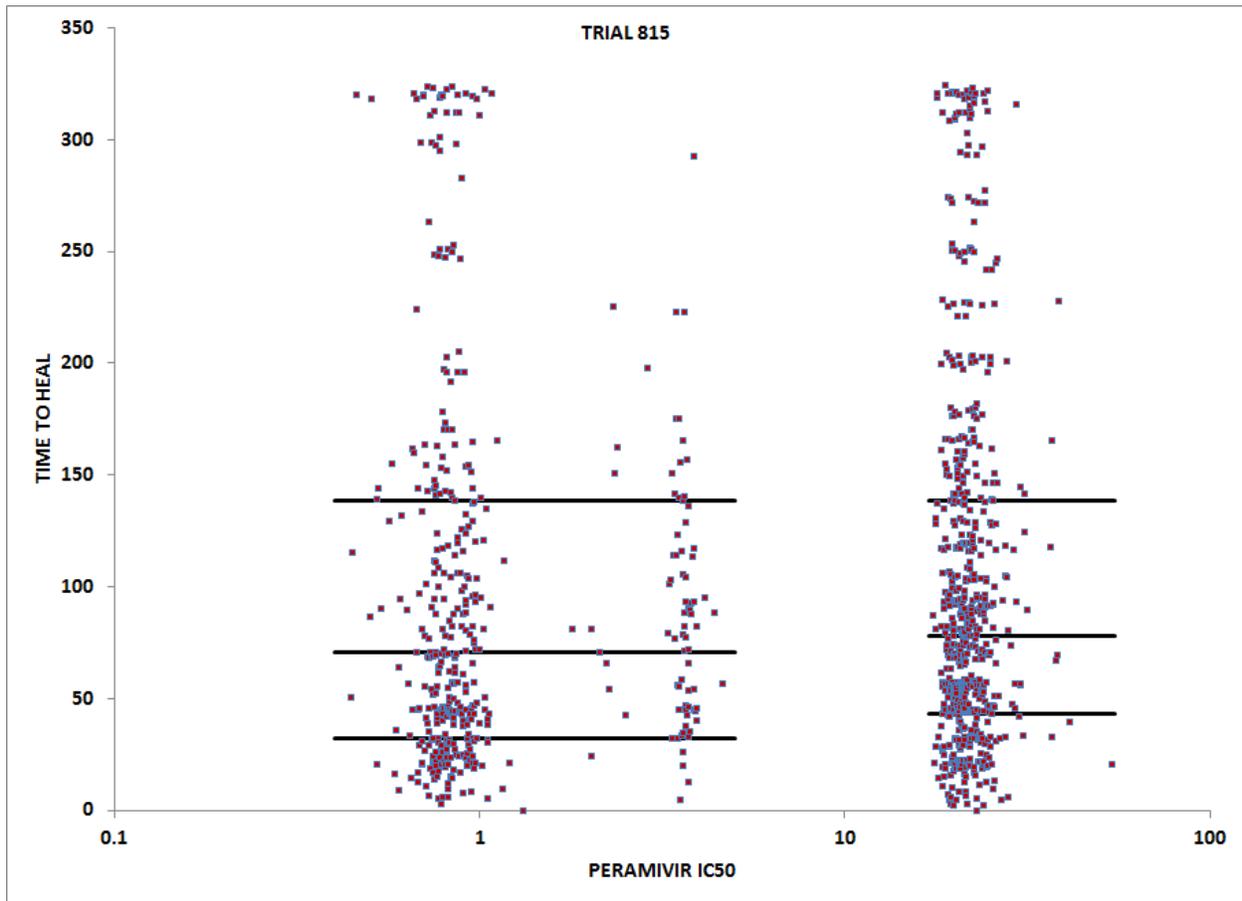
IC50	COUNT	IC50	COUNT	IC50	COUNT
0.45	0	2.33	6	12.16	0
0.53	7	2.75	3	14.34	0
0.62	10	3.24	1	16.92	0
0.73	46	3.82	57	19.96	136
0.86	163	4.51	12	23.55	359
1.02	101	5.32	1	27.78	113
1.20	17	6.28	0	32.77	18
1.42	2	7.40	0	38.66	5
1.67	0	8.74	0	45.61	2
1.97	1	10.31	0	53.80	1

One will notice that the IC50's form a trimodal distribution with two major peaks around .8 and 23 and a minor peak around 3.8. There are no subjects with IC50 between 5 and 15. All but one of the type A subjects with subtype H3N2 had IC50<5; all but one of those with subtype H1N1 had IC50>15.

TABLE 3.2.3 B
TRIAL 815, IC50 BY SUBTYPE

SUBTYPE	NUMBER WITH	
	IC50<5	IC50>15
B	70	0
Indeterminate	31	39
A-H1N1, Wild Type	1	593
A-H3N2	324	1

This graph is a plot of time to heal versus the IC50 for peramivir (on a log scale).
 Figure 3.2.3 C



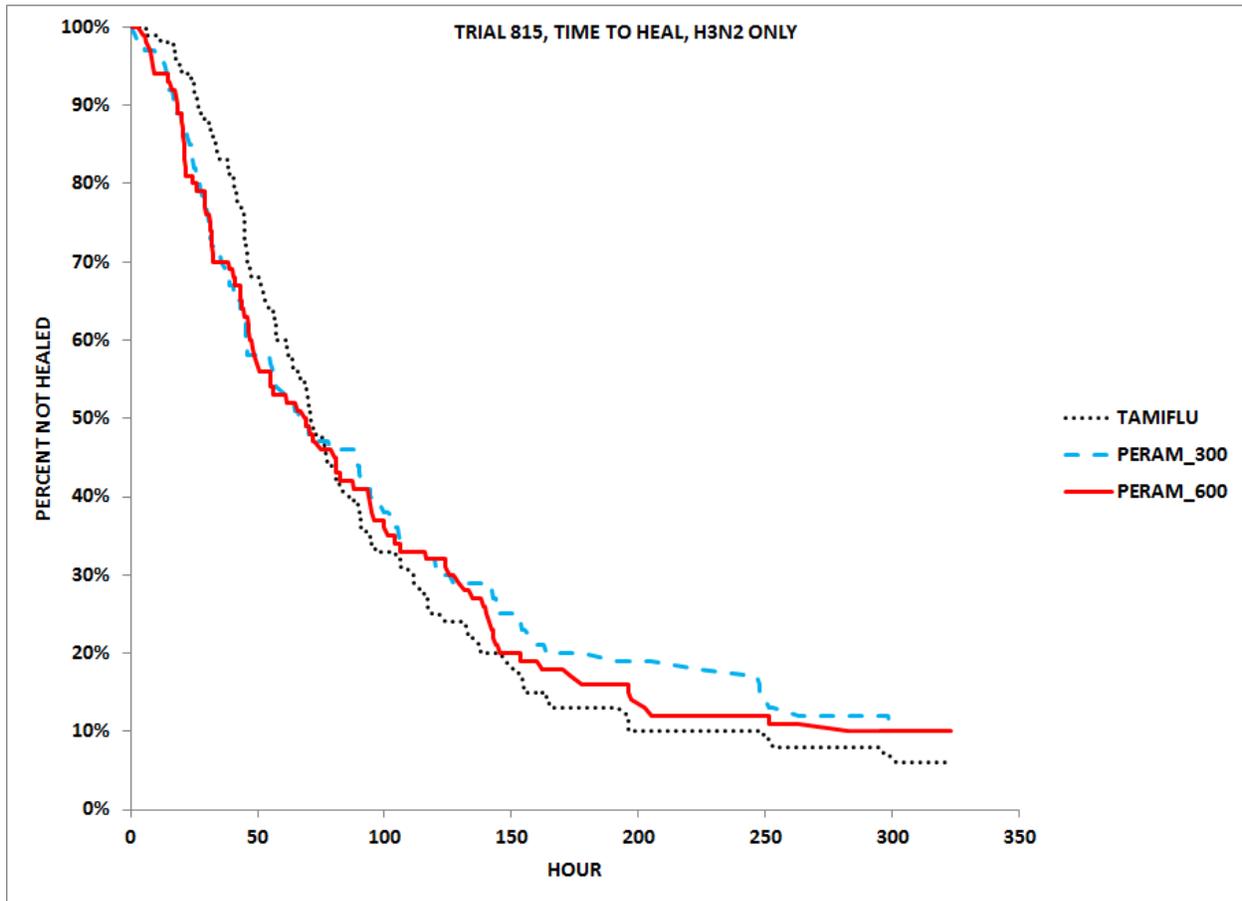
The trimodal shape of the IC50 is visible here too. It is a little harder to see the association between IC50 and time to healing. The horizontal black lines are the medians and quartiles of the times to healing for subjects with IC50<5 and with IC50>15. There are 7-11 hour reductions in 1st quartile and median of times to healing for subjects with IC50<5 compared to subjects with IC50>15. The third quartiles are about the same in both groups.

TABLE 3.2.3 C
 TIME TO HEALING, TRIAL 815

	1ST QUARTILE	MEDIAN	3RD QUARTILE
IC50<5	32.3 hrs	70.5	138.5
IC50>15	43.2 hrs	77.8	138.6

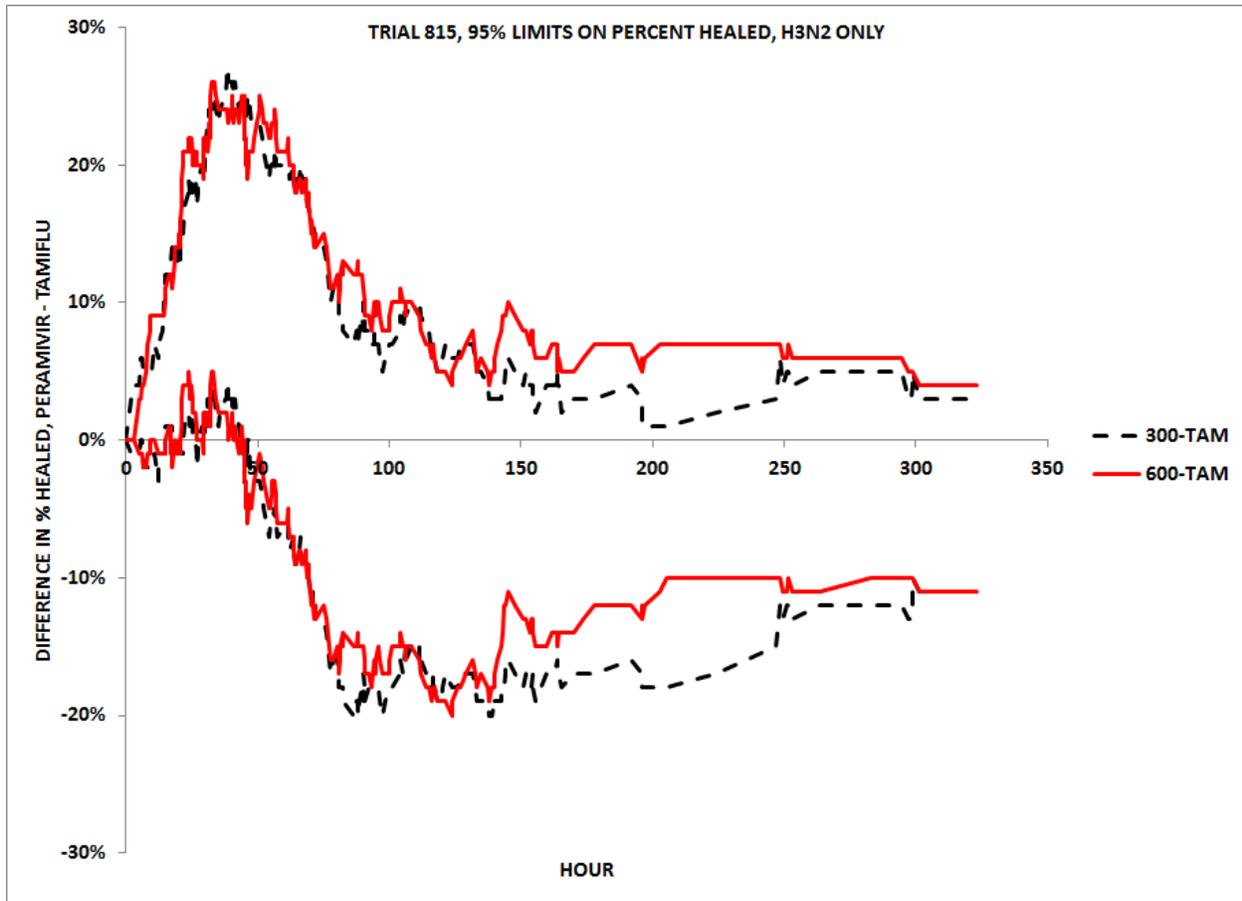
The Kaplan-Meier curves for time to healing and the 95% confidence bands for the difference between peramivir and Tamiflu in H3N2 subjects (almost the subjects with $IC_{50} < 5$) are given below in figures D-F.

Figure 3.2.3 D



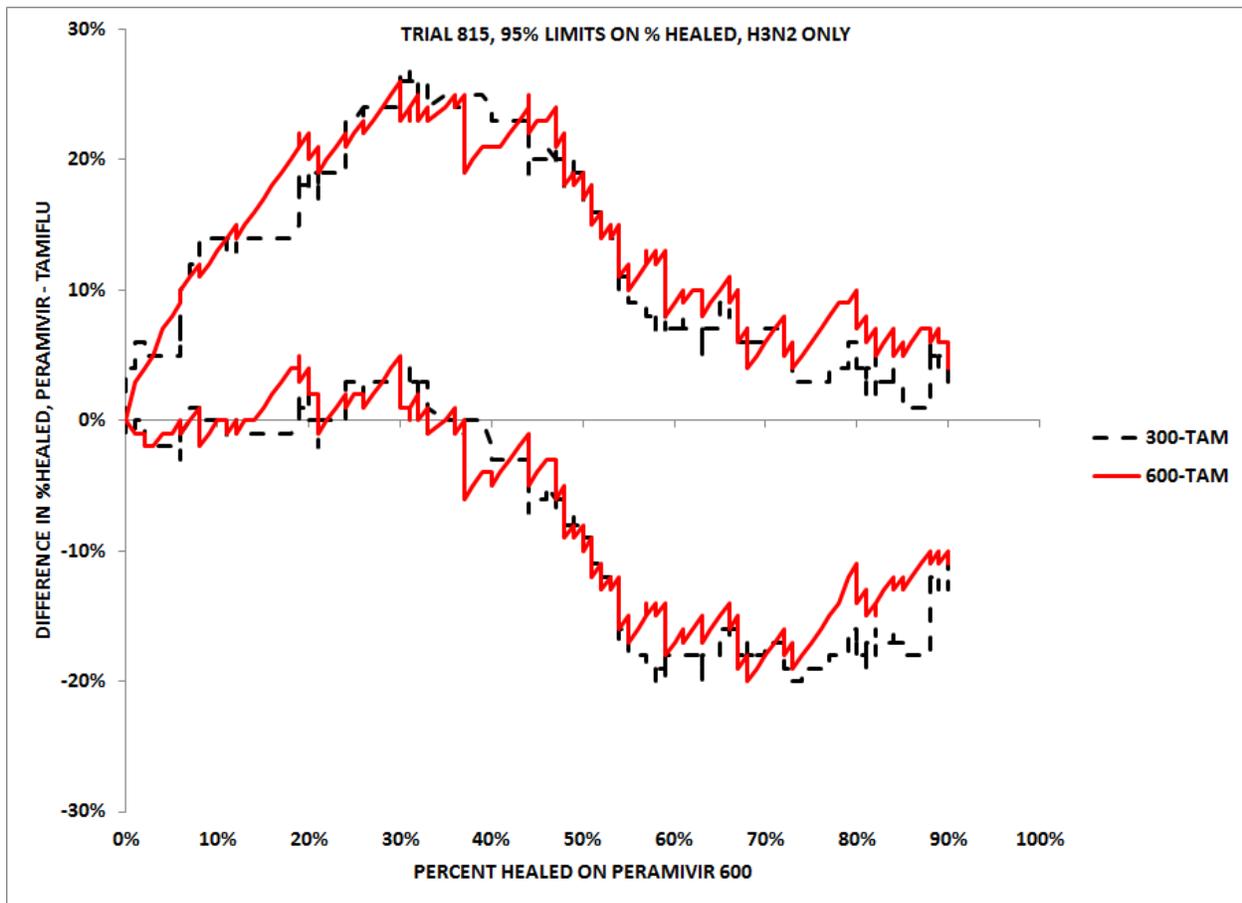
One will notice there is somewhat greater separation than in the previous graph (figure 3.2.3 A) which used all 1093 subjects. Also the Tamiflu arm shows slower healing for the first half of subjects but faster healing for the 50% of subjects healing more slowly.

Figure 3.2.3 E



In the plot of the confidence bands for the differences (figures E and F), it is quite noticeable that peramivir starts off ahead of Tamiflu but is lagging behind for the longer healing subjects. If one plots the confidence bands for the differences in percent healed not against time but against percent healed on peramivir 600mg (figure F), it is clear that the peramivir speeds up healing for subjects with below median healing time and delays it for subjects with above median healing time.

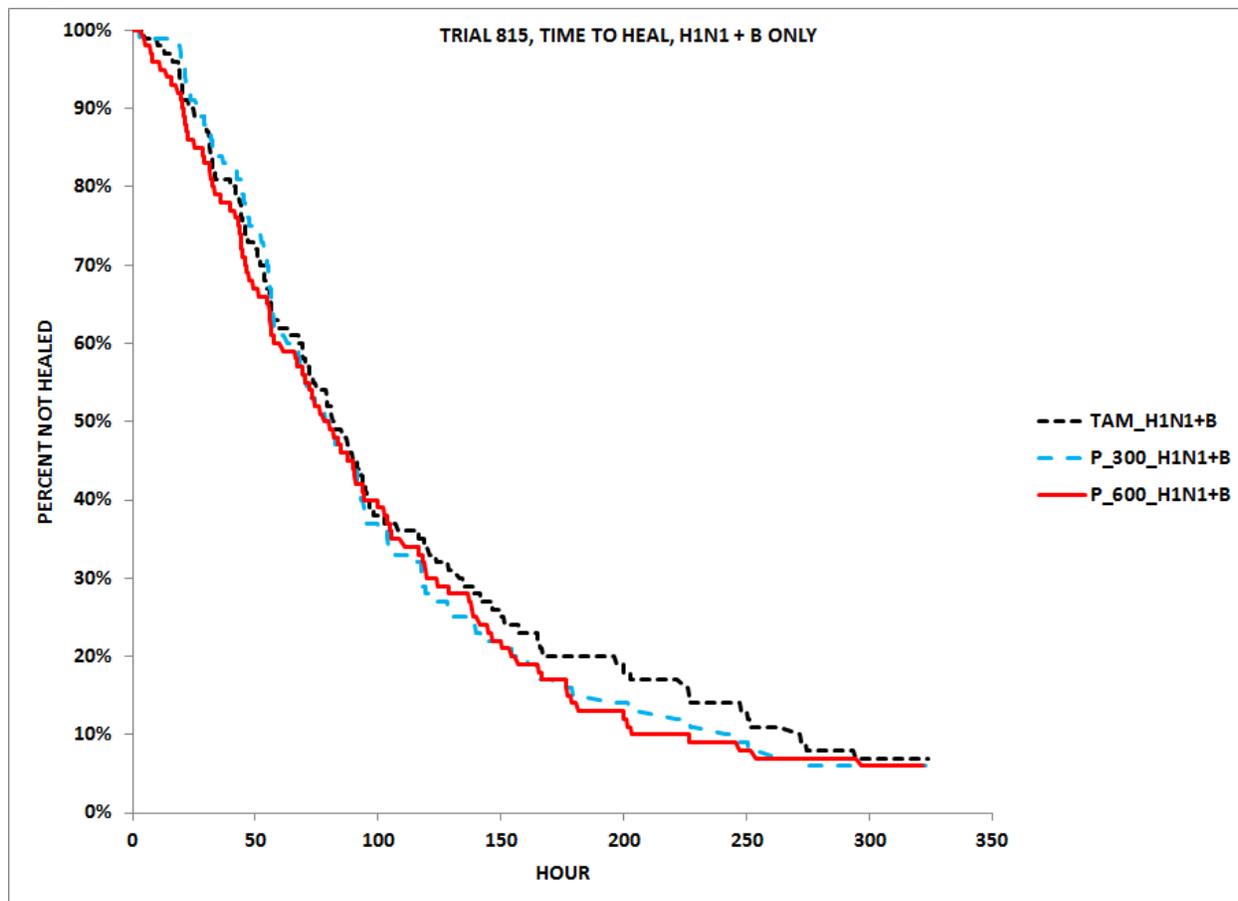
Figure 3.2.3 F



Over the entire group, peramivir and Tamiflu produce approximately the same effect but the uncertainty is fairly large; the percent healed on peramivir may be 20% larger than the percent on Tamiflu at around the first quartile; it may be 20% smaller at around the third quartile.

If one looks at the complementary group of subjects with type B or type A, H1N1, the Kaplan-Meier curves look nearly identical until about 60% of patients are healed, after which the Tamiflu subjects took longer to heal.

Figure 3.2.3 G



Assuming that Tamiflu was ineffective in these subjects, this graph suggests that peramivir conferred only a slight reduction in the time to healing for the longer healing subjects. In section 3.2.5 below, it will be seen that the median times to healing in the H1N1+B subgroup is nearly the same in all three arms in trial 815.

The estimation of the hazard ratio of peramivir to placebo from the data in trial 815, with a Tamiflu control, is performed as follows. Directly from trial 815, one can calculate the log hazard ratios of peramivir to Tamiflu for time to healing, and their standard errors. These are given in table 3.2.3 D for both peramivir arms and by four methods: using FDA corrected times and applicant's original times, and using all subjects.

TABLE 3.2.3 D
LOG HAZARD RATIOS, PERAMIVIR TO TAMIFLU, TRIAL 815

DATA, METHOD	LOG HAZARD	
	RATIO	SEE
ALL SUBJECTS		
FDA_RESULTS_300_MG	0.04198	0.07879
APPLICANT_RESULTS_300_MG	0.06269	0.07911
FDA_RESULTS_600_MG	0.04817	0.07820
APPLICANT_RESULTS_600_MG	0.03902	0.07872
H3N2 ONLY		
FDA_RESULTS_300_MG	-0.1394	0.15011
APPLICANT_RESULTS_300_MG	-0.1490	0.14951
FDA_RESULTS_600_MG	-0.0518	0.14926
APPLICANT_RESULTS_600_MG	-0.0590	0.14966

The first observation one can make here concerns the hazard ratio of peramivir to Tamiflu for all subjects, most of whom were type B and type A H1N1 subjects. Table 3.2.3 D shows that within this subgroup, the log hazard ratios were close to zero, corresponding to peramivir being very slightly superior to Tamiflu. If one were to assume that Tamiflu was effectively a placebo in this subgroup, one would also conclude that peramivir was only slightly effective in this group.

The non-inferiority computation will only be carried out using the H3N2 subgroup. First, one has the peramivir-tamiflu hazard ratios from table 3.2.3 D. The fact that these log ratios are negative shows peramivir was estimated to be slightly inferior to Tamiflu in this subgroup and supports the idea that Tamiflu is effective in this subgroup.

Second, one can calculate from the data in the Tamiflu NDA, the estimated log hazard ratios of Tamiflu to placebo for the same time to healing and their standard errors. These are presented in table 3.2.3 E.

TABLE 3.2.3 E
LOG HAZARD RATIOS, TAMIFLU TO PLACEBO, TAMIFLU NDA

TRIAL	LOG HAZARD RATIO	SEE
JV15823	0.75	0.0944
WV15670	0.75	0.0842
WV15671	0.69	0.0893
ALL	0.73	0.0510

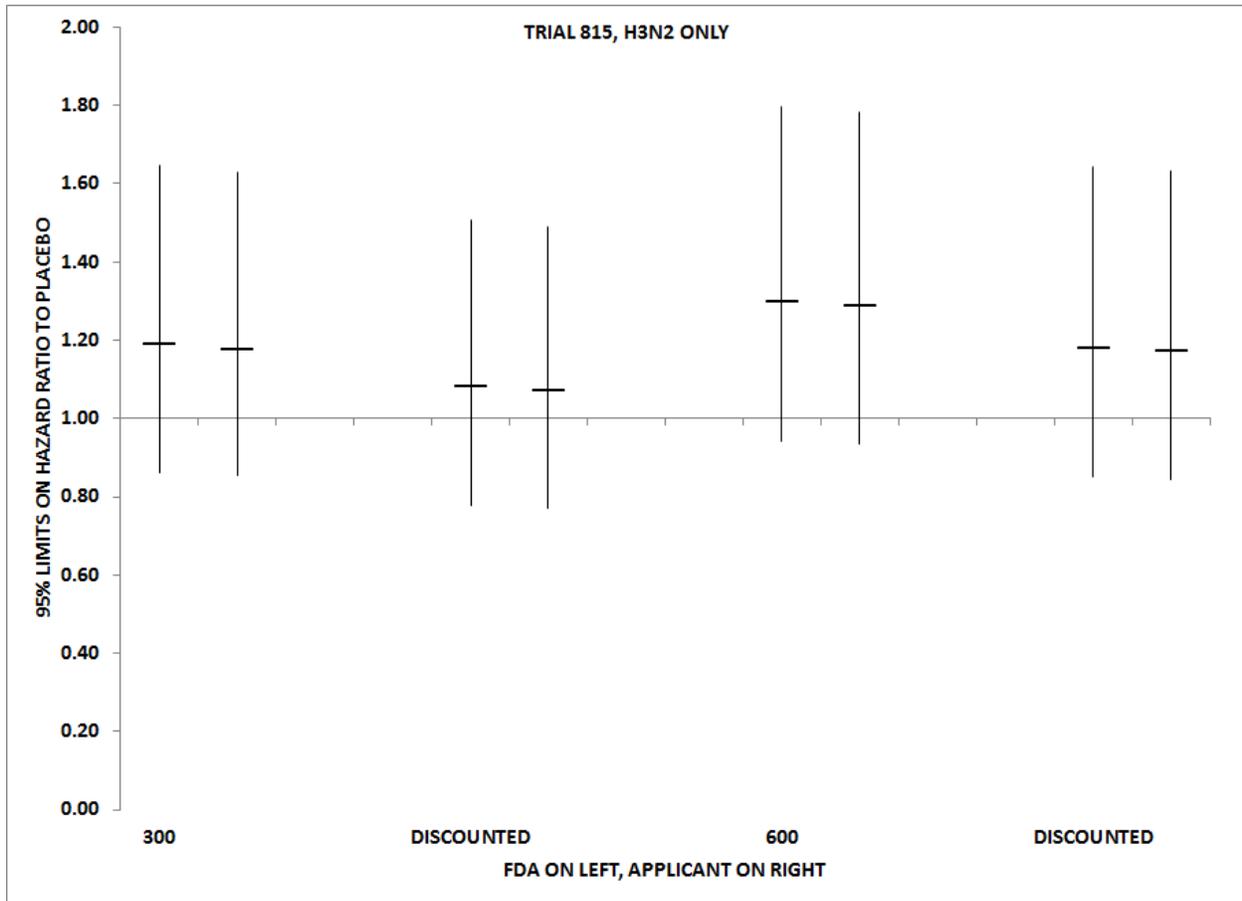
The final results for the estimated log hazard ratio of peramivir to placebo is just the sum of the log hazard ratio of peramivir to Tamiflu plus the log hazard ratio of Tamiflu to placebo. The standard error is just the square root of the sum of the squares of the standard errors of the two terms. The problem with doing this is that there is evidence that the log hazard ratio, Tamiflu to placebo, of .73 is only correct for the H3N2 cases in trial 815; the log hazard ratio, Tamiflu to placebo, is most likely zero for the H1N1 cases in trial 815. There will be further comment on this point near the end of the section.

There is still a problem with doing this so simply. The computation just described assumes that the Tamiflu/placebo hazard ratio would have been constant between the trials in the Tamiflu NDA and what would have been seen in the Tamiflu susceptible subset of trial 815, had there been a second, placebo, control in that trial. One can see that the log hazard ratio of Tamiflu to placebo varied from .69 to .75 and that the standard error varied from .084 to .094 among the three trials in the Tamiflu NDA. (These trials preceded the approval of Tamiflu so it is quite likely that there were no Tamiflu resistant strains in these trials.) This would suggest that, in the interest of prudence, one might chose an estimate of the Tamiflu/placebo hazard ratio slightly closer to one and chose an estimate of the standard error that is slightly larger. This will provide greater assurance that inter-trial variability in Tamiflu efficacy isn't producing a spurious conclusion of peramivir superiority to placebo.

Looking at the inter-trial variability in the estimates and standard errors of the Tamiflu/placebo hazard ratio within the Tamiflu NDA, the current FDA reviewer calculated a sensitivity analysis in which the hazard ratio of Tamiflu to placebo is moved 10% closer to one (compare the change from .69 to .75 in table 3.2.3 E) and the standard error of that log hazard ratio is enlarged by 10% (compare the change from .084 to .094 in table 3.2.3 E). (The FDA non-inferiority guidance recommends providing a protection against inter-trial variability by using the sum of the standard errors instead of the square root of the sum of their squares. This simple sum is always larger but doesn't use the available data on inter-trial variability from the three trials in the Tamiflu NDA.)

The results of this sensitivity analysis are presented in the following graph. This gives the estimated log hazard ratio of peramivir to placebo, for each dose of peramivir in the H3N2 (=probably Tamiflu susceptible) subset of trial 815 for each of the four estimates of the peramivir/Tamiflu hazard ratio in table 3.2.3 C combined with the estimated Tamiflu/placebo hazard ratio, using the pooled Tamiflu NDA data (last row of table 3.2.3 D) and using that same pooled data with 10% decrease in the point estimate and 10% increase in the standard error. The 10% is suggested as plausible in the previous paragraph.

Figure 3.2.3 H



One will notice that the point estimates, even with the 10% discounting, are positive and the 600mg dose is marginally significant. The one-sided p-values, with discounting, are .055 for 600mg and .15 for 300mg. More conservatively, the discounted p-values are .16 for 600mg and .32 for 300mg. (Remember that statistical significance requires one-sided p-value < .025.)

Table 3.2.3 F summarizes the exploration of efficacy testing for peramivir in trial 815. The two complementary subgroups, B+A,H1N1 and A,H3N2 were analyzed separately, with Tamiflu acting as placebo in the first subgroup and with activity relative to placebo given by the Tamiflu NDA in the second subgroup.

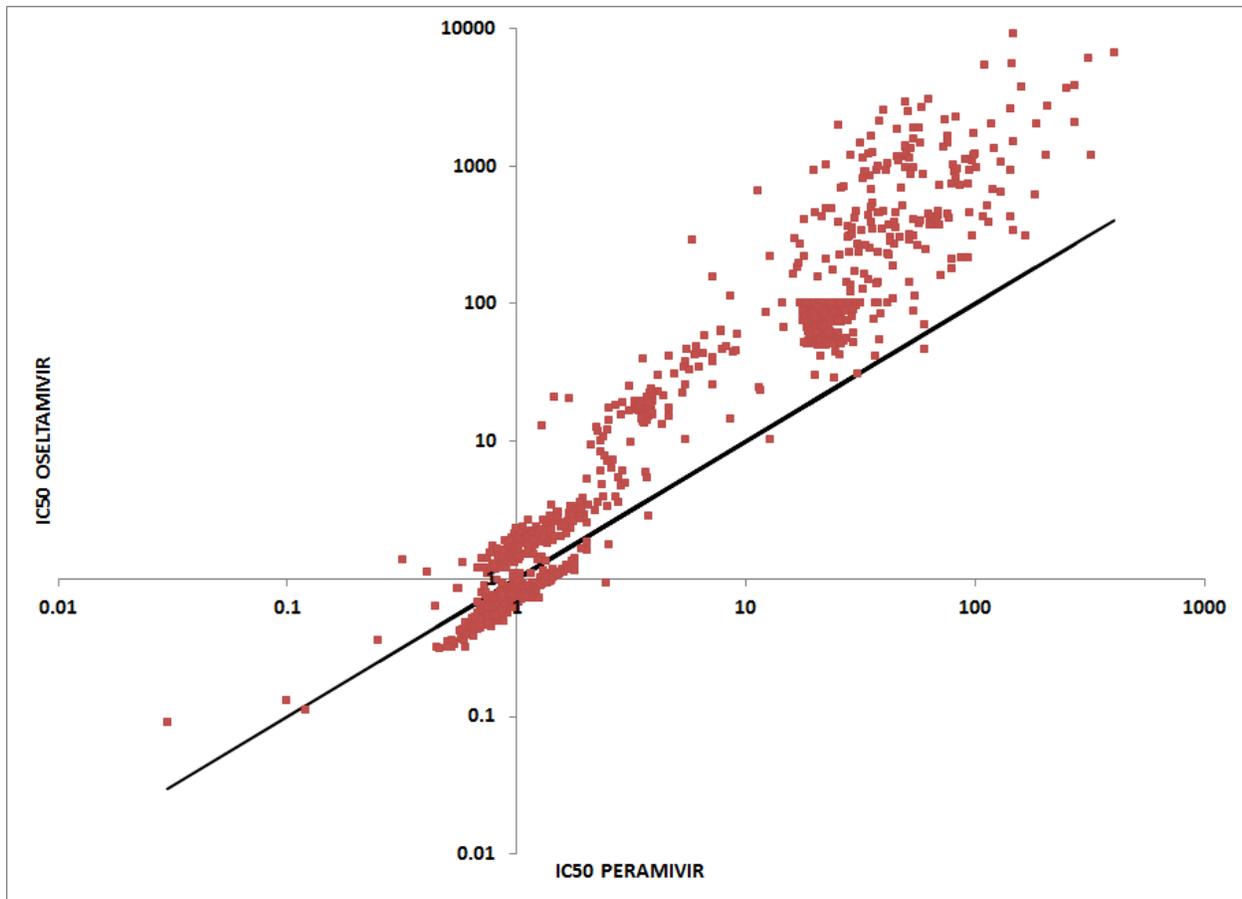
TABLE 3.2.3 F
 INFERRED P-VALUES FOR PERAMIVIR OVER PLACEBO,
 TRIAL 815, H3N2 SUBJECTS

	LOG HAZARD RATIO	STANDARD ERROR	P-VALUE
300_mg	0.175	0.165	.145
300_mg_discounted	0.080	0.168	.317
600_mg	0.263	0.165	.055
600_mg_discounted	0.168	0.168	.159

The cumulative weight of evidence from this trial is mildly in favor of efficacy for peramivir for Tamiflu susceptible subjects.

One additional observation should be made here, concerning peramivir and Tamiflu resistant virus. The following graph shows the scatterplot of IC50 for Tamiflu plotted against IC50 for peramivir, both on a log scale.

Figure 3.2.3 I



The positive association between the two is quite strong. This supports both the conclusions that type A H3N2 subjects in this trial also had Tamiflu susceptible virus and that type A H1N1 subjects had Tamiflu resistant virus.

Thus the strongest conclusions that one can draw from trial 815 are that

- 1) the data suggest peramivir is effective against strains of type A influenza that are also susceptible to Tamiflu and
- 2) peramivir does not appear to be very effective against Tamiflu resistant strains of type A influenza. The medians are nearly the same with the Kaplan-Meier curves diverging only after about 60% of subjects have healed. The strong correlation in the IC50 between the two drugs mildly argues against efficacy in the Tamiflu resistant subgroup.

3.2.4 Comparison of Hazard Ratio of Peramivir to Placebo for Types A and B

The following graph (figure 3.2.4 A) is an initial exploration of the difference in the efficacy of peramivir relative to control between type A and type B influenza. The applicant has been inconsistent in whether results should use all ITTI subjects randomized (which results in larger sample sizes and smaller confidence intervals in trials 211 and 311) or whether only type A infected subjects should be included (which results in more favorable results in trial 212). The following graph gives the point estimates and 95% confidence intervals for the hazard ratio of time to symptom alleviation for peramivir to placebo in trials 211, 311, 411=results from 211 and 311 pooled, and trial 212. In trial 212, the type A H3N2 and H1N1 Wild type subjects only were used; the subjects with H275Y substitution are not included in this graph. Trial 722 used almost exclusively type A subjects; trial 815 has a Tamiflu control and the hazard ratio of Tamiflu to placebo by type is not available.

The graph gives two 95% confidence intervals for each trial and peramivir dose. The left interval in each pair is the hazard ratio to placebo for type A infected subjects, the right interval is for type B infected subjects. One will notice that, in every case, the hazard ratio for type A infected subjects is greater than 1. In trials 211 and in the pooling of 211 and 311, the confidence intervals are about 3/4 above zero, with one-side p-values in the range .1 to .25.

In contrast, the results for type B influenza are uniformly unfavorable. The intervals are wide enough to be statistically compatible with zero effect so no one should conclude peramivir is inferior to placebo for type B.

Figure 3.2.4 A

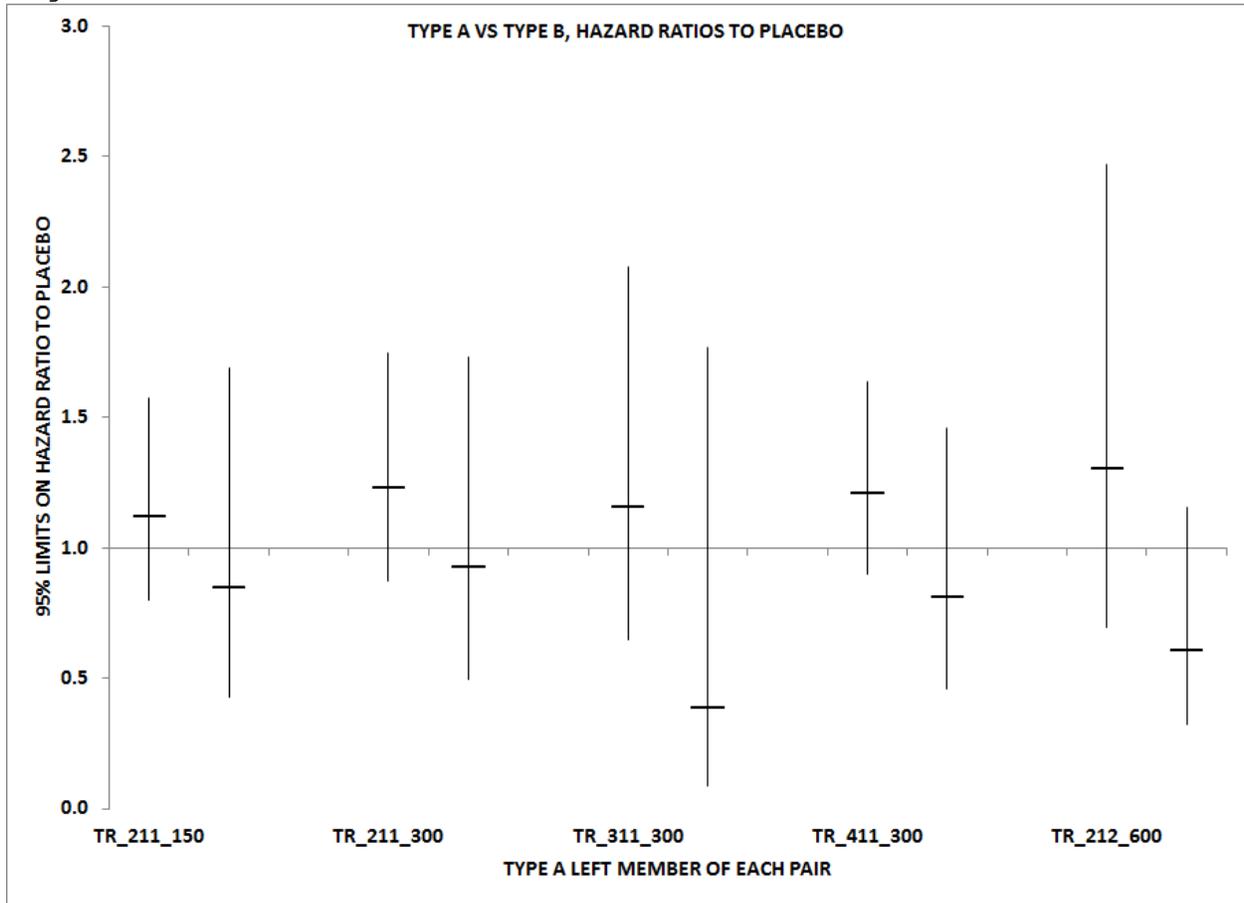


Table 3.2.4 A gives the frequency of influenza types by trial and arm, i.e. the sample sizes that go into the above intervals. Trial 212 in this table uses only H3N2 and H1N1 Wild type for type A.

TABLE 3.2.4 A

STUDY, DOSE	FREQUENCY OF INFLUENZA TYPES BY TRIAL AND ARM			
	TRT	TYPE_A	TYPE_B	TYPE_UNK
TR_211_150	IM Placebo	80	24	0
TR_211_150	IM 150 mg qd	81	19	0
TR_211_300	IM 300 mg qd	76	25	1
TR_311_300	IM Placebo	21	4	0
TR_311_300	IM 300 mg qd	46	6	2
TR_411_300	IM Placebo	101	28	0
TR_411_300	IM 300 mg qd	122	31	3
TR_212_600	IM Placebo	23	25	0
TR_212_600	IM 600 mg qd	26	27	0
TR_722_300	Placebo	100	0	0
TR_722_300	Peramivir 300	97	2	0
TR_722_600	Peramivir 600	97	1	0
TR_815_300	Oseltamivir	327	23	15
TR_815_300	Peramivir 300	330	21	13
TR_815_600	Peramivir 600	325	26	13

3.2.5 Confidence Intervals for Medians from Kaplan-Meier Curves

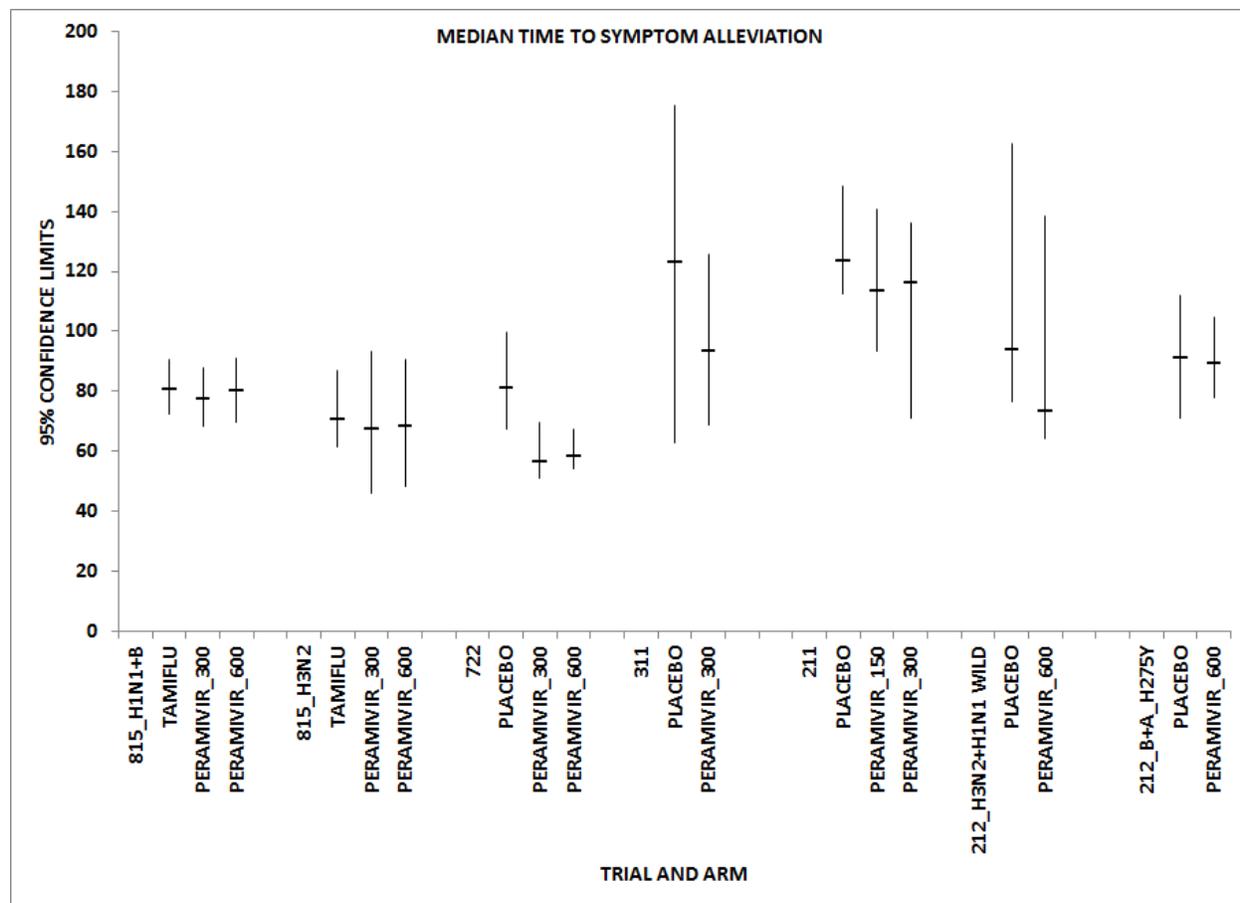
Table 3.2.5 A gives the 95% confidence intervals for the median time to symptom alleviation for each trial and arm. The results are computed twice. The first row uses the protocol and FDA definition of time to healing as the first of two consecutive symptom free entries. The second row uses the applicant's computation as the first of two consecutive symptom free entries after every entry which has a moderate or severe or missing symptom (i.e. after every suspected relapse).

The table splits trials 815 and 212 in two. For trial 815, the split is type A H3N2 vs type B+ type A, H1N1. One will notice that for all three arms, the H3N2 medians are shorter, supporting the contention that the two drugs are more effective against that sub-type. For trial 212, the split is Tamiflu susceptible subjects with type A H1N1 Wild or H3N2 vs the second half includes type A H1N1 with Tamiflu resistant H275Y and type B. For this trial, one should notice that the difference between placebo and peramivir is larger for the Wild +H3N2 half of the data.

TABLE 3.2.5 A
 MEDIAN HOURS TO SYMPTOM ALLEVIATION
 WITH 95% CONFIDENCE INTERVALS

TRIAL_815		
Tamiflu	Peramivir 300	Peramivir 600
All		
78.4 (70.8, 86.9)	75.8 (67.1, 87.1)	77.5 (68.8, 88.1)
82.1 (73.6, 91.2)	78.7 (68.7, 88.1)	81.0 (73.2, 91.6)
B+H1N1		
81.1 (72.4, 90.5)	78.1 (68.2, 88.1)	80.7 (69.8, 91.3)
86.4 (76.7, 94.1)	79.3 (69.5, 89.2)	87.3 (76.5, 98.7)
H3N2		
70.9 (61.7, 87.2)	67.6 (45.9, 93.2)	68.8 (48.1, 90.6)
75.1 (64.6, 92.9)	69.5 (54.6, 94.9)	70.8 (47.7, 92.2)
TRIAL_722		
Placebo	Peramivir 300	Peramivir 600
81.8 (67.5, 99.9)	57.1 (50.9, 69.8)	58.6 (54.0, 67.5)
81.8 (67.9, 101.5)	58.1 (50.9, 74.1)	60.2 (54.4, 70.2)
TRIAL_211_Wise_Excluded		
Placebo	Peramivir 150	Peramivir 300
124.0 (112.5, 148.4)	114.0 (93.6, 140.8)	116.8 (71.3,
136.0)		
130.4 (113.9, 161.3)	126.5 (96.3, 147.4)	118.2 (77.8,
135.8)		
TRIAL_311_Wise_Excluded		
Placebo	Peramivir 300	
123.6 (63.0, 175.5)	94.0 (68.7, 125.8)	
117.9 (64.9, 172.6)	111.0 (80.3, 127.9)	
TRIAL_211_311_POOLED_Wise_Excluded		
Placebo	Peramivir 150	Peramivir 300
124.2 (112.9, 142.4)	114.0 (93.6, 140.8)	110.3 (78.0,
128.2)		
125.8 (113.2, 159.3)	126.5 (96.3, 147.4)	114.5 (87.5,
130.4)		
TRIAL_212_H3N2+WILD_Wise_Excluded		
Placebo	Peramivir 600	
94.4 (76.4, 162.5)	73.7 (64.1, 138.5)	
95.7 (76.1, 162.3)	72.1 (64.2, 138.3)	
TRIAL_212_H275Y+B_Wise_Excluded		
Placebo	Peramivir 600	
91.4 (70.8, 112.0)	89.8 (78.1, 104.8)	
109.8 (90.0, 125.0)	93.9 (82.9, 114.5)	

The following graph gives a quick overview of the preceding table, using the medians computed from the FDA reviewer's healing time (= first of two consecutive diary entries with alleviation). Figure 3.2.5 A



One can see, either from the graph or the table, that in trial 722 peramivir was clearly superior to placebo since the confidence intervals for the arm medians don't even overlap. There is also a suggestion of superiority of peramivir in trials 311, 211, and 212 (H3N2 and H1N1 Wild subset). In all of these, the placebo median is longer but there is considerable overlap in the intervals.

It would be desirable to have the 95% confidence intervals for the difference in the medians in order to assess statistical significance. That is not easy. The medians and their confidence intervals are read from the Kaplan-Meier curves. In those curves, there are standard errors for the percent healed at each time but there are not standard errors for the time at which a given percent (say 50%) are healed. The confidence interval for the median of an individual arm can be computed indirectly by looking for the time points for which upper and lower confidence curves cross 50% healed. Since those two points are not computed using a standard error on the time scale, it is not possible to obtain a standard error and a confidence interval for the difference in the medians of two arms. Instead statistical significance has to be assessed by looking at the confidence bands for the differences in percent healed, as was done in section 3.2.1 above.

A confidence interval for the difference in the medians can be obtained by the method of bootstrapping. To compute the bootstrap intervals for every trial and every peramivir-placebo comparison requires substantial computer time. The determination of statistical significance by use of the Cox regression is preferable.

3.2.6 Pooling Evidence from Trials Excluding Tamiflu-Controlled 815

One could be content with the observation made above that trial 722 by itself is a convincing demonstration of efficacy. One could also dismiss the lack of statistical significance in trials 211, 311, and 212 as a result of small sample sizes and lower doses. Nonetheless, somewhat greater comfort with the overall conclusion may result from looking at a statistically valid pooling of the results of the placebo controlled trials.

Tables 3.2.6 A and B give the Fisher p-values for pooling results from trials 722, 211, 311, and 212. Each individual study produced a p-value obtained from the log hazard rates of a peramivir arm to placebo obtained by stratified Cox proportional hazards regressions. This produces one p-value for each study and peramivir dose. These were synthesized into a single p-value by Fisher's method, which sums up the values of $-2 \cdot \log(\text{individual p-value})$ and compares that to a chi-square distribution on degrees of freedom = $2 \cdot \text{number of individual p-values}$. This was done 4 times, using the FDA reviewer's and the applicant's time to healing and using all subjects in the ITTI group or using only types expected to be susceptible. Expected susceptible types were type A in trials 211 and 311, type A H1N1 Wild or type A H3N2 in trial 212. The Fisher pooled p-value is computed for trials 211, 311, and 212 alone and for those three trials plus trial 722.

TABLE 3.2.6 A
 FISHER P-VALUES POOLING ALL TRIALS
 WITH PLACEBO CONTROLS, FDA RESULTS

USING ALL SUBJECTS	TRIAL	P-VALUE	FISHER POOLED P-VALUE
Trial 722 300mg	.0039	3.9×10^{-4}	
Trial 722 600mg	.0006		
Trial 211 150mg	.30	.29*	
Trial 211 300mg	.19		
Trial 311 300mg	.49		
Trial 212 600mg	.29		
USING SUSCEPTIBLE TYPES ONLY			
Trial 722 300mg	.0039	1.3×10^{-4}	
Trial 722 600mg	.0006		
Trial 211 150mg	.250	.127*	
Trial 211 300mg	.12		
Trial 311 300mg	.31		
Trial 212 600mg	.20		

* Result pooling 211, 311, 212 only

TABLE 3.2.6 B
FISHER P-VALUES POOLING ALL TRIALS
WITH PLACEBO CONTROLS, FDA RESULTS

APPLICANT_RESULTS_STRATIFIED				
USING ALL SUBJECTS	TRIAL	P-VALUE	FISHER POOLED P-VALUE	
Trial 722 300mg	.0021		1.9×10^{-4}	
Trial 722 600mg	.0005			
Trial 211 150mg	.32		.259*	
Trial 211 300mg	.14			
Trial 311 300mg	.49			
Trial 212 600mg	.29			
USING SUSCEPTIBLE TYPES ONLY				
Trial 722 300mg	.0021		6.1×10^{-5}	
Trial 722 600mg	.0005			
Trial 211 150mg	.250		.110*	
Trial 211 300mg	.11			
Trial 311 300mg	.30			
Trial 212 600mg	.18			

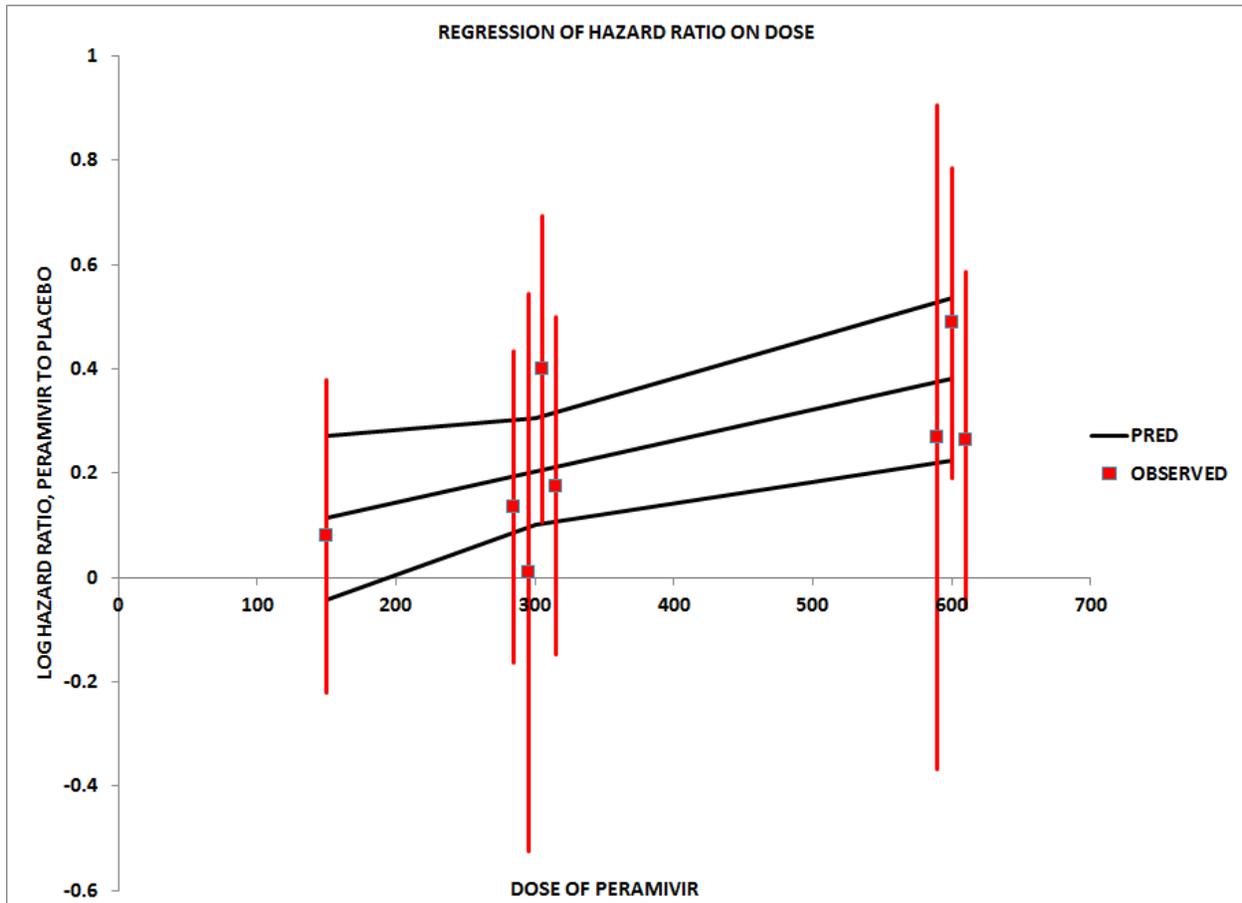
* Result pooling 211, 311, 212 only

One can see that the three BioCryst trials have p-values of .26-.29 after pooling by Fisher's method if all subjects are included but if only types expected to be susceptible are included, the pooled p-values are .11-.13, nearly marginally significant. When the results of trial 722 are included, the pooled p-values are all smaller than 4×10^{-4} , which is smaller than the 6.25×10^{-3} customary for two successful trials. In other words, the BioCryst studies are not collectively significant but they are supportive of efficacy in type A influenza without the H1N1 H275Y substitution. Combined with trial 722, they are sufficient to demonstrate efficacy against Tamiflu susceptible strains of type A influenza.

3.2.7 Dose Response Modelling

As part of the exploration of the efficacy of peramivir, the FDA reviewer fitted a dose-response model to all the trials. For each trial and peramivir dose, the reviewer computed the log hazard ratio of peramivir to placebo, with its confidence interval. (These have already been seen above.) For trial 212, only type A H3N2 and H1N1 wild type subjects were used. For trial 815, only type A H3N2 subjects were used and then combined with the Tamiflu/placebo ratio as described in section 3.2.3 above. To these 8 points (211 at 150mg, 211 at 300mg, 311 at 300mg, 722 at 300mg, 815 at 300mg, 212 at 600mg, 722 at 600mg, and 815 at 600mg), the reviewer fit a linear regression of log hazard ratio on dose. The following graph (figure 3.2.7 A) gives the fitted curve together with 95% confidence bands on the fitted curve (in black) together with the observed log hazard ratios and their by-study-and-dose confidence intervals in red. The order of the 8 studies from left to right is as listed above.

Figure 3.2.7 A

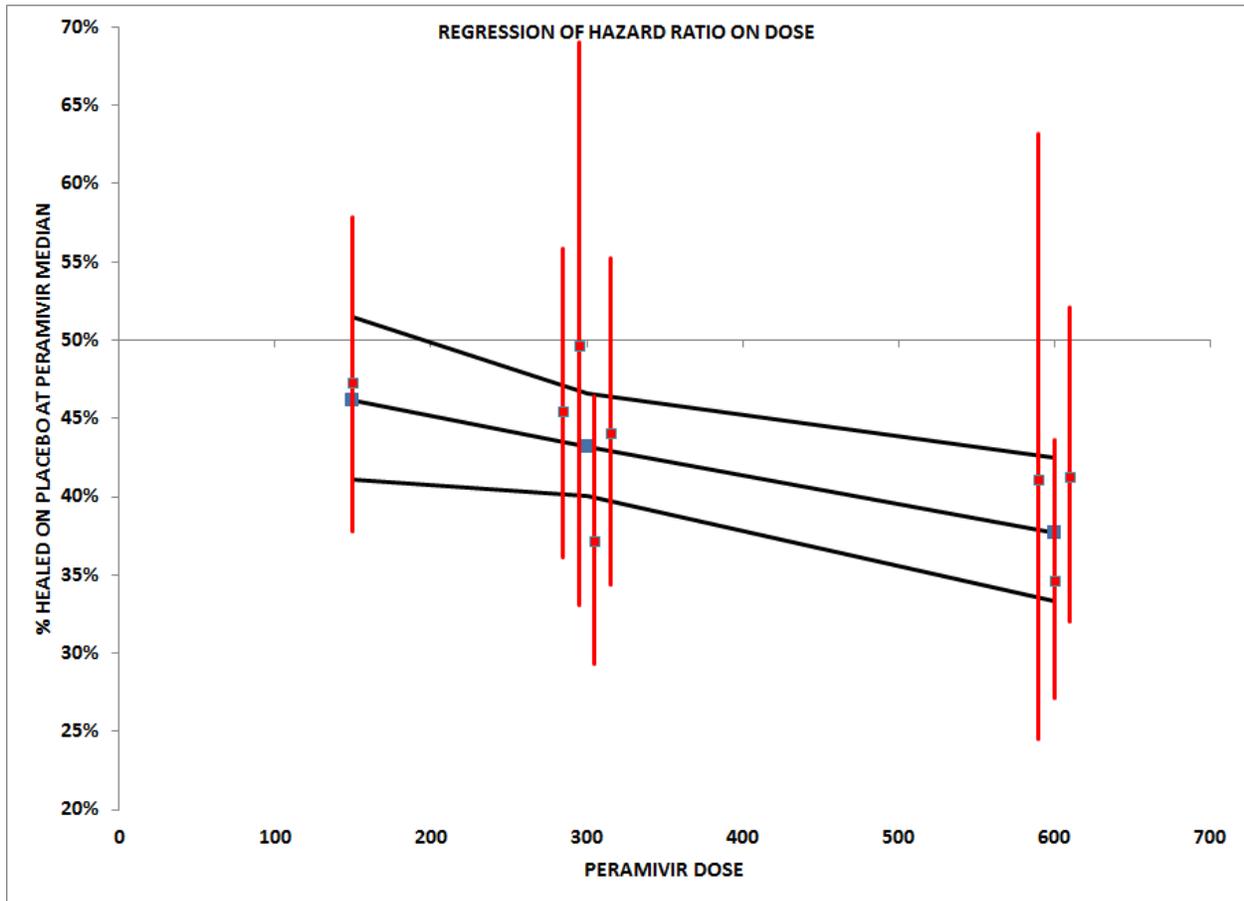


One can see that the dose response curve fits the observed hazard ratios fairly well with a statistically significantly positive slope. This supports the conclusion of peramivir efficacy (the slope would be zero for an ineffective drug). There was no attempt to fit a non-linear curve so the possibility that the curve flattens out between 300mg and 600mg cannot be ruled out. It seems likely that the 150mg dose is less effective than the 600mg dose; one cannot be certain with respect to the difference between 300mg and 600mg.

It can be difficult to interpret exactly what the log hazard ratio means. Mathematically, the percent not healed on peramivir = percent not healed on placebo, raised to the power of the hazard ratio. E.g. if the hazard ratio were 1.12, then when 50% of placebo subjects had not healed, only 46% ($=.5^{1.12}$) of peramivir subjects would not be healed; when 50% of peramivir subjects were not healed, 53.8% ($=.5^{(1/1.12)}$) of placebo subjects would not be healed.

The mathematics of this example has been applied to the graph in figure A to give the plots of the observed and modelled percentages of subjects who would heal on placebo by the time 50% of subjects had healed on peramivir (=median healing time for peramivir).

Figure 3.2.7 B



One can see that the percent of patients healed on placebo is estimated to be only 45% at the median healing time for peramivir 150mg and to be a bit less than 40% at the median healing time for peramivir 600mg.

3.3 FDA Analysis of Time to Fever Resolution

In addition to the material on symptom alleviation presented in section 3.2 above, the applicant also calculated times to fever resolution. Fever resolution required subjects to have an axillary temperature above 38.0 C at treatment start and to have temperature decline to below 37.20 C. This is a secondary endpoint expected to be somewhat correlated with time to symptom alleviation. The establishment of a treatment effect on time to fever resolution would lend support to the efficacy conclusions with respect to symptom alleviation.

In the Shionogi trial, 722, there were two separate times to fever resolution. In what follows, FEVER will be the abbreviations used for the endpoints in the previous paragraph. FEVER12 will be the abbreviation for the endpoint which required axillary temperature to be below 37.00 C for at least 12 hours and for no anti-pyretic medicine to be used in the previous 4 hours.

Table 3.3 A gives the correlations between the two fever endpoints and the time to symptom alleviation. The two following graphs (figures 3.3 A and B) give the scatter plots of times to fever resolution and symptom alleviation. One will notice that FEVER12 is more highly correlated with time to symptom alleviation but even that correlation is weak, as the scatter plots confirm. The association between fever resolution and symptom alleviation is statistically significant but not particularly clinically important.

TABLE 3.3 A
CORRELATIONS OF FEVER RESOLUTION AND SYMPTOM ALLEVIATION TIMES
TRIAL 722

Treatment	RSQ	R	95% CONFIDENCE		N
			LOWER	UPPER	
FEV12					
IV_Placebo	0.12927	0.35954	0.17551	0.51929	100
IV_300_mg_q.d.	0.03651	0.19107	-0.00659	0.37436	99
IV_600_mg_q.d.	0.04955	0.22260	0.02530	0.40322	98
Pooled	0.096999	0.31145	0.20490	0.41070	297
FEV					
IV_Placebo	0.033082	0.18188	-0.01508	0.36526	100
IV_300_mg_q.d.	0.014310	0.11962	-0.07967	0.30973	99
IV_600_mg_q.d.	0.007696	0.08773	-0.11162	0.28029	99
Pooled	0.027172	0.16484	0.052195	0.27334	298

Figure 3.3 A

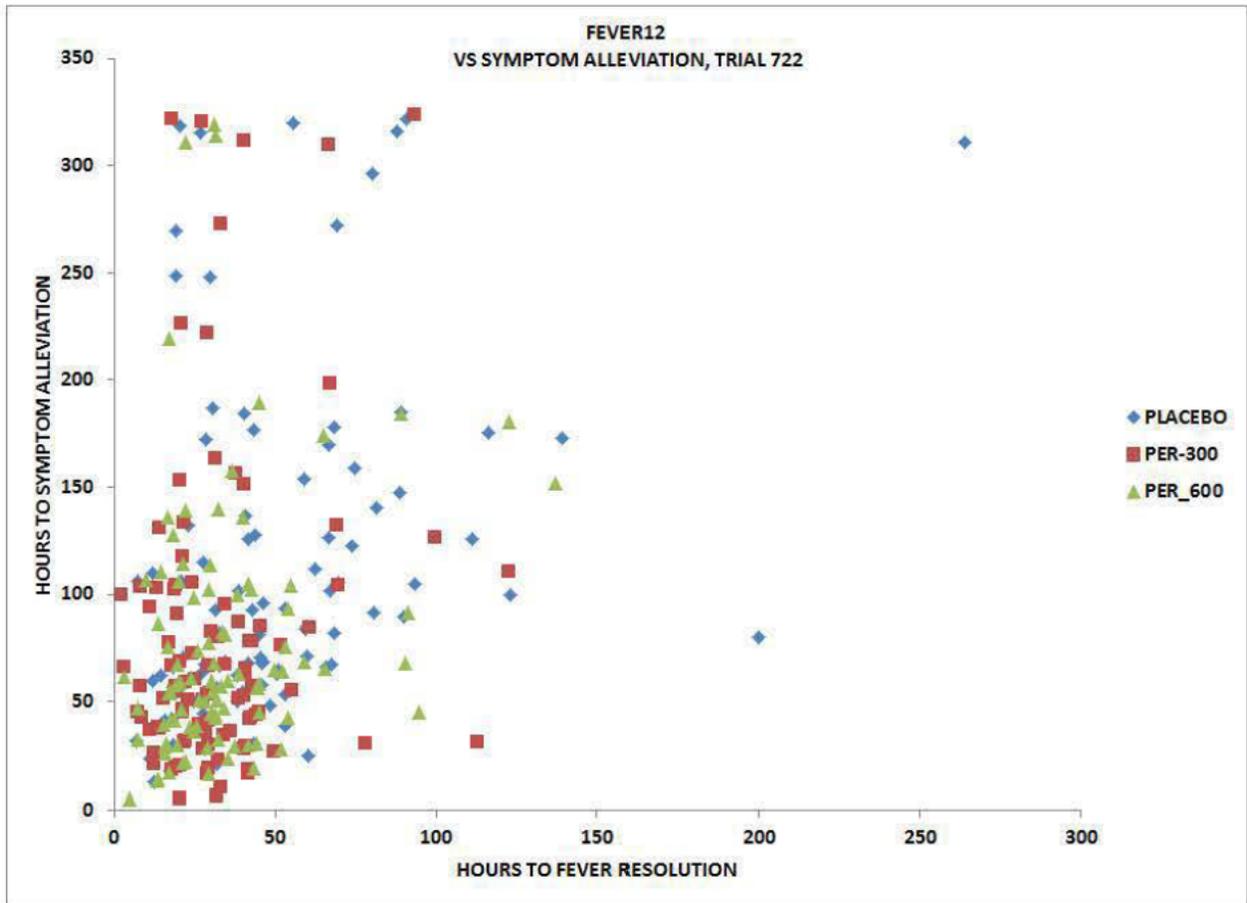
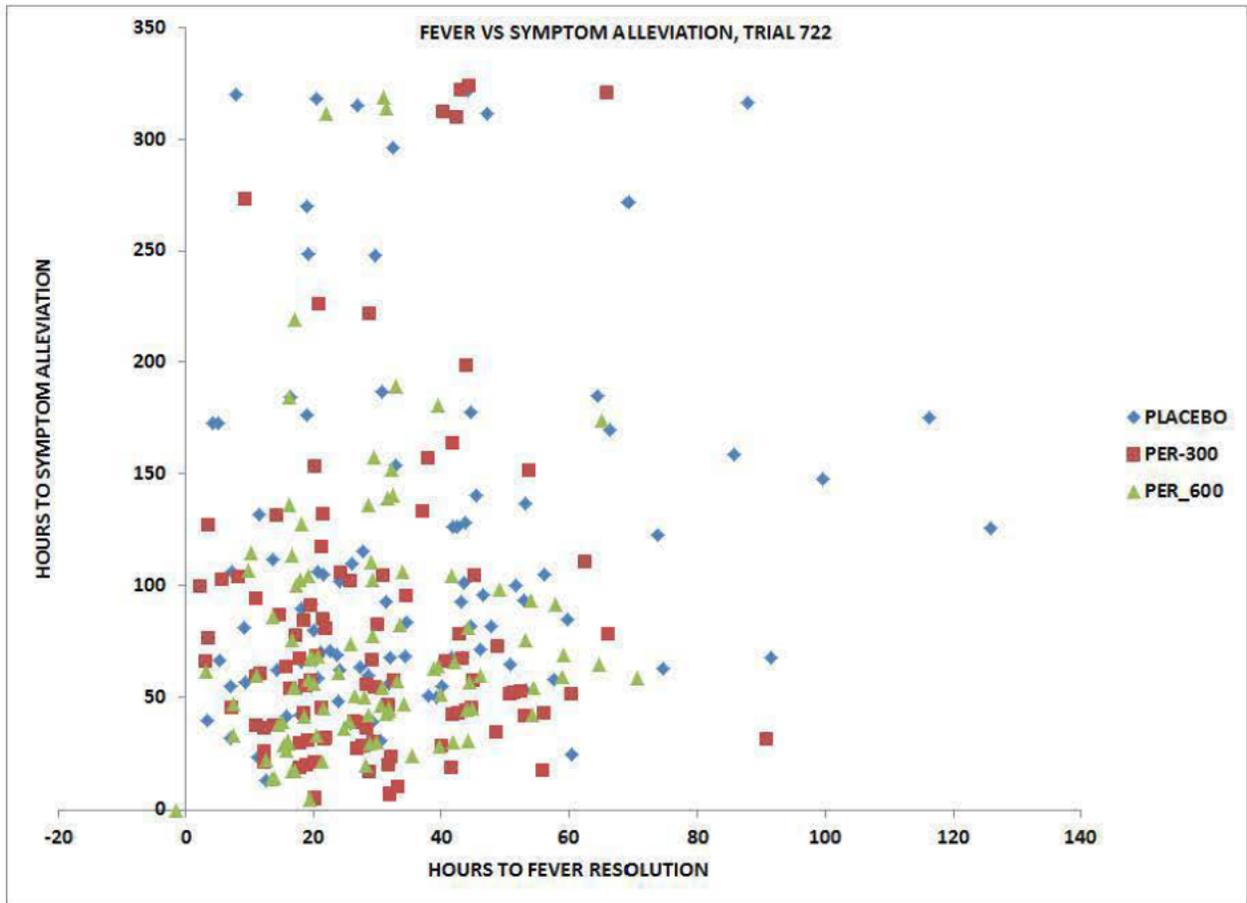
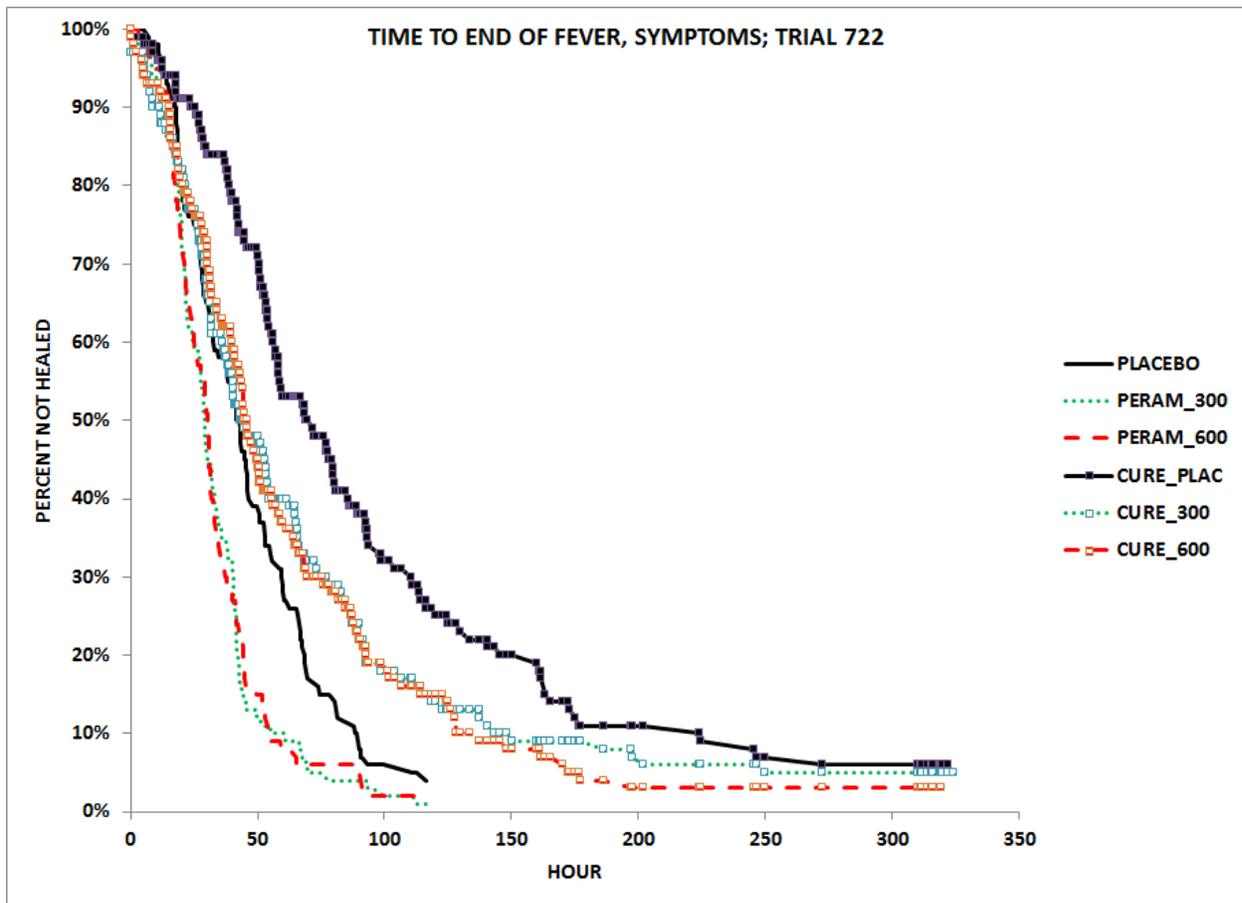


Figure 3.3 B



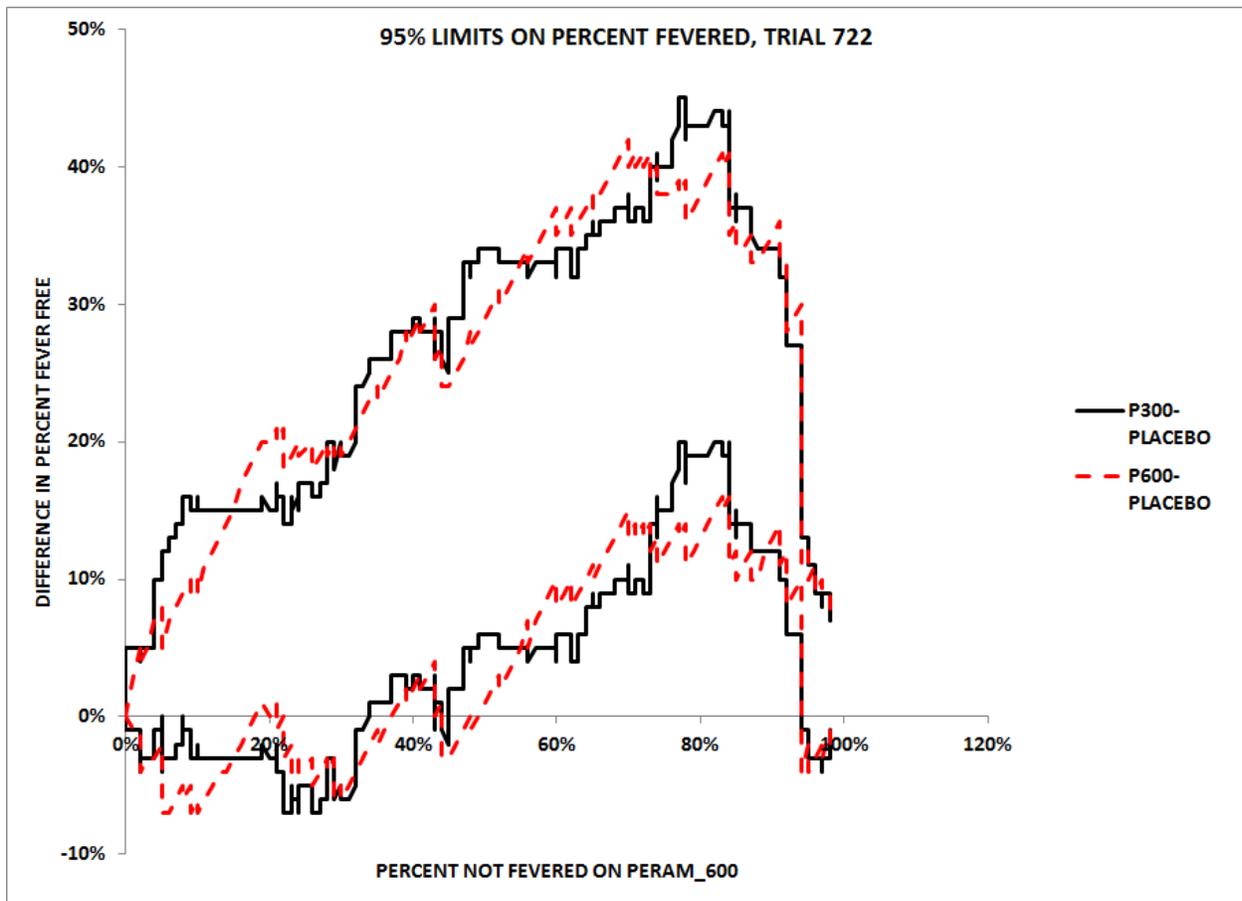
There is a statistically significant treatment effect on the FEVER12 measure for time to fever resolution in trial 722. The treatment effect with the FEVER version of the endpoint is not significant. These conclusions can be seen in the following graphs which show the Kaplan-Meier curves (figures C and E) for time to fever resolution for all three arms in trial 722 and the 95% confidence bands for the difference in percent without fever (figures D and F) between the peramivir and placebo arms.

Figure 3.3 C: Results with the FEVER12 endpoint:



For comparison purposes, the Kaplan-Meier curves for time to symptom alleviation are also included in this graph. In each pair of curves for a given arm, the time to symptom alleviation is the right curve, the one that takes longer to reach zero. In each triplet of curves corresponding to one endpoint, one can see that placebo is clearly separated from the two peramivir arms and takes longer to reach zero; the two peramivir arms are effectively superimposed for each endpoint.

Figure 3.3 D



The 95% confidence bands for the peramivir-placebo differences in percent fever free are plotted with percent fever-free on peramivir 600 mg rather than hour on the x-axis. As was explained in section 3.2 where similar graphs for time to symptom alleviation were given, using percent resolved on one arm may be more informative because time to healing varies among different influenza seasons and strains. One can see that there is a statistically significantly higher percent of subjects fever free on peramivir from the 20th percentile on fever resolution on.

Figure 3.3 E: Results with FEVER

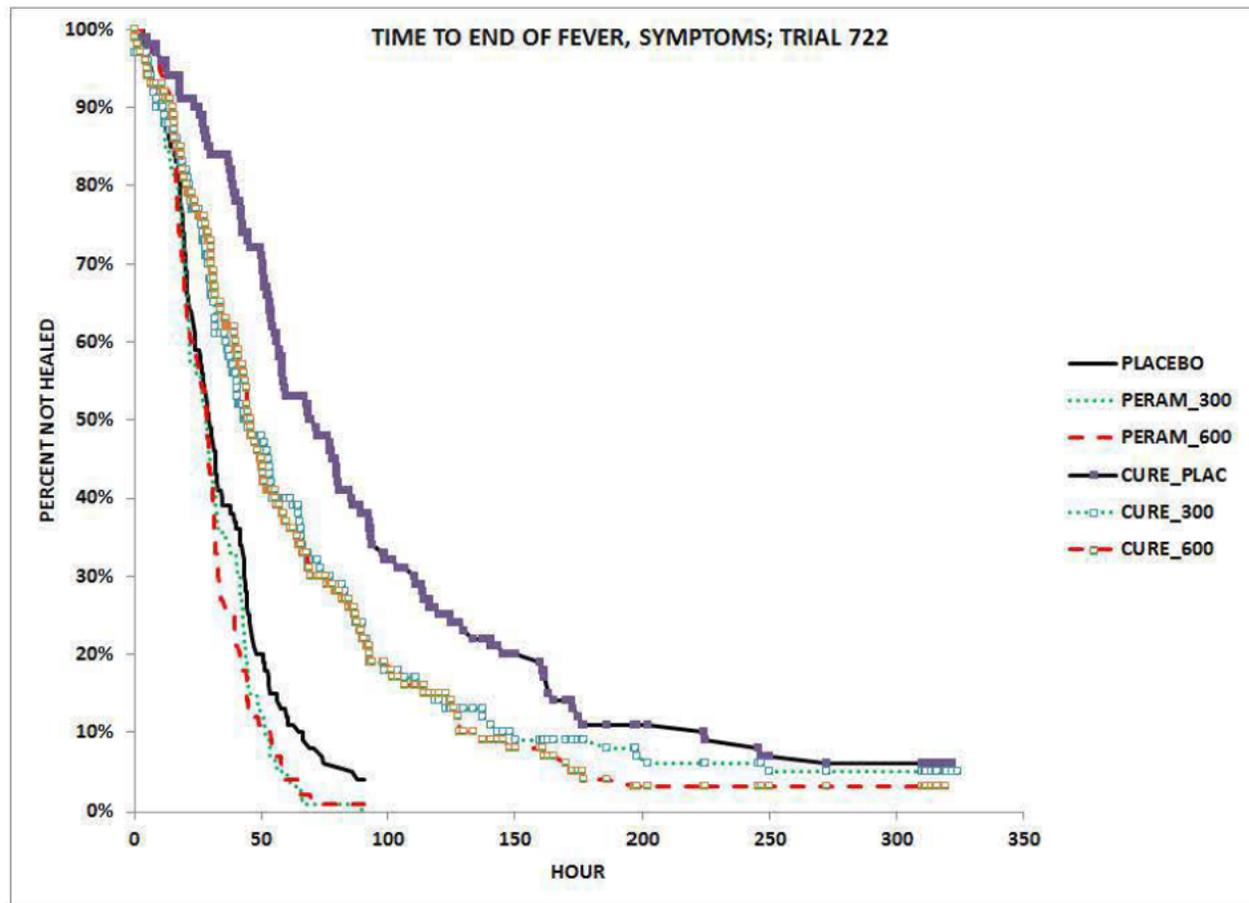
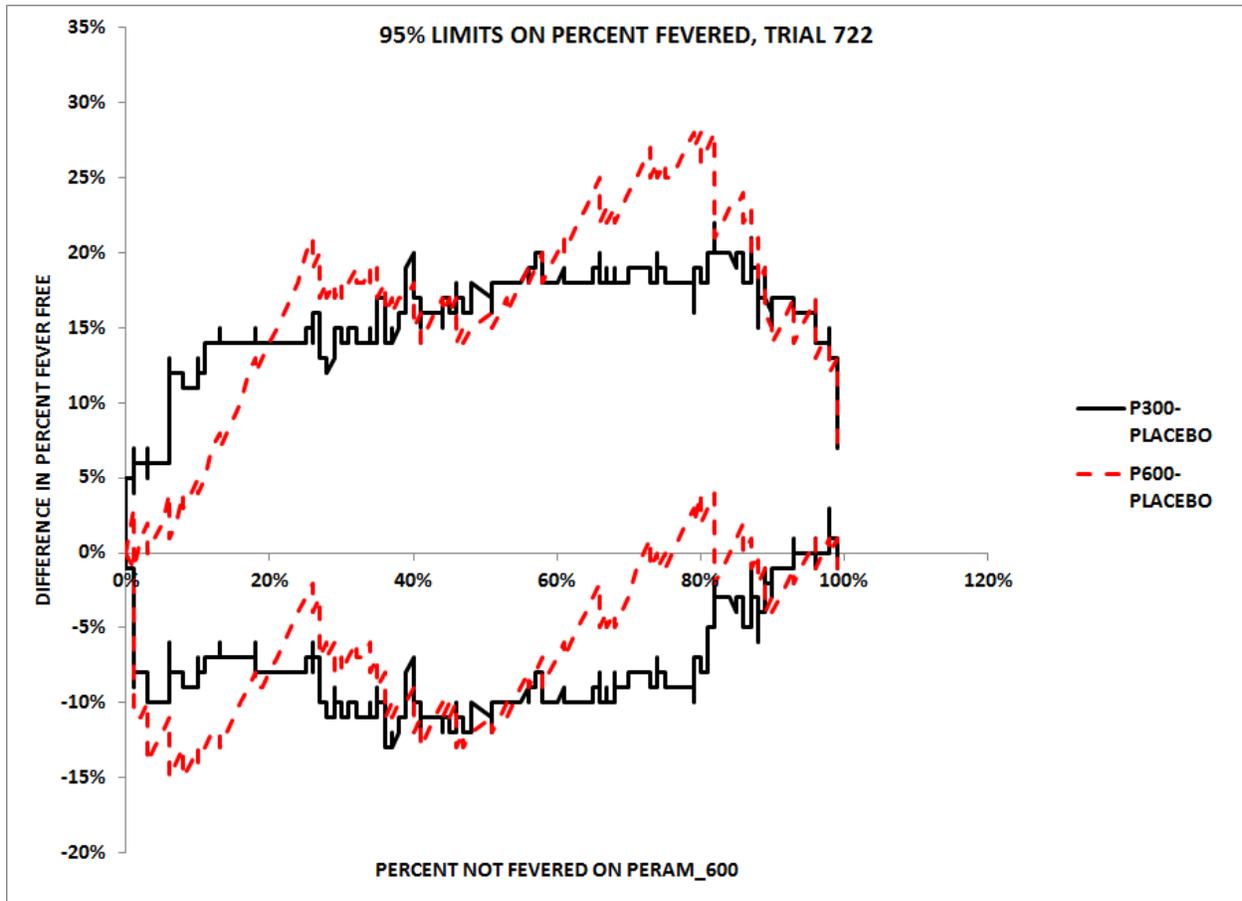


Figure 3.3 F



One will notice that with the second choice of fever resolution measure, the separation of placebo from peramivir is weaker and, as the confidence band graph shows, not statistically significant.

There are no results on the differences between Tamiflu and placebo with respect to time to fever resolution so it is impossible to get any relevant conclusions from trial 815 on this endpoint. BioCryst only used the FEVER12 endpoint, not the FEVER endpoint. In fact, Shionogi's initial analysis used the FEVER endpoint in trial 722 and BioCryst re-analyzed the Shionogi data with FEVER12.

There is another problem with the endpoint in the BioCryst studies. In these three studies, fever was identified at screening and a respectable number of subjects resolved their fever before treatment. Specifically 13 out of 82 subjects (16%) in trial 311, 59 out of 342 (17%) in trial 211, and 106 out of 399 (27%) in trial 212. The denominators in the above list are the number with any record for fever resolution. Subjects without fever at the initiation of treatment cannot show a treatment effect. Thus, in the analyses that follow only the 69, 283, and 339 subjects with positive times to fever resolution in trial 311, 211, and 212 are included.

For completeness, we give the Kaplan-Meier curves and the plots of 95% confidence bands for trials 211, 311, and 212 below. (Data from John Wise is excluded in these results.)

Figure 3.3 G

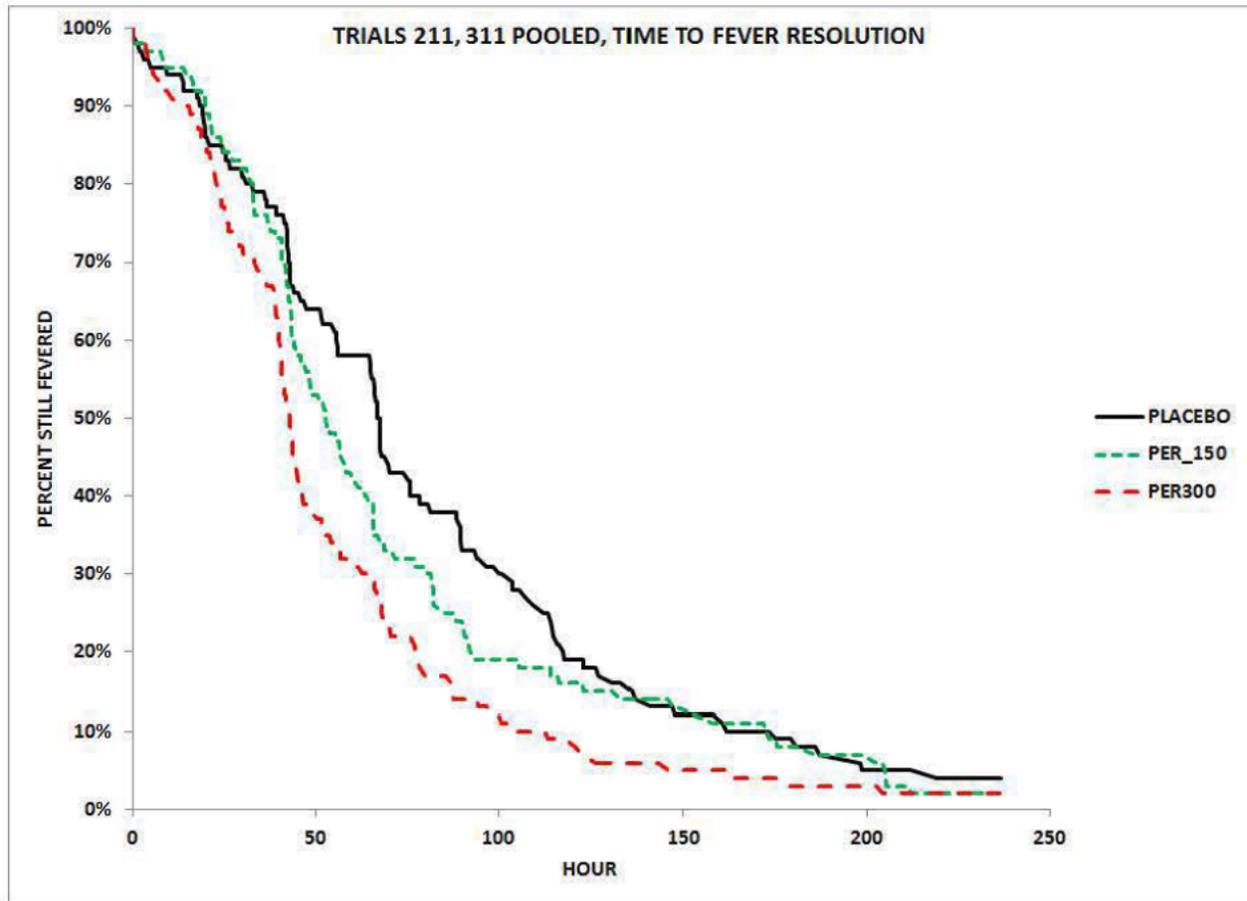
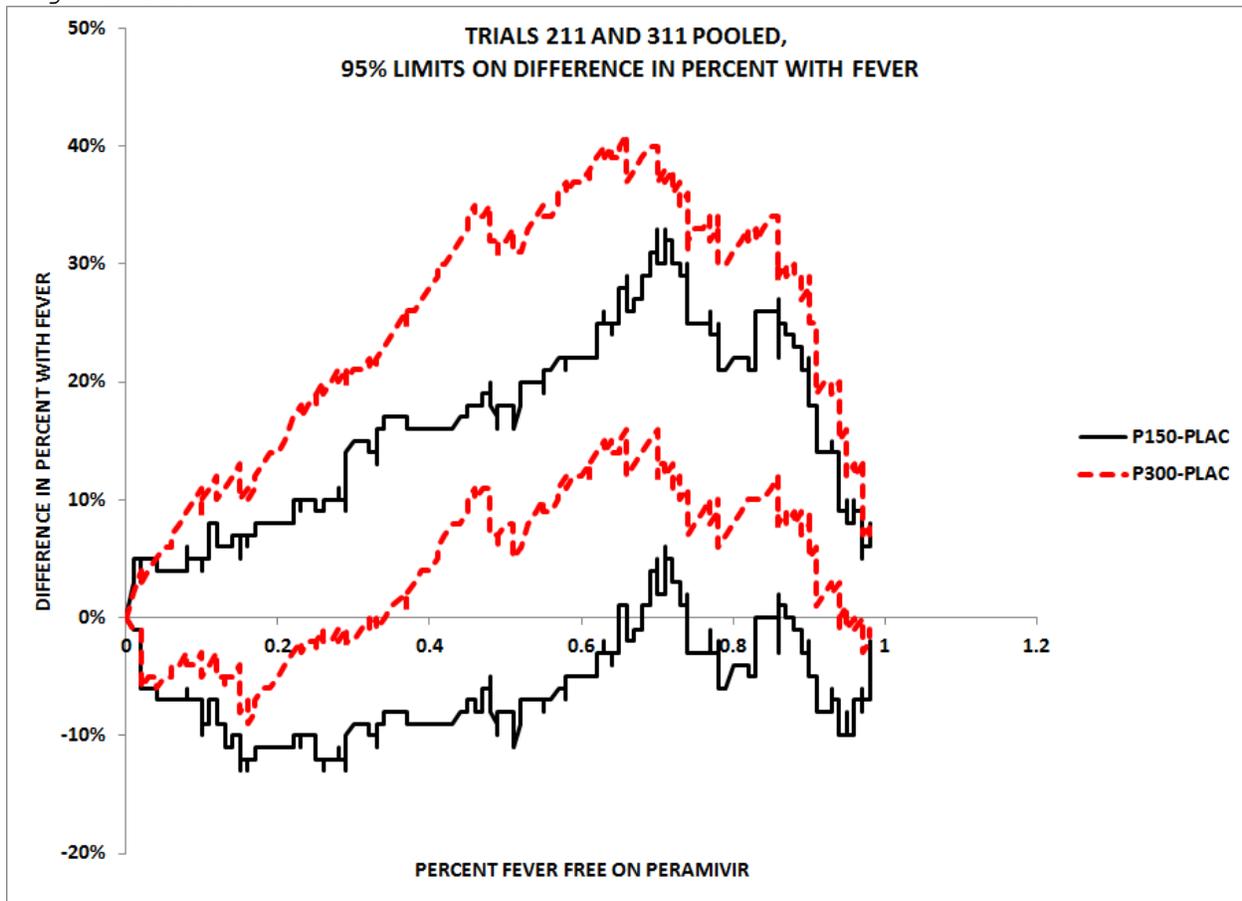


Figure 3.3 H



There is a statistically significant reduction in time to fever resolution for the 300 mg peramivir arm in trials 211 and 311 but the slight apparent effect in the 150 mg arm was not significant. In trial 212, there was no treatment effect even though the dose was higher. This accords with the observations made in section 3.2 above that trial 212 enrolled mainly subjects with Tamiflu resistant (and, therefore, likely peramivir resistant) influenza strains.

Figure 3.3 I

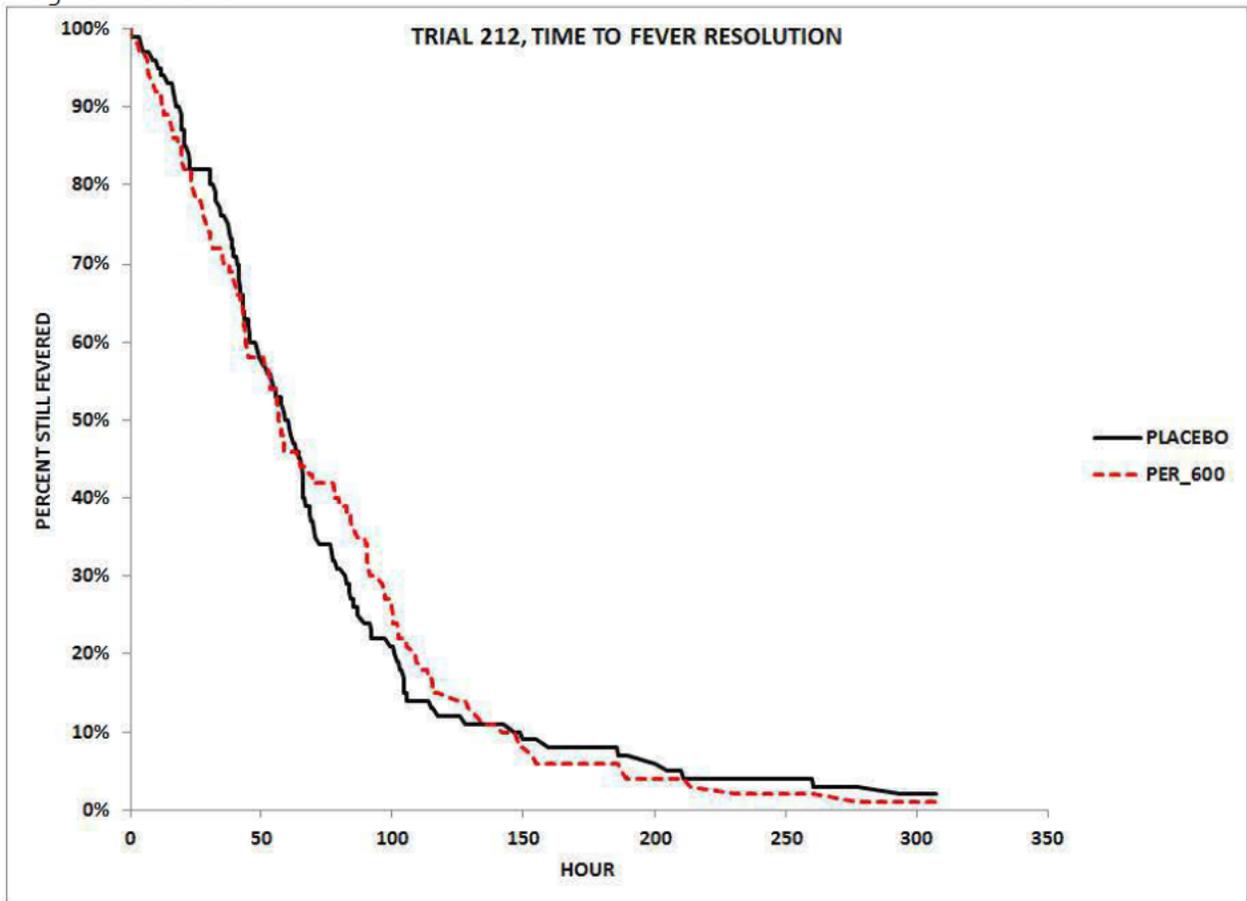


Figure 3.3 J

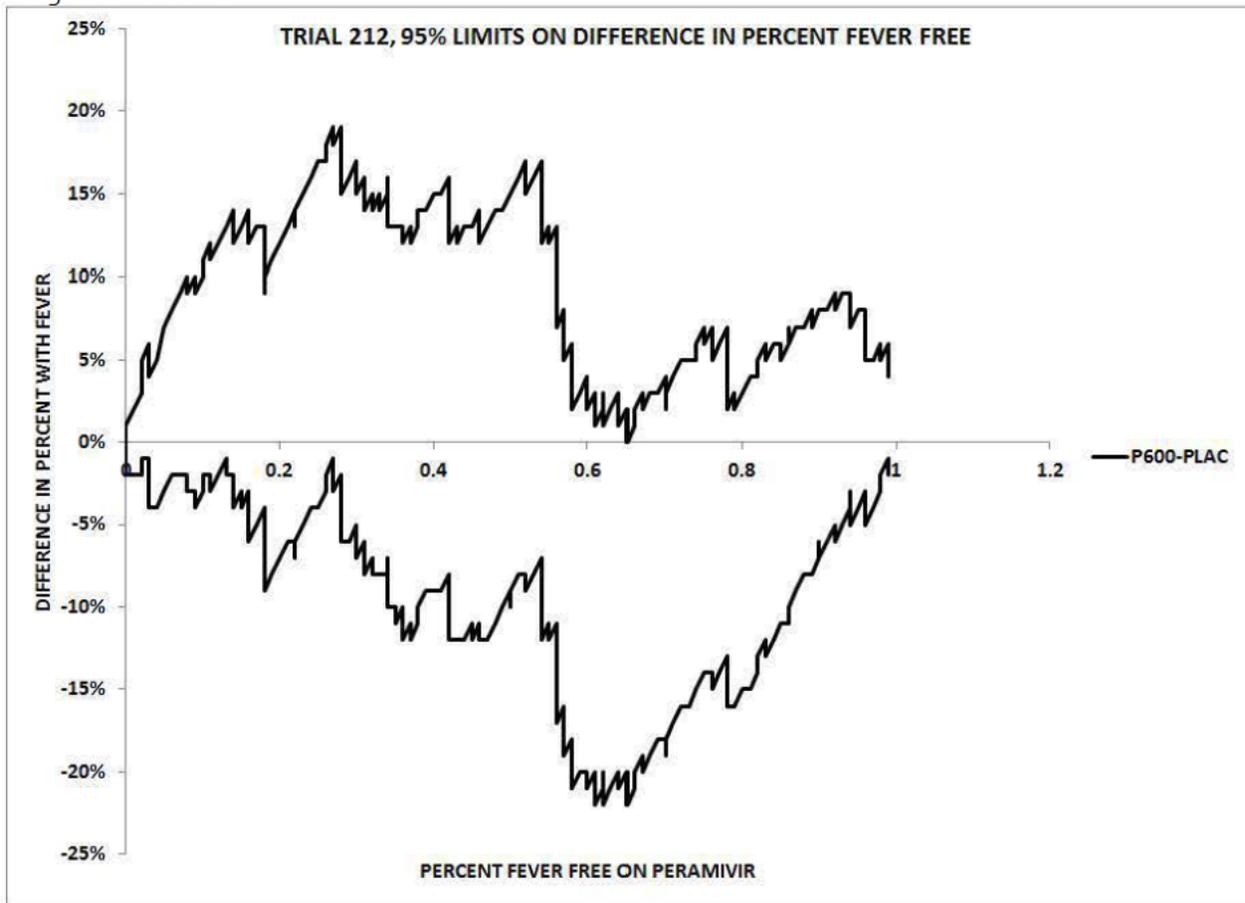


Table 3.3 B gives the median times to fever resolution and their 95% confidence intervals, as read off the Kaplan-Meier plots.

TABLES 3.3. B
 MEDIAN TIMES TO FEVER RESOLUTION
 WITH 95% CONFIDENCE INTERVALS

FEVER		
TRIAL 722		
PLACEBO	PERAMIVIR 300	PERAMIVIR 600
29.3 (24.0, 34.5)	28.3 (21.7, 31.9)	28.4 (22.9, 30.75)
FEVER12		
TRIAL 722		
PLACEBO	PERAMIVIR 300	PERAMIVIR 600
42.4 (32.7, 46.5)	29.2 (25.8, 33.2)	30.2 (25.4, 31.8)
TRIAL 311		
PLACEBO	PERAMIVIR 300	
54.3 (31.3, 77.9)	40.7 (34.6, 44.9)	
TRIAL 211		
PLACEBO	PERAMIVIR 150	PERAMIVIR 300
67.5 (59.8, 83.4)	51.4 (43.5, 60.3)	43.7 (40.25, 52.1)
TRIAL 211_311_POOLED		
PLACEBO	PERAMIVIR 150	PERAMIVIR 300
67.1 (56, 75.9)	52.9 (43.8, 63.6)	42.8 (40.25, 45.6)
TRIAL 212		
PLACEBO	PERAMIVIR 600	
59.7 (49.3, 65.8)	56.7 (45.6, 76.6)	

For the FEVER12 endpoint in trials 722, 311, and 211, there was around a 12 hour improvement in time to fever resolution; for the FEVER endpoint in trial 722 and FEVER12 in trial 212, there was no noticeable improvement.

In summary, there is some reduction in time to fever resolution with peramivir (at least with strains of influenza which are Tamiflu and peramivir susceptible). The improvement is less than with symptom alleviation (12 hours instead of about 24). Since fever resolves quite a bit quicker than the suite of all seven symptoms, this reduction in magnitude is not surprising. Overall, fever resolution confirms the conclusion of peramivir efficacy in some strains of influenza A.

3.4 FDA Analysis of Relapses Post Healing

A certain number of subjects reported symptoms that returned to moderate or severe subsequent to their nominal healing, at least as healing was specified in the protocol and calculated by the FDA reviewer. The fact that the applicant's computation deferred healing until all such relapses and yet gave results which differed from those obtained from the FDA's computation by practically inconsequential amounts gives one confidence that the overall efficacy conclusions are robust to concerns about relapses. Nonetheless, the FDA reviewer has included a further exploration of such relapses here.

The following tables examine the instances of relapse in trial 722. This trial has been considered by itself because it is the largest trial and the one with statistically significant results. Relapse is defined as the occurrence of a moderate or severe diary entry for at least one symptom post healing. Table 3.4 A gives the percent of healed subjects, by arm and by symptom, who had at least one moderate or at least one severe entry after their healing.

TABLE 3.4 A
PERCENT OF HEALED SUBJECTS WITH
SYMPTOM_EVER_MODERATE/SEVERE_POST_HEAL

	PLAC	PER_300	PER_600
EVER MODERATE			
COUGH	12.8%	8.6%	15.1%
SORE THROAT	5.3%	4.3%	10.8%
HEADACHE	7.4%	2.2%	2.2%
NASAL CONGESTION	4.3%	4.3%	8.6%
FEELING FEVERISH	3.2%	0.0%	2.2%
ACHES AND PAINS	3.2%	1.1%	2.2%
FATIGUE	6.4%	3.2%	3.2%
EVER SEVERE			
COUGH	0.0%	1.1%	0.0%
SORE THROAT	0.0%	1.1%	1.1%
HEADACHE	0.0%	0.0%	1.1%
NASAL CONGESTION	0.0%	1.1%	0.0%
FEELING FEVERISH	1.1%	0.0%	1.1%
ACHES AND PAINS	2.1%	0.0%	1.1%
FATIGUE	0.0%	0.0%	0.0%

Not surprisingly, cough is the symptom which most often recurs. In all three arms, cough had the highest rate of relapse to at least moderate. Other symptoms varied in high rates of relapse to at least moderate. Headache and fatigue recurred more commonly on placebo, sore throat and nasal congestion on 600 mg peramivir. Aside from cough, no conspicuous relapses occurred on 300 mg peramivir. The overall lack of pattern suggests that cough is the only commonly recurring symptom. The bottom half of table 3.4 A shows that nothing relapses to severe except in isolated cases.

Table 3.4 B shows the same percentages of relapsers as in table 3.4 A but separated by smoking status. (Given the rarity of relapse to severe, only the relapse to at least moderate is shown.) Here it is surprising that smoking status does not particularly explain the high relapse rate on cough. Smokers resume coughing than non-smokers a lot more on peramivir 300 mg, a little bit more on placebo, and a lot less on peramivir 600 mg. For placebo and 300 mg placebo, smoking is somewhat correlated with higher relapse rates but this correlation breaks down for 600 mg peramivir. This leads one to be skeptical about the strength of the smoking/relapse association.

TABLE 3.4 B
 PERCENT OF HEALED SUBJECTS WITH SYMPTOM_EVER_MODERATE/SEVERE_POST_HEAL
 BY SMOKING STATUS

EVER MODERATE	SMOKER	PLAC	PER_300	PER_600
COUGH	NO	11.7%	4.9%	23.0%
	YES	14.7%	15.6%	0.0%
SORE THROAT	NO	6.7%	4.9%	13.1%
	YES	2.9%	3.1%	6.3%
HEADACHE	NO	3.3%	0.0%	1.6%
	YES	14.7%	6.3%	3.1%
NASAL CONGESTION	NO	1.7%	4.9%	8.2%
	YES	8.8%	3.1%	9.4%
FEELING FEVERISH	NO	3.3%	0.0%	1.6%
	YES	2.9%	0.0%	3.1%
ACHES AND PAINS	NO	3.3%	1.6%	1.6%
	YES	2.9%	0.0%	3.1%
FATIGUE	NO	3.3%	4.9%	3.3%
	YES	11.8%	0.0%	3.1%

A further question one could ask after identifying subjects with at least one moderate or severe symptom post healing is whether those relapses are at isolated diary entries or persist for a while. Table 3.4 C gives the fraction of post healing diary entries which had at least a moderate or a severe score for each given symptom. The percentages in the table are the average fraction of high diary entries, averaged only over subjects who had at least one high diary entry. The fractions would obviously much lower if subjects who had zero high diary entries post healing. The table also includes the number of subjects who had at least one high diary entry.

TABLE 3.4 C
 AMONG SUBJECTS EVER MODERATE OR SEVERE POST HEALING
 ON GIVEN SYMPTOM

FRACTION_OF_POST_HEAL_ENTRIES_WITH_HIGH_SYMPTOM	MEAN FRACTION HIGH ENTRIES			NUMBER OF SUBJECTS EVER HIGH		
	PLAC	PER_300	PER_600	NPLAC	N300	N600
AT LEAST MODERATE						
COUGH	16.3%	25.6%	22.0%	12	8	14
SORE THROAT	12.8%	35.8%	9.4%	5	4	10
HEADACHE	24.2%	8.9%	7.6%	7	2	2
NASAL CONGESTION	13.4%	15.5%	14.5%	4	4	8
FEELING FEVERISH	12.2%	5.1%	3	2		
ACHES AND PAINS	29.2%	6.3%	5.1%	3	1	2
FATIGUE	20.8%	10.0%	6.0%	6	3	3
SEVERE						
COUGH		22.2%			1	
SORE THROAT		22.2%	5.3%		1	1
HEADACHE			5.3%			1
NASAL CONGESTION		5.6%			1	
FEELING FEVERISH	14.3%		5.3%	1		1
ACHES AND PAINS	10.3%		5.3%	2		1

The fraction of post-healing entries with high symptom ratings are all <30% of the subject's total post-healing entries, indicating that relapses are either short-lived or sporadic. The subjects in the placebo arm tended to have slightly higher fraction of high entries across all symptoms. The peramivir arms did have more high entries with respect to cough and sore throat. Also cough does seem to persist once it recurs in all three arms to a somewhat greater extent than other symptoms. Relapse to severe occurs so rarely that nothing can be said about it.

Table 3.4 D looks at the persistence of relapse in a different way by giving the average wait from healing to the first time and last time that a moderate/severe diary entry occurs. These are averaged only over those subjects who actually have a relapse for a given symptom. The number of subjects with at least one moderate entry for each symptom is also given in the table.

TABLE 3.4 D

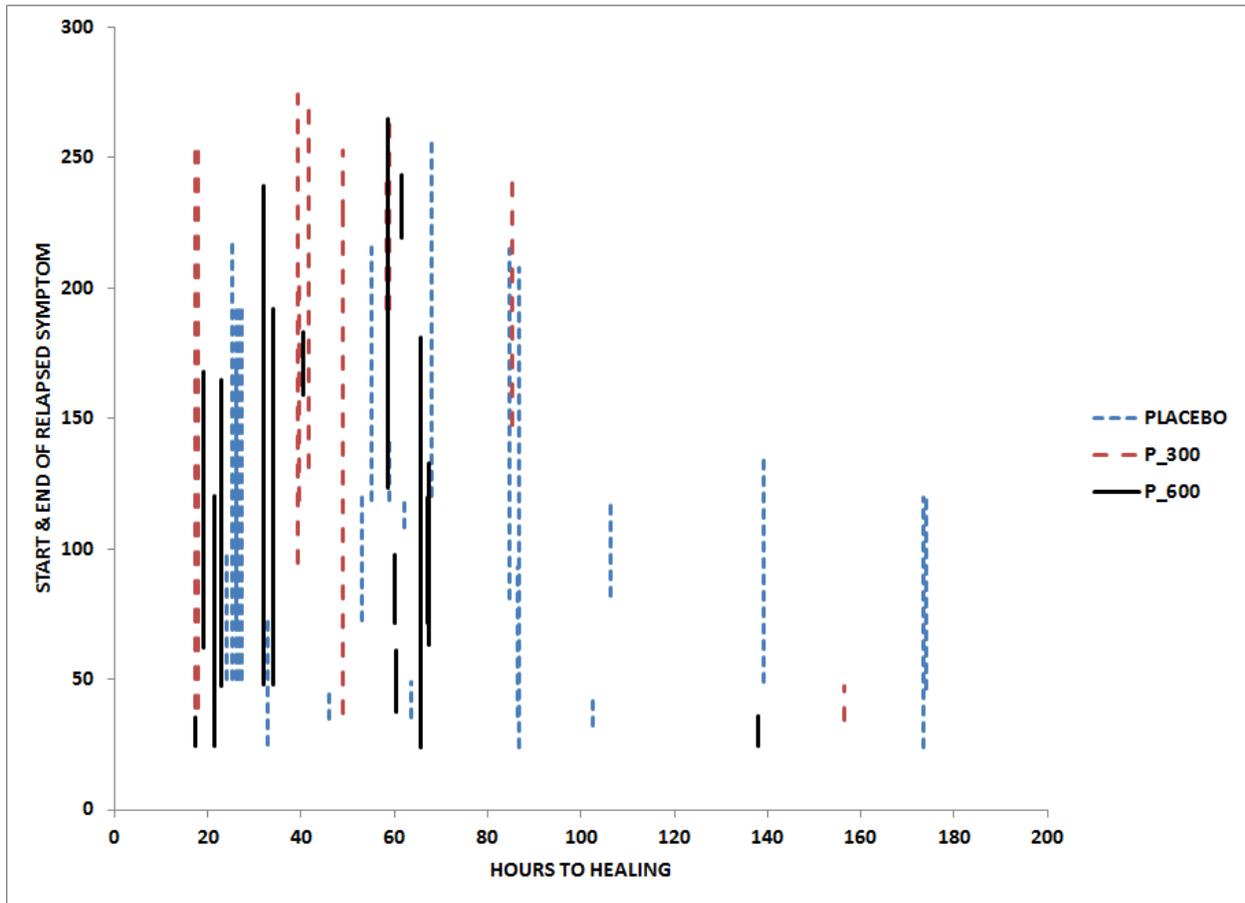
HOURS POST HEALING UNTIL FIRST, LAST HIGH SYMPTOM ENTRY

	PLACEBO			PERAM_300			PERAM_600		
	#	1ST_TIME	LAST_TIME	#	1ST_TIME	LAST_TIME	#	1ST_TIME	LAST_TIME
AT LEAST MODERATE									
COUGH	12	72	100.3	8	120.3	188.8	14	72.6	125.9
SORE THROAT	5	106.9	134.2	4	90.3	188.4	10	87.5	94.5
HEADACHE	7	80.7	143.3	2	37.6	37.6	2	101.5	160.1
NASAL CONGESTION	4	105.1	147.1	4	74.7	138	8	96.9	143
FEELING FEVERISH	3	103.9	119.5	2	101.5	101.5			
ACHES AND PAINS	3	72.3	165.6	1	205.8	205.8	2	101.5	101.5
FATIGUE	6	57.7	125.8	3	143.3	172.7	3	78.3	78.3
SEVERE									
COUGH				1	180.2	250.2			
SORE THROAT				1	131.2	250.2	1	155.5	155.5
HEADACHE							1	155.5	155.5
NASAL CONGESTION				1	228.7	228.7			
FEELING FEVERISH	1	46.5	46.5				1	155.5	155.5
ACHES AND PAINS	2	127.7	127.7				1	155.5	155.5

It is difficult to see much pattern here. Cough recurs as early as 72 hours after healing in placebo and 600 mg peramivir but not until 120 hours post healing on 300 mg peramivir. The last cough recurrence was, on average, 100 hours post healing on placebo versus 126 to 188 hours post healing on peramivir. (Of course, healing is earlier on peramivir.) No obvious conclusions can be drawn from such patterns. The samples are quite small, particularly for the six non-cough symptoms so the absence of any striking pattern is not surprising.

The following graph (figure 3.4 A) is an attempt to display the relationship between relapsed symptoms, time of healing and treatment arm. For every subject who had a relapse to at least moderate, for every symptom with a relapse, there is a vertical line showing the time from first entry with relapse to last entry with relapse, measured in hours since healing. The horizontal coordinate of each line is the time of healing. If two symptoms occur in the same subject (or in two subjects with the same time of healing), the horizontal coordinate is jittered slightly to make both symptoms visible. One will notice a couple of vertical lines with thicker marks, corresponding of this occurrence.

Figure 3.4 A



There doesn't seem to be anything noticeable pattern here either. The one thing one can see is that the solid black lines for peramivir 600 and the dashed red lines for peramivir 300 are further to the left (earlier healing) than the dotted blue lines for placebo. Of course, one knew already that peramivir healing times are earlier than placebo healing times.

An overall summary is that 8-15% of subjects will have a few episodes of moderate (rarely, severe) cough during the week or so following healing. Other symptoms may rebound somewhat less often. Smokers are at a bit higher risk of a relapse. One should also observe that recurrence of one or two symptoms 50 or 100 hours after healing might not be recurrence of influenza but rather an incidence of some other, unrelated cause.

4. Results in Special Populations

There are some difficulties with analyzing the data by subgroup. Two trials, 722 and 815 both conducted by Shionogi, are large enough that one can sub-divide the results by levels of baseline covariates and still get meaningful results. The three trials conducted by BioCryst, 211, 311, and 212 are so small that subdividing the results by levels of a baseline covariate produces estimated effects with uncertainty too large to permit meaningful comments.

Consequently, the FDA statistical reviewer has conducted the analyses by sub-groups several ways. First, all trials were pooled together and stratified Cox proportional hazards regressions were run at each level of interesting baseline covariates.

Four analyses were conducted with different choices of stratifying variable. In the first analysis, the stratifying variable was study. (In detail, this would mean that, for example, a Cox regression is run for males in each study and then the weighted average of the results is used as the final result for males. This repeated for females and for each level of each interesting baseline covariate.) Because there is no placebo arm in trial 815, no results from this trial appear in the results of this analysis.

Second, all trials were pooled together and stratified Cox regressions were run using group rather than study for stratification. Here group is defined as Shionogi (trial 722) or BioCryst (other 3 trials). The peramivir arms of both trials 722 and 815 are all pooled together and compared to the placebo arm of trial 722 when using this stratification. This accounts for the slight, but noticeable, downward shift between the first and second forest plots below.

Third, Cox regressions by subgroup were run in trial 722 alone. Fourth, Cox regressions by subgroup were run in the pooled three BioCryst studies, stratifying by study.

In all these analyses, all peramivir arms were treated as the same, despite the variability in dose from 150mg IM to 600 mg IV. Many of the results above suggest that there is little dose response variation in the range 150 to 600 mg and breaking down results by dose yields even smaller subgroups and even more highly variable estimates.

The overall results may best be summarized by forest plots which show the 95% confidence limits for the log hazard ratio of peramivir to placebo for each covariate level. The following five graphs give the forest results in the order: all trials pooled and regressions stratified by study, all trials pooled and regressions stratified by group, trial 722 alone, three BioCryst trials stratified by study.

The covariates explored are age (divided by the quartiles of all studies pooled), sex, race, influenza subtype, duration of illness prior to treatment, baseline total symptom score (CSS), country, and smoking status. Some subgroups have been deleted because the estimates were too variable and the confidence interval too wide to be useful. Figure 4 A

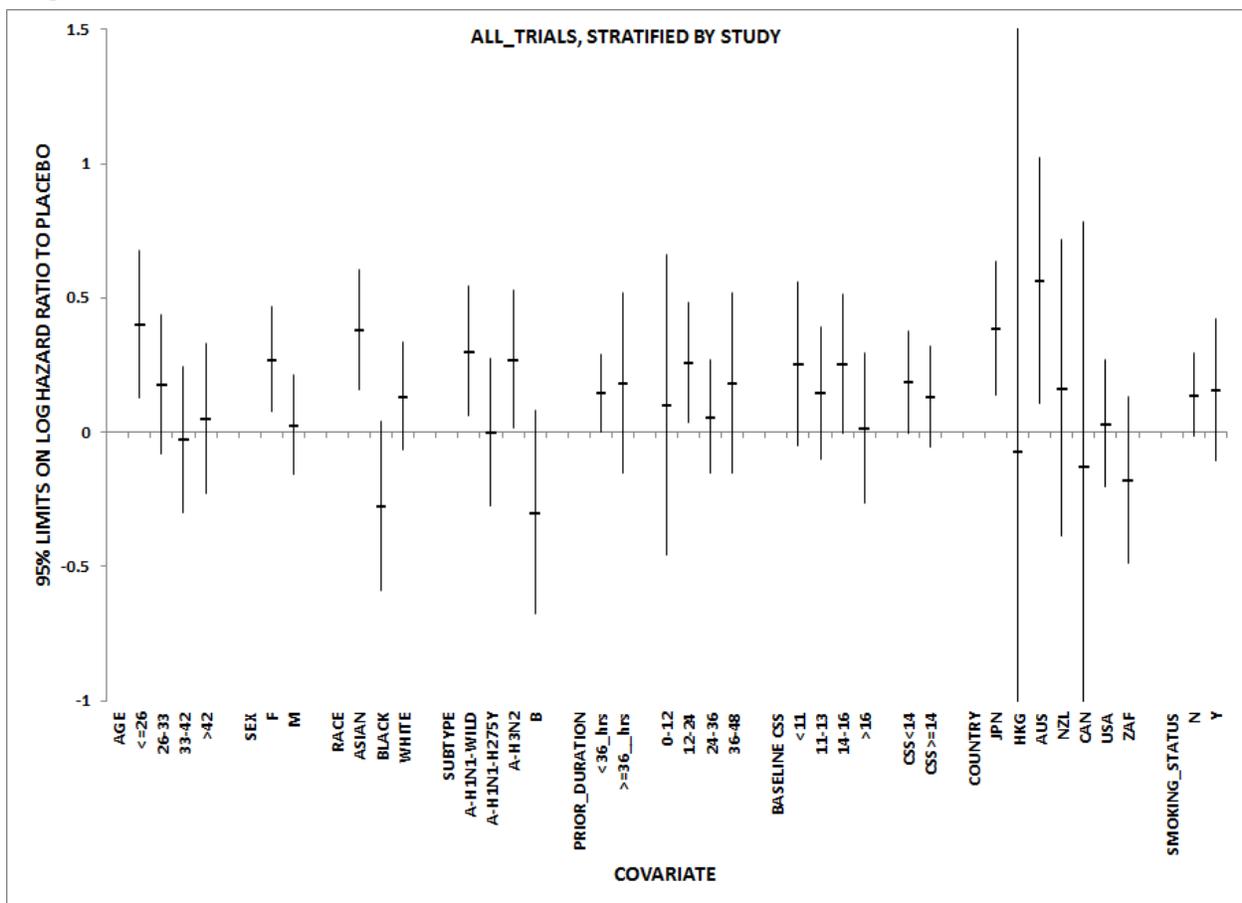


Figure 4 B

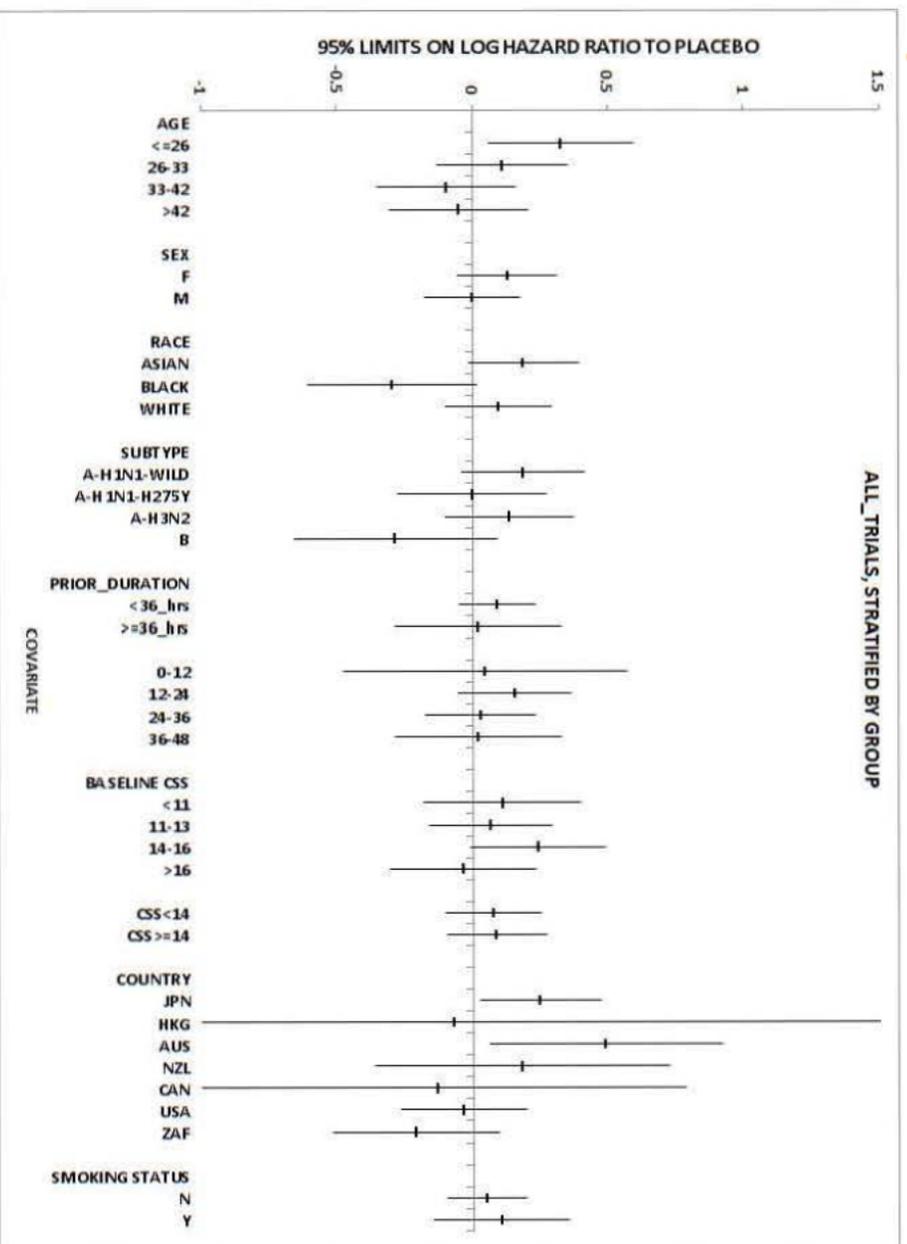


Figure 4 C

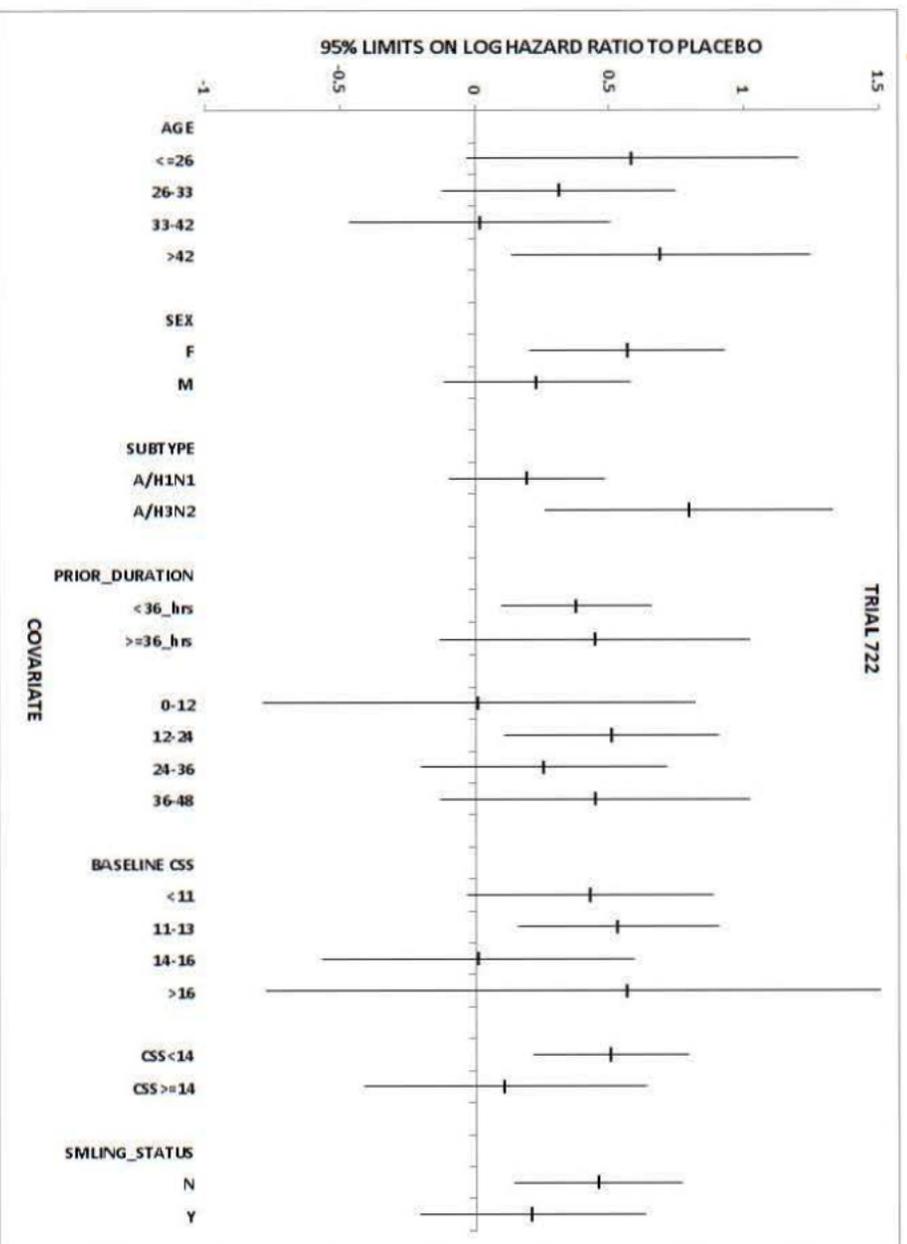
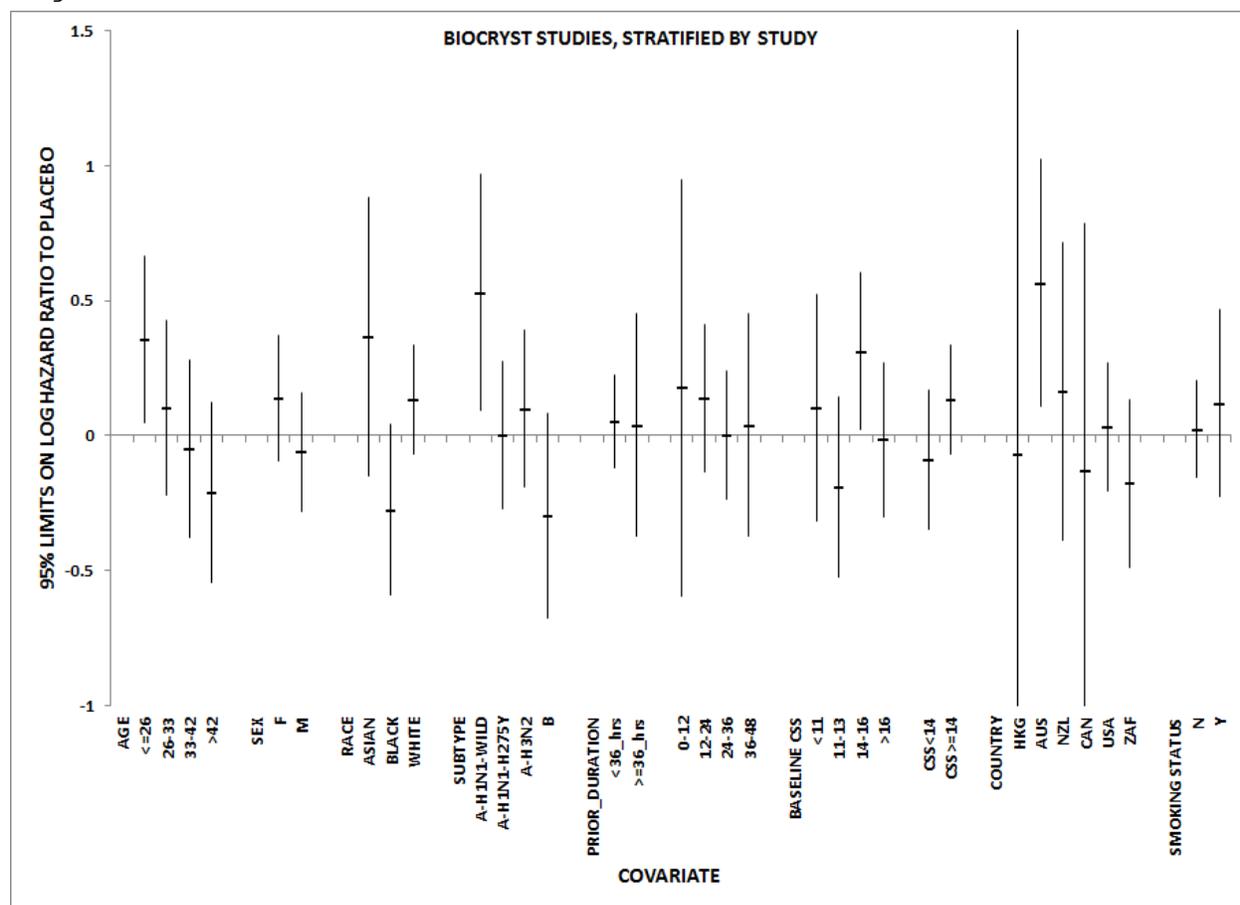


Figure 4 D



There are three subgroups noticeable within these plots that look to be poor performers: influenza type B, influenza type A with H275Y substitution, and Black race. The first two results are expected; it is unclear why Blacks should do worse. Two other patterns may also be seen: peramivir's performance relative to placebo declines with increasing age and does better in Asia than in the rest of the world. (Countries in this plot are in the order Asia, Oceania, America, South Africa from right to left.) As has been remarked above, Asians tend to seek treatment earlier and with lower baseline symptom scores than Americans and Africans so this effect of country is a surrogate for those effects.

4.1 Gender, Race, and Age

The following tables give the numeric results of corresponding to the above plots. Tables 4.1 A-D give the point estimates and 95% confidence limits of log hazard ratios, peramivir to placebo, in each covariate subgroup, along with the sample size in that subgroup when the trials used in each of the four analyses are pooled.

TABLE 4.1 A
LOG HAZARD RATIOS BY AGE, SEX, RACE, COUNTRY
ALL TRIALS, STRATIFIED BY STUDY

COVARIATE	LOG HAZARD		N
	RATIO	95% LIMITS	
AGE_STRATIFIED_BY_STUDY			
≤26	0.404	(0.129, 0.679)	427
26-33	0.180	(-0.080, 0.439)	465
33-42	-0.026	(-0.298, 0.245)	418
>42	0.050	(-0.231, 0.332)	450
AGE_STRATIFIED_BY_STUDY			
18-64	0.156	(0.022, 0.291)	1728
65-74	0.481	(-1.945, 2.908)	25
SEX			
F	0.271	(0.075, 0.466)	870
M	0.027	(-0.159, 0.213)	890
RACE			
ASIAN	0.382	(0.158, 0.607)	1099
BLACK	-0.274	(-0.589, 0.041)	193
WHITE	0.134	(-0.066, 0.335)	468
ETHNICITY			
Hisp	-0.451	(-1.269, 0.367)	34
Not_Hisp	0.289	(0.109, 0.469)	1325
Unk	-0.004	(-0.214, 0.206)	401
COUNTRY_STRATIFIED_BY_STUDY			
AUS	0.565	(0.107, 1.024)	104
CAN	-0.129	(-1.043, 0.785)	25
HKG	-0.069	(-2.067, 1.928)	5
JPN	0.386	(0.137, 0.635)	793
NZL	0.165	(-0.388, 0.718)	61
USA	0.031	(-0.205, 0.268)	341
ZAF	-0.177	(-0.488, 0.135)	197

TABLE 4.1 B
LOG HAZARD RATIOS BY AGE, SEX, RACE, COUNTRY
ALL TRIALS, STRATIFIED BY GROUP

AGE STRATIFIED_BY_GROUP			
<=26	0.321	(0.055, 0.588)	427
26-33	0.069	(-0.169, 0.307)	465
33-42	-0.096	(-0.348, 0.157)	418
>42	-0.041	(-0.294, 0.212)	450
AGE STRATIFIED_BY_GROUP			
18-64	0.063	(-0.064, 0.189)	1728
65-74	-0.573	(-2.192, 1.046)	25
SEX			
F	0.100	(-0.083, 0.283)	870
M	0.004	(-0.168, 0.177)	890
RACE			
ASIAN	0.153	(-0.046, 0.351)	1099
BLACK	-0.296	(-0.607, 0.015)	193
WHITE	0.096	(-0.101, 0.293)	468
ETHNICITY			
Hisp	-0.405	(-1.213, 0.403)	34
Not_Hisp	0.147	(-0.018, 0.313)	1325
Unk	-0.019	(-0.227, 0.188)	401
COUNTRY_STRATIFIED_BY_GROUP			
AUS	0.491	(0.061, 0.921)	104
CAN	-0.129	(-1.043, 0.785)	25
HKG	-0.069	(-2.067, 1.928)	5
JPN	0.211	(-0.005, 0.428)	793
NZL	0.182	(-0.361, 0.725)	61
USA	-0.033	(-0.263, 0.197)	341
ZAF	-0.211	(-0.517, 0.096)	197

TABLE 4.1 C
LOG HAZARD RATIOS BY AGE, SEX, TRIAL 722

AGE TRIAL_722			
<=26	0.586	(-0.029, 1.202)	63
26-33	0.315	(-0.118, 0.749)	89
33-42	0.021	(-0.462, 0.505)	75
>42	0.693	(0.138, 1.249)	70
SEX			
F	0.568	(0.207, 0.930)	146
M	0.232	(-0.113, 0.577)	151

TABLE 4.1 D
 LOG HAZARD RATIOS BY AGE, SEX, RACE, COUNTRY
 BIOCRYST TRIALS, STRATIFIED BY STUDY

AGE_BIOCRYST_TRIALS			
<=26	0.356	(0.046, 0.666)	202
26-33	0.103	(-0.220, 0.426)	177
33-42	-0.048	(-0.377, 0.280)	175
>42	-0.210	(-0.543, 0.123)	181
AGE_BIOCRYST_TRIALS			
18-64	0.057	(-0.103, 0.217)	719
65-74	0.481	(-1.945, 2.908)	11
SEX			
F	0.140	(-0.093, 0.373)	375
M	-0.061	(-0.284, 0.162)	360
RACE			
ASIAN	0.367	(-0.149, 0.883)	74
BLACK	-0.274	(-0.589, 0.041)	193
WHITE	0.134	(-0.066, 0.335)	468
ETHNICITY			
Hisp	-0.451	(-1.269, 0.367)	34
Not_Hisp	0.180	(-0.080, 0.440)	300
Unk	-0.004	(-0.214, 0.206)	401
COUNTRY_BIOCRYST_TRIALS			
AUS	0.565	(0.107, 1.024)	104
CAN	-0.129	(-1.043, 0.785)	25
HKG	-0.069	(-2.067, 1.928)	5
NZL	0.165	(-0.388, 0.718)	61
USA	0.031	(-0.205, 0.268)	341
ZAF	-0.177	(-0.488, 0.135)	197

Tables 4.1 E-H give the median times to symptom alleviation in each arm for E) all studies pooled, stratified by study, F) all studies pooled stratified by group, G) trial 722, and H) BioCryst studies stratified by study. The stratified median in any subgroup-arm is the weighted average of the medians in each stratum of the subgroup-arm with the sample size in the stratum subgroup being the weight. The medians are computed treating all times as if they were observed. The stratum sub-groups are too small to permit computing the medians from Kaplan-Meier curves in each stratum subgroup-arm. The tables also give the sample size in each subgroup-arm. As elsewhere in this section, all peramivir doses are treated as the same. These tables include some subgroups which were too small to permit computation of hazard ratios.

TABLE 4.1 E
 MEDIAN, SAMPLE SIZES IN EACH SUBGROUP
 ALL TRIALS, STRATIFIED BY STUDY
 AGE, SEX, RACE, COUNTRY

COVARIATE	PERAMIVIR		PLACEBO	
	MEDIAN	N	MEDIAN	N
AGE STRATIFIED BY STUDY				
≤26	80.0	331	114.2	95
26-33	82.6	354	99.3	108
33-42	84.2	315	94.4	101
>42	67.0	350	89.1	100
AGE STRATIFIED BY STUDY				
12-17	79.7	1	.	.
18-64	78.4	1320	100.3	402
65-74	79.1	23	103.3	2
≥75	90.9	6	.	.
SEX				
F	90.9	667	116.4	200
M	71.6	683	86.8	204
RACE				
ASIAN	70.7	963	99.9	135
BLACK	103.3	102	79.8	87
WHITE	93.8	285	111.0	182
ETHNICITY				
Hispanic	134.3	22	77.2	12
Not_Hispanic	75.2	1121	104.8	202
Unk	88.4	207	96.6	190
COUNTRY STRATIFIED BY STUDY				
AUS	84.9	63	136.6	41
CAN	110.8	18	112.5	7
GBR	.	.	222.5	2
HKG	191.5	2	121.3	3
JPN	68.2	693	81.8	100
KOR	52.6	70	.	.
NZL	138.0	35	151.9	26
TWN	82.8	162	.	.
USA	93.0	199	97.8	140
ZAF	103.4	108	100.7	85

TABLE 4.1 F
MEDIANS, SAMPLE SIZES IN EACH SUBGROUP
ALL TRIALS, STRATIFIED BY GROUP
AGE, SEX, RACE, COUNTRY

COVARIATE	PERAMIVIR		PLACEBO	
	MEDIAN	N	MEDIAN	N
AGE STRATIFIED BY GROUP				
<=26	75.4	331	115.3	95
26-33	81.2	354	103.3	108
33-42	83.1	315	98.7	101
>42	66.9	350	89.1	100
AGE STRATIFIED BY GROUP				
I2_17	79.7	1	.	.
18-64	76.8	1320	103.6	402
65-74	84.7	23	103.3	2
>=75	120.3	6	.	.
SEX				
F	87.4	667	119.2	200
M	69.2	683	87.4	204
RACE				
ASIAN	70.2	963	90.6	135
BLACK	101.8	102	71.4	87
WHITE	92.1	285	113.8	182
ETHNICITY				
Hisp	133.1	22	80.0	12
Not_Hisp	74.9	1121	104.8	202
Unk	88.3	207	93.1	190
COUNTRY STRATIFIED BY GROUP				
AUS	90.0	63	139.3	41
CAN	110.8	18	112.5	7
GBR	.	.	222.5	2
HKG	191.5	2	121.3	3
JPN	68.1	693	81.8	100
KOR	52.6	70	.	.
NZL	139.1	35	140.4	26
TWN	82.8	162	.	.
USA	89.0	199	93.7	140
ZAF	103.5	108	101.3	85

TABLE 4.1 G
MEDIANS, SAMPLE SIZES IN EACH SUBGROUP,
TRIAL 722, AGE, SEX

COVARIATE	PERAMIVIR		PLACEBO	
	MEDIAN	N	MEDIAN	N
AGE TRIAL_722				
<=26	55.2	49	88.9	14
26-33	58.9	52	81.9	37
33-42	62.2	48	67.5	27
>42	54.2	48	85.5	22
SEX				
F	57.1	97	89.8	49
M	58.8	100	70.6	51

TABLE 4.1 H
 MEDIAN, SAMPLE SIZES IN EACH SUBGROUP
 BIOCRYST TRIALS, STRATIFIED BY STUDY
 AGE, SEX, RACE, COUNTRY

COVARIATE	PERAMIVIR		PLACEBO	
	MEDIAN	N	MEDIAN	N
AGE_BIOCRYST_TRIALS				
<=26	88.9	120	118.6	81
26-33	99.8	103	108.3	71
33-42	112.6	99	104.2	74
>42	91.8	103	90.1	78
AGE_BIOCRYST_TRIALS				
12_17	79.7	1	.	.
18-64	97.4	411	106.5	302
65-74	125.0	9	103.3	2
>=75	119.7	4	.	.
SEX				
F	106.6	221	125.0	151
M	88.6	204	92.3	153
RACE				
ASIAN	114.7	38	151.6	35
BLACK	103.3	102	79.8	87
WHITE	93.8	285	111.0	182
ETHNICITY				
Hisp	134.3	22	77.2	12
Not_Hisp	105.1	196	127.3	102
Unk	88.4	207	96.6	190
COUNTRY_BIOCRYST_TRIALS				
AUS	84.9	63	136.6	41
CAN	110.8	18	112.5	7
GBR	.	.	222.5	2
HKG	191.5	2	121.3	3
NZL	138.0	35	151.9	26
USA	93.0	199	97.8	140
ZAF	103.4	108	100.7	85

4.2 Other Covariates

Tables 4.2 A-D give the point estimates and 95% confidence limits of log hazard ratio, peramivir to placebo, for subgroups defined by influenza subtype, prior duration of illness, baseline symptom score, and smoking status.

TABLE 4.2 A
LOG HAZARD RATIOS BY SUBTYPE, PRIOR DURATION,
BASELINE SEVERITY, COUNTRY, SMOKING STATUS
ALL TRIALS, STRATIFIED BY STUDY

COVARIATE	LOG HAZARD RATIO	95% LIMITS	N
SUBTYPE STRATIFIED BY STUDY			
A-H1N1-WILD	0.302	(0.061, 0.543)	733
A-H1N1-H275Y	0.001	(-0.273, 0.275)	230
A-H3N2	0.271	(0.015, 0.527)	535
B	-0.299	(-0.679, 0.081)	180
Indeter	0.655	(-0.627, 1.938)	82
SUBTYPE BY STUDY			
A	0.216	(0.072, 0.359)	1551
B	-0.299	(-0.679, 0.081)	180
SUBTYPE STRATIFIED BY STUDY			
A/H1N1	0.302	(0.061, 0.543)	733
A/H3N2	0.271	(0.015, 0.527)	535
A/Indet	1.180	(-0.207, 2.568)	53
B	-0.380	(-0.763, 0.004)	177
A/H1N1-H275Y	0.001	(-0.273, 0.275)	230
PRIOR DURATION OF ILLNESS STRATIFIED BY STUDY			
<36 hrs	0.145	(-0.001, 0.292)	1363
>=36 hrs	0.183	(-0.155, 0.521)	397
PRIOR DURATION OF ILLNESS STRATIFIED BY STUDY			
0-12	0.101	(-0.459, 0.660)	129
12-24	0.260	(0.038, 0.483)	625
24-36	0.058	(-0.153, 0.269)	609
36-48	0.183	(-0.155, 0.521)	396
BASELINE SYMPTOM TOTAL STRATIFIED BY STUDY			
<11	0.256	(-0.051, 0.562)	419
11-13	0.146	(-0.103, 0.395)	543
14-16	0.254	(-0.007, 0.514)	444
>16	0.014	(-0.266, 0.294)	354
BASELINE SYMPTOM TOTAL STRATIFIED BY STUDY			
CSS<14	0.188	(-0.004, 0.379)	956
CSS>=14	0.131	(-0.058, 0.320)	798
SMOKING STATUS STRATIFIED BY STUDY			
N	0.138	(-0.017, 0.293)	1281
Y	0.158	(-0.109, 0.425)	479

TABLE 4.2 B
 LOG HAZARD RATIOS BY SUBTYPE, PRIOR DURATION,
 BASELINE SEVERITY, COUNTRY, SMOKING STATUS
 ALL TRIALS, STRATIFIED BY GROUP

SUBTYPE STRATIFIED BY GROUP			
A-H1N1-WILD	0.144	(-0.078, 0.366)	733
A-H1N1-H275Y	0.001	(-0.273, 0.275)	230
A-H3N2	0.128	(-0.107, 0.363)	535
B	-0.283	(-0.659, 0.092)	180
Indeter	0.957	(-0.073, 1.988)	82
SUBTYPE STRATIFIED BY GROUP			
A	0.093	(-0.041, 0.228)	1551
B	-0.283	(-0.659, 0.092)	180
SUBTYPE STRATIFIED BY GROUP			
A/H1N1	0.144	(-0.078, 0.366)	733
A/H3N2	0.128	(-0.107, 0.363)	535
A/Indet	1.030	(-0.048, 2.108)	53
B	-0.363	(-0.741, 0.016)	177
A/H1N1-H275Y	0.001	(-0.273, 0.275)	230
PRIOR DURATION OF ILLNESS STRATIFIED BY GROUP			
<36 hrs	0.068	(-0.071, 0.206)	1363
>=36 hrs	0.046	(-0.256, 0.349)	397
PRIOR DURATION OF ILLNESS STRATIFIED BY GROUP			
0-12	0.000	(-0.525, 0.525)	129
12-24	0.123	(-0.084, 0.331)	625
24-36	0.019	(-0.182, 0.219)	609
36-48	0.044	(-0.258, 0.347)	396
BASELINE SYMPTOM TOTAL STRATIFIED BY GROUP			
<11	0.117	(-0.167, 0.401)	419
11-13	0.005	(-0.216, 0.227)	543
14-16	0.225	(-0.024, 0.474)	444
>16	-0.028	(-0.294, 0.238)	354
BASELINE SYMPTOM TOTAL STRATIFIED BY GROUP			
CSS<14	0.047	(-0.127, 0.220)	956
CSS>=14	0.089	(-0.091, 0.270)	798
SMOKING STATUS STRATIFIED BY GROUP			
N	0.052	(-0.094, 0.197)	1281
Y	0.060	(-0.189, 0.309)	479

TABLE 4.2 C
 LOG HAZARD RATIOS BY SUBTYPE, PRIOR DURATION,
 BASELINE SEVERITY, COUNTRY, SMOKING STATUS
 TRIAL 722

SUBTYPE TRIAL_722			
A-H1N1-WILD	0.194	(-0.096, 0.484)	216
A-H3N2	0.796	(0.262, 1.331)	70
Indeter	1.243	(-0.478, 2.963)	8
SUBTYPE TRIAL_722			
A/H1N1	0.194	(-0.096, 0.484)	216
A/H3N2	0.796	(0.262, 1.331)	70
A/Indet	1.243	(-0.478, 2.963)	8
PRIOR_DURATION_OF_ILLNESS TRIAL_722			
<36_hrs	0.378	(0.098, 0.657)	242
>=36_hrs	0.446	(-0.128, 1.020)	55
PRIOR_DURATION_OF_ILLNESS TRIAL_722			
0-12	0.014	(-0.788, 0.815)	35
12-24	0.508	(0.110, 0.906)	124
24-36	0.255	(-0.202, 0.712)	83
36-48	0.446	(-0.128, 1.020)	55
BASELINE_SYMPTOM_TOTAL TRIAL_722			
<11	0.427	(-0.029, 0.883)	96
11-13	0.529	(0.157, 0.902)	133
14-16	0.011	(-0.570, 0.592)	54
>16	0.562	(-0.776, 1.900)	14
BASELINE_SYMPTOM_TOTAL TRIAL_722			
CSS<14	0.504	(0.217, 0.791)	229
CSS>=14	0.111	(-0.415, 0.636)	68
SMOKING_STATUS TRIAL_722			
N	0.456	(0.145, 0.767)	196
Y	0.212	(-0.205, 0.630)	101

TABLE 4.2 D
 LOG HAZARD RATIOS BY SUBTYPE, PRIOR DURATION,
 BASELINE SEVERITY, COUNTRY, SMOKING STATUS
 BIOCRYST TRIALS, STRATIFIED BY STUDY

SUBTYPE BIOCRYST TRIALS			
A-H1N1-WILD	0.531	(0.095, 0.967)	119
A-H1N1-H275Y	0.001	(-0.273, 0.275)	230
A-H3N2	0.101	(-0.190, 0.391)	244
B	-0.299	(-0.679, 0.081)	130
Indeter	-0.128	(-1.939, 1.683)	12
SUBTYPE BIOCRYST TRIALS			
A	0.126	(-0.050, 0.303)	602
B	-0.299	(-0.679, 0.081)	130
SUBTYPE BIOCRYST TRIALS			
A/H1N1	0.531	(0.095, 0.967)	119
A/H3N2	0.101	(-0.190, 0.391)	244
A/Indet	1.063	(-1.263, 3.388)	9
B	-0.380	(-0.763, 0.004)	127
A/H1N1-H275Y	0.001	(-0.273, 0.275)	230
PRIOR DURATION OF ILLNESS BIOCRYST TRIALS			
<36 hrs	0.053	(-0.120, 0.225)	608
>=36 hrs	0.039	(-0.374, 0.453)	127
PRIOR DURATION OF ILLNESS BIOCRYST TRIALS			
0-12	0.179	(-0.593, 0.951)	37
12-24	0.139	(-0.134, 0.411)	254
24-36	0.003	(-0.235, 0.242)	317
36-48	0.039	(-0.374, 0.453)	127
BASELINE SYMPTOM TOTAL BIOCRYST TRIALS			
<11	0.103	(-0.318, 0.524)	106
11-13	-0.189	(-0.524, 0.146)	168
14-16	0.313	(0.022, 0.604)	218
>16	-0.014	(-0.302, 0.273)	243
BASELINE SYMPTOM TOTAL BIOCRYST TRIALS			
CSS<14	-0.090	(-0.350, 0.169)	268
CSS>=14	0.134	(-0.068, 0.336)	461
SMOKING STATUS BIOCRYST TRIALS			
N	0.025	(-0.155, 0.205)	582
Y	0.120	(-0.228, 0.467)	153

Tables 4.2 E-H give the median times to symptom alleviation and sample sizes for each treatment arm for subgroups defined by influenza subtype, prior duration of illness, baseline symptom score, and smoking status.

TABLE 4.2 E
 MEDIAN, SAMPLE SIZES IN EACH SUBGROUP
 ALL TRIALS, STRATIFIED BY STUDY
 SUBTYPE, PRIOR DURATION, BASELINE SYMPTOMS, SMOKING

COVARIATE	PERAMIVIR		PLACEBO	
	MEDIAN	N	MEDIAN	N
SUBTYPE STRATIFIED BY STUDY				
A-H1N1-WILD	75.0	621	112.7	110
A-H1N1-H275Y	85.1	104	90.6	124
A-H3N2	72.9	423	101.7	110
B	100.3	127	110.6	53
Indeter	85.8	75	214.8	7
SUBTYPE STRATIFIED BY STUDY				
.	92.8	29	.	.
A	76.3	1194	99.8	351
B	100.3	127	110.6	53
SUBTYPE STRATIFIED BY STUDY				
A/H1N1	75.0	621	112.7	110
A/H3N2	72.9	423	101.7	110
A/Indet	81.8	46	214.8	7
B	103.5	125	110.4	52
A+B	30.5	2	341.5	1
Indet	92.8	29	.	.
A/H1N1-H275Y	85.1	104	90.6	124
PRIOR DURATION OF ILLNESS STRATIFIED BY STUDY				
<36 hrs	78.6	1016	99.5	341
>=36 hrs	80.1	334	117.0	63
PRIOR DURATION OF ILLNESS STRATIFIED BY STUDY				
0-12	62.6	107	81.5	22
12-24	77.1	464	101.2	156
24-36	84.9	445	101.1	163
36-48	80.2	333	117.0	63
>=48	23.7	1	.	.
BASELINE SYMPTOM TOTAL STRATIFIED BY STUDY				
<11	57.0	342	73.2	71
11-13	81.6	430	95.9	113
14-16	87.1	337	108.2	107
>16	102.2	241	123.5	113
BASELINE SYMPTOM TOTAL STRATIFIED BY STUDY				
CSS<14	71.1	772	82.4	184
CSS>=14	93.3	578	116.6	220
SMOKING STATUS STRATIFIED BY STUDY				
N	85.4	968	102.2	309
Y	64.7	382	92.8	95

TABLE 4.2 F
MEDIANS, SAMPLE SIZES IN EACH SUBGROUP
ALL TRIALS, STRATIFIED BY GROUP
SUBTYPE, PRIOR DURATION, BASELINE SYMPTOMS, SMOKING

COVARIATE	PERAMIVIR		PLACEBO	
	MEDIAN	N	MEDIAN	N
SUBTYPE STRATIFIED BY GROUP				
A-H1N1-WILD	75.0	621	110.6	110
A-H1N1-H275Y	85.1	104	90.6	124
A-H3N2	73.4	423	106.9	110
B	104.1	127	113.3	53
Indeter	82.3	75	199.5	7
SUBTYPE STRATIFIED BY GROUP				
.	97.3	29	.	.
A	75.1	1194	100.1	351
B	104.1	127	113.3	53
SUBTYPE STRATIFIED BY GROUP				
A/H1N1	75.0	621	110.6	110
A/H3N2	73.4	423	106.9	110
A/Indet	78.0	46	199.5	7
B	108.2	125	113.3	52
A+B	30.5	2	341.5	1
Indet	97.3	29	.	.
A/H1N1-H275Y	85.1	104	90.6	124
PRIOR DURATION OF ILLNESS STRATIFIED BY GROUP				
<36 hrs	77.0	1016	100.7	341
>=36 hrs	79.8	334	113.6	63
PRIOR DURATION OF ILLNESS STRATIFIED BY GROUP				
0-12	62.2	107	74.2	22
12-24	73.7	464	103.2	156
24-36	83.8	445	102.0	163
36-48	79.9	333	113.6	63
>=48	23.7	1	.	.
BASELINE SYMPTOM TOTAL STRATIFIED BY GROUP				
<11	56.3	342	71.5	71
11-13	79.4	430	96.5	113
14-16	87.2	337	109.9	107
>16	102.6	241	118.5	113
BASELINE SYMPTOM TOTAL STRATIFIED BY GROUP				
CSS<14	67.8	772	80.6	184
CSS>=14	91.4	578	113.4	220
SMOKING STATUS STRATIFIED BY GROUP				
N	81.5	968	108.3	309
Y	65.4	382	85.9	95

TABLE 4.2 G
 MEDIAN, SAMPLE SIZES IN EACH SUBGROUP, TRIAL 722,
 SUBTYPE, PRIOR DURATION, BASELINE SYMPTOMS, SMOKING

COVARIATE	PERAMIVIR		PLACEBO	
	MEDIAN	N	MEDIAN	N
SUBTYPE TRIAL 722				
A-H1N1-WILD	57.1	144	81.4	72
A-H3N2	56.6	46	77.8	24
B	104.0	3	.	.
Indeter	59.5	4	220.7	4
SUBTYPE TRIAL 722				
A	57.1	194	81.8	100
B	104.0	3	.	.
SUBTYPE TRIAL 722				
A/H1N1	57.1	144	81.4	72
A/H3N2	56.6	46	77.8	24
A/Indet	59.5	4	220.7	4
B	104.0	3	.	.
PRIOR DURATION OF ILLNESS TRIAL 722				
<36 hrs	57.4	164	87.4	78
>=36 hrs	57.1	33	67.3	22
PRIOR DURATION OF ILLNESS TRIAL 722				
0-12	63.4	27	57.3	8
12-24	52.8	84	93.7	40
24-36	67.1	53	87.4	30
36-48	57.1	33	67.3	22
BASELINE SYMPTOM TOTAL TRIAL 722				
<11	54.4	69	67.3	27
11-13	57.4	86	92.7	47
14-16	73.5	32	81.5	22
>16	86.9	10	222.0	4
BASELINE SYMPTOM TOTAL TRIAL 722				
CSS<14	56.1	155	76.0	74
CSS>=14	75.5	42	96.7	26
SMOKING STATUS TRIAL 722				
N	61.3	130	93.1	66
Y	51.4	67	66.4	34

TABLE 4.2 H
MEDIANS, SAMPLE SIZES IN EACH SUBGROUP
BIOCRYST TRIALS, STRATIFIED BY STUDY
SUBTYPE, PRIOR DURATION, BASELINE SYMPTOMS, SMOKING

COVARIATE	PERAMIVIR		PLACEBO	
	MEDIAN	N	MEDIAN	N
SUBTYPE_BIOCRYST_TRIALS				
A-H1N1-WILD	108.1	79	171.9	38
A-H1N1-H275Y	85.1	104	90.6	124
A-H3N2	87.8	156	108.4	86
B	114.5	77	110.6	53
Indeter	150.5	9	206.9	3
SUBTYPE_BIOCRYST_TRIALS				
.	249.3	3	.	.
A	95.2	345	107.0	251
B	114.5	77	110.6	53
SUBTYPE_BIOCRYST_TRIALS				
A/H1N1	108.1	79	171.9	38
A/H3N2	87.8	156	108.4	86
A/Indet	113.1	6	206.9	3
B	120.3	75	110.4	52
A+B	30.5	2	341.5	1
Indet	249.3	3	.	.
A/H1N1-H275Y	85.1	104	90.6	124
PRIOR_DURATION_OF_ILLNESS_BIOCRYST_TRIALS				
<36_hrs	93.7	339	103.1	263
>=36_hrs	118.3	86	143.6	41
PRIOR_DURATION_OF_ILLNESS_BIOCRYST_TRIALS				
0-12	75.8	23	95.3	14
12-24	97.7	133	103.8	116
24-36	96.9	183	104.2	133
36-48	118.3	86	143.6	41
BASELINE_SYMPTOM_TOTAL_BIOCRYST_TRIALS				
<11	64.1	56	76.8	44
11-13	119.3	102	98.2	66
14-16	93.4	133	115.1	85
>16	107.8	134	119.9	109
BASELINE_SYMPTOM_TOTAL_BIOCRYST_TRIALS				
CSS<14	94.4	158	86.7	110
CSS>=14	99.3	267	119.2	194
SMOKING_STATUS_BIOCRYST TRIALS				
N	101.2	335	104.7	243
Y	95.3	90	107.6	61

5. Summary and Conclusions:

The applicant has conducted five trials on the efficacy of peramivir. Four of them were placebo controlled and one Tamiflu controlled. One of the large placebo controlled trials, 722, showed a sufficient superiority of peramivir to placebo to provide statistical evidence equivalent to two studies statistically significant at the conventional .025 level.

None of the other studies achieved formal statistical significance at level .025. Nonetheless, they all provide evidence supporting the conclusion from trial 722. The median times to healing were clinically meaningfully shorter for peramivir than for placebo in all trials: 30 hours shorter for both 300mg and 600mg in trial 722, 8 hours shorter for 150mg and 14 hours shorter for 300mg in the pooled results of trials 211 and 311, 21 hours shorter for 600mg in trial 212 (provided one looked only at the Tamiflu susceptible subjects in trial 212: type A H3N2 and H1N1 Wild).

The dose response pattern seen in the difference in the medians was part of a statistically significant pattern in the log hazard ratios that could be found by comparing the 150, 300, and 600mg responses in all four placebo controlled trials.

Results on time to resolution of fever confirmed results on the primary endpoint of time to healing. The issue of robustness of results to possible relapses has been examined and found not to be a cause for doubt about the primary efficacy conclusion.

The three BioCryst trials also suggest that peramivir does not work in type B influenza. In all three trials, the log hazard ratios of peramivir to placebo were positive for type A and negative for type B, although all the confidence intervals straddled zero. There were almost no type B cases in trial 722 so all the evidence concerning efficacy in type B comes from these three trials.

With respect to baseline covariates, five covariates showed a suggestion of an interaction with treatment. Peramivir's performance relative to placebo declines with increasing age and does better in Asia than in the rest of the world. It also declines with later start of treatment and higher baseline symptom score. Asians tend to seek treatment earlier and with lower baseline symptom scores than

Americans and Africans so this effect of country is a surrogate for those effects. Blacks also did worse on peramivir relative to placebo; if this is anything other than random, the reason for it is unclear.

Finally, there are reasons to doubt whether peramivir will work against Tamiflu resistant strains of type A influenza. In trial 212, positive results for peramivir were obtained in the sub-group with type A, H1N1 wild or H3N2. Results were negative for the sub-group with type B or type A, H1N1 with H275Y substitution. In trial 815, there was no testing for the H275Y substitution but all subjects with type A H1N1 had $IC_{50} > 15$, all subjects with type A H3N2 had $IC_{50} < 5$. One will recall from table 3.2.5 A that the median healing times for all three arms were about 11 hours shorter among H3N2 subjects than among H1N1 or B subjects. Furthermore, the correlation between IC_{50} for Tamiflu and IC_{50} for peramivir was high.

In short summary, peramivir at 600mg is convincingly effective against Tamiflu susceptible strains of influenza A with about a 30 hour reduction in symptom duration. Peramivir may be effective down to doses as low as 150mg against those influenza strains but the reduction in symptom duration is only about 8 hours. Finally, it is unlikely that peramivir is very effective against influenza B or against strains of influenza A that are Tamiflu resistant.

Thomas Hammerstrom, Ph.D.
Mathematical Statistician

Concur: Dr. Soon

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

THOMAS S HAMMERSTROM
08/22/2014

GUOXING SOON
08/22/2014

DIONNE L PRICE
08/22/2014

Concur with the overall conclusion that the 600 mg IV formulation is efficacious.

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

NDA Number: 206426 Applicant: BioCryst Pharmaceutic Stamp Date: 12/23/2013

Drug Name: Peramavir NDA/BLA Type: NDA

On **initial** overview of the NDA/BLA application for RTF:

	Content Parameter	Yes	No	NA	Comments
1	Index is sufficient to locate necessary reports, tables, data, etc.	X			
2	ISS, ISE, and complete study reports are available (including original protocols, subsequent amendments, etc.)	X			
3	Safety and efficacy were investigated for gender, racial, and geriatric subgroups investigated (if applicable).	X			
4	Data sets in EDR are accessible and do they conform to applicable guidances (e.g., existence of define.pdf file for data sets).	X			

IS THE STATISTICAL SECTION OF THE APPLICATION FILEABLE? Yes

If the NDA/BLA is not fileable from the statistical perspective, state the reasons and provide comments to be sent to the Applicant.

Please identify and list any potential review issues to be forwarded to the Applicant for the 74-day letter.

Content Parameter (possible review concerns for 74-day letter)	Yes	No	NA	Comment
Designs utilized are appropriate for the indications requested.	X			
Endpoints and methods of analysis are specified in the protocols/statistical analysis plans.	X			
Interim analyses (if present) were pre-specified in the protocol and appropriate adjustments in significance level made. DSMB meeting minutes and data are available.	X			
Appropriate references for novel statistical methodology (if present) are included.			X	
Safety data organized to permit analyses across clinical trials in the NDA/BLA.			X	
Investigation of effect of dropouts on statistical analyses as described by applicant appears adequate.	X			

STATISTICS FILING CHECKLIST FOR A NEW NDA/BLA

Thomas Hammerstrom 1/27/14

Reviewing Statistician Date

Greg Soon

Supervisor/Team Leader Date

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

THOMAS S HAMMERSTROM
01/27/2014

GUOXING SOON
01/27/2014