CENTER FOR DRUG EVALUATION AND RESEARCH

APPROVAL PACKAGE FOR:

APPLICATION NUMBER

21-335/S-004

Statistical Review(s)

DEPARTMENT OF HEALTH AND HUMAN SERVICES
PUBLIC HEALTH SERVICE
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH

STATISTICAL REVIEW AND EVALUATION

Medical Division: Oncology Drug Products (HFD-150)
Biometrics Division: Division of Biometrics I (HFD-710)

STATISTICAL KEY WORDS:	**Active control trials; Progression-free survival; major cytogenetic response; accelerated approval**
NDA NUMBER:	**21-335**
SERIAL NUMBER:	**000**
DATE RECEIVED BY CENTER:	**6/28/02**
DRUG NAME:	**Gleevec (STI571; Imatinib mesylate)**
INDICATION:	**First-line Philadelphia chromosome positive (Ph+) chronic myelogenous leukemia (CML) in chronic phase**
SPONSOR:	**Novartis**
DOCUMENTS REVIEWED:	**Volumes 2, 11, 12 and 33**
STATISTICAL REVIEWER:	**Mark Rothmann, Ph.D. (HFD-710)**
STATISTICAL TEAM LEADER:	**Gang Chen, Ph.D. (HFD-710)**
CLINICAL REVIEWER:	**Peter Bross, M.D. (HFD-150)**
PROJECT MANAGER:	**Ms. Ann Staten (HFD-150)**
DATE SUBMITTED TO DFS:	**November 29, 2002**

Distribution: NDA 21-335
HFD-150/Staten
HFD-150/Bross
HFD-150/Johnson
HFD-150/Pazdur
HFD-710/Rothmann
HFD-710/Chen
HFD-710/Mahjoob
HFD-710/Chi
HFD-710/Anello

File Directory: C:/fda.pro/NDA_21335/June 2002/Review.doc

Table of Contents

STATISTICAL REVIEW AND EVALUATION

1 Executive Summary and Statistical Findings

1.1 Conclusions and Recommendations

Results for time to progression (includes deaths as events) and time to progression of accelerated phase or blast crisis (could include death as an events) are both quite statistically significant favoring Gleevec – as are results for complete hematological response (CHR) and major cytogenetic response (MCyR). Those p-values for time to progression, CHR, and MCyR are much less than (<<) 0.0000001.

1.2 Overview of the Studies Reviewed

This review will concentrate solely on the results of study CSTI571 0106. Study CSTI571 0106 was a randomized, open-label, multicenter, phase III study comparing the experimental treatment of Gleevec (STI571; Imatinib for injection) with the active-control (standard therapy) of Interferon-α (IFN) combined with Cytarabine (Ara-C) in patients with newly diagnosed previously untreated Philadelphia chromosome positive (Ph+) chronic myelogenous leukemia in chronic phase (CML-CP). One thousand one hundred six (1106) patients at 177 centers in 16 countries were randomized to the two arms. One thousand thirty-two (1032) was the planned number of patients. The first patient was randomized June 16, 2000 and the analysis cutoff date is January 31, 2002.

Patients in the Gleevec arm were to receive a once daily dose of 400 mg Gleevec. Patients in the control arm were to receive subcutaneously 5 mg/m^2/day of IFN and for ten days each month also 20 mg/m^2/day of Ara-C.

Patients were to receive the randomized therapy until there was no evidence of lack of response, disease progression or intolerance. Patients could have been offered the possibility of receiving the therapy of the other arm for any of the following: loss of complete hematological response (CHR), loss of major cytogenetic response (MCyR), increasing white blood cell count, intolerance of treatment, failure to achieve a CHR by 12 months, or failure to receive a MCyR by 12 months.

This submission presents interim results, 12 months after the last patient was randomized into the study. The interim analysis has MCyR rate at 12 months as the primary endpoint.

For approval, the primary endpoint is time to progression (TTP). Time to progression was defined as the time from randomization to the first of the following:

- Death (due to any cause when reported as primary reason for discontinuation of treatment).
- Progression to accelerated phase (AP) or blast crisis (BC).
- Loss of MCyR.
- Loss of CHR.

- Increase in white blood cell count (if approved by the Study Management Committee).

Otherwise a patient's TTP will be censored at the last examination/assessment date. If patients discontinue study medication, the last date the patient used study medication will be used. If patients crossed over, the last date on first-line treatment was used as the date of last examination.

The FDA medical reviewer regards time to progression to AP or BC as an important endpoint. This endpoint is defined as the time from randomization to the first of death (due to CML when reported as the primary reason for discontinuation of treatment), or progression to AP or BC.

1.3 Some Statistical and Technical Issues

Results for time to progression and major cytogenetic response are greatly statistically significant.

- There was rather poor dose compliance in the IFN+Ara-C arm. For the IFN+Ara-C arm 87.2% of the patients had a dose change from there initial dose compared to 44.8% for the Gleevec arm. See the medical reviewer's review for further details.
- This submission gives positive results for one study – not two studies. There is no replicate study.
- Interim analysis of time to progression is an unplanned analysis.
- There was a statistically significant difference in the censoring distributions for time to progression (a greater censoring distribution for the Gleevec arm).
- No type I error rate (level of significance) was spent for an interim analysis of time to progression at this time.

1.4 Principal Findings

The following table gives the sponsor's results for unconfirmed major and complete cytogenetic response (CCyR):

Table 1.4.1: Sponsor's Summary of MCyR and CCyR.

	Gleevec (n=553)	IFN+Ara-C (n=553)
MCyR and CCyR		
Number of MCyR	457 (82.6%)	220 (39.8%)
95% C.I.	0.792-0.857	0.357-0.440
p-value	< 0.001	
Number of CCyR	375 (67.8%)	110 (19.9%)
p-value	< 0.001	
MCyR and CCyR (up to crossover)		
Number of MCyR	457 (82.6%)	112 (20.3%)
95% C.I.	0.792-0.857	0.170-0.238
p-value	< 0.001	
Rate of CCyR	375 (67.8%)	41 (7.4%)
p-value	< 0.001	
MCyR and CCyR (second-line treatment)		
Patients starting second-line treatment	7	218
Number	0 (0.0%)	116 (53.2%)
95% C.I.	0-0.410	0.464-0.600

For major cytogenetic response, the results are quite statistically significant favoring the Gleevec arm. For the ITT population, the relative frequency of a major cytogenetic response in the Gleevec was more than twice that in the IFN+Ara-C arm. Many of the major cytogenetic responses (108/220) in the IFN+Ara-C arm occurred after crossover to Gleevec. None of the major cytogenetic responses in the Gleevec arm occurred after crossover to IFN+Ara-C.

Reviewer Comment:

1. Valid comparisons cannot be made for second-line treatment, since those that received second-line treatment were not randomly divided between the two arms.

2. Table 1.4.2 below gives some Fisher's exact test p-values and exact 95% confidence intervals (not provided by the sponsor) based on this reviewer's calculations for MCyR and CCyR. The p-values are quite small.

Table 1.4.2: Reviewer's Summary of some Fisher's Exact Test P-values and C.I.'s for MCyR and CCyR.

	Gleevec (n=553)	IFN+Ara-C (n=553)
MCyR and CCyR		
Number of MCyR	457	220
p-value [1]	<< 0.0000001	
Number of CCyR	375	110
95% C.I.	0.637-0.717	0.166-0.235
p-value [2]	<< 0.0000001	
MCyR and CCyR (up to crossover)		
Rate of MCyR	457	112
p-value [3]	<< 0.0000001	
Rate of CCyR	375	41
95% C.I.	0 637-0 717	0.054-0.099
p-value [4]	<< 0.0000001	

[1] p-value = 4.69×10^{-50}. [2] p-value = $3\ 70 \times 10^{-60}$. [3] p-value = $2\ 22 \times 10^{-102}$ [4] p-value = 3.95×10^{-105}.

For major cytogenetic response, the results are quite statistically significant favoring the Gleevec arm.

The following table gives the sponsor's results for duration of major cytogenetic response for first-line treatment.

Table 1.4.3: Sponsor's Summary for Duration of MCyR.

	Gleevec (n=553)	IFN+Ara-C (n=553)
Number of MCyR	457	112
Number of patients who lost MCyR	7 (1.5%)	9 (8.0%)
Estimated % (95% CI) MCyR maintained after 6 months	98.8% (97, 100)	90.0% (83, 97)
Estimated % (95% CI) MCyR maintained after 9 months	98.8% (97, 100)	90.0% (83, 97)
Estimated % (95% CI) MCyR maintained after 12 months	97.7% (95, 100)	82.5% (67, 98)

The duration of major cytogenetic response is well maintained for both arms with a greater duration for the Gleevec arm.

Reviewer Comment:

The confidence intervals presented by the sponsor for duration of major cytogenetic response are based on normal theory. This is not an appropriate way of constructing confidence intervals in this setting (with estimates far away from 0.5). More appropriate

methods of constructing 95% confidence intervals here give for both treatment arms lower limits and upper limits smaller than those presented by the sponsor.

The following table gives the sponsor's results (more detailed) for major and complete cytogenetic response for those patients that received first-line IFN+Ara-C and second-line Gleevec.

Table 1.4.4: Sponsor's Summary of MCyR and CCyR for Second-line Gleevec.

Responses to first-line IFN+Ara-C	Responses to second-line Gleevec			
	Complete	Partial	No response	Total
Complete	2	0	1	3
Partial	3	3	4	10
No response	66	42	97	205
Total	71	45	102	218

We have that most of those patients (108/205 = 53%) that crossed over from treatment of IFN+Ara-C without responding to Gleevec responded.

The following table gives the sponsor's results for confirmed major and complete cytogenetic response up to crossover.

Table 1.4.5: Sponsor's Summary of Confirmed MCyR and CCyR (up to crossover).

	Gleevec (n=553)	IFN+Ara-C (n=553)
MCyR and CCyR (up to crossover)		
Number of MCyR	419 (75.8%)	67 (12.1%)
95% C.I.	0.720-0.793	0.095-0.151
p-value	< 0.001	
Number of CCyR	297 (53.7%)	15 (2.7%)

For this population, the relative frequency of a confirmed major cytogenetic response in the Gleevec arm was more than six times that in the IFN+Ara-C arm.

Reviewer Comment:
Table 1.4.6 below gives some Fisher's exact test p-values and exact 95% confidence intervals (not provided by the sponsor) based on this reviewer's calculations for confirmed MCyR and CCyR. The results are quite statistically significant. The p-values are quite small.

Table 1.4.6: Reviewer's Summary of Fisher's Exact Test P-values for Confirmed MCyR and CCyR (up to crossover).

	Gleevec (n=553)	IFN+Ara-C (n=553)
MCyR and CCyR (up to crossover)		
Number of MCyR	419	67
p-value [1]	<< 0.0000001	
Number of CCyR	297	15
95% C.I.	0.495-0.579	0.015-0.044
p-value [2]	<< 0.0000001	

[1] p-value = 3.21×10^{-109}. [2] p-value = 2.56×10^{-91}.

At the time of analysis, according to the sponsor, there were 24 and 103 events of progression respectively for the Gleevec and IFN+Ara-C arms. The sponsor reported a p-value (log-rank test or Wilcoxon test) of less than 0.001. From this reviewer's calculations the log-rank test p-value is much less than 0.0000001 (p-value = 3.76×10^{-17}) with an estimated Gleevec vs. IFN+Ara-C hazard ratio of 0.183 (95% C.I. of (0.117, 0.285)).

From the FDA medical officers assessments there were 24 and 117 events of progression respectively for the Gleevec and IFN+Ara-C arms with a corresponding log-rank test p-value (based on this reviewer's calculations) that is much less than 0.0000001 (p-value = 3.62×10^{-21}) with an estimated Gleevec vs. IFN+Ara-C hazard ratio of 0.158 (95% C.I. of (0.102, 0.245)).

Reviewer Comment:
For time to progression (sponsor's submission and FDA medical officers assessments) there was a statistically significant difference (p-values = 0.014 and 0.020, respectively) in the censoring distributions for time to progression. The censoring distribution was larger for the Gleevec arm. This difference does not seem to have a great impact on the difference between arms in time to progression.

At the time of analysis, based on the sponsor's submission, there were 10 and 36 events of progression to AP or BC respectively for the Gleevec and IFN+Ara-C arms. None of these events were due to deaths. The sponsor reported a p-value (log-rank test or Wilcoxon test) of less than 0.001 favoring Gleevec. From this reviewer's calculations the log-rank test p-value is approximately 0.000014 with an estimated Gleevec vs. IFN+Ara-C hazard ratio of 0.239 (95% C.I. of (0.119, 0.482)). Results are similar for time to progression to AP (also 10 and 36 events respectively for the Gleevec and IFN+Ara-C arms).

From the FDA medical officers assessments there were 10 and 37 events of progression to AP respectively for the Gleevec and IFN+Ara-C arms with a corresponding log-rank test p-value of approximately 0.000008 (based on this reviewer's calculations) and an estimated Gleevec vs. IFN+Ara-C hazard ratio of 0.233 (95% C.I. of (0.116, 0.469)). None of these events were due to deaths.

Reviewer Comment:
For time to progression to AP (sponsor's submission and FDA medical officers assessments) there was a statistically significant difference (p-values = 0.017 and 0.018, respectively) in the censoring distributions for time to progression to AP. The censoring distribution was larger for the Gleevec arm. This difference does not seem to have a great impact on the difference between arms in time to progression to AP.

At the time of analysis, there were 11 and 20 events of death respectively for the Gleevec and IFN+Ara-C arms. The sponsor reported that the results were not statistically significant.

2 Statistical Review and Evaluation of Evidence

2.1 Introduction

Study CSTI571 0106 was a randomized, open-label, multicenter, phase III study comparing the experimental treatment of Gleevec (STI571; Imatinib for injection) with the active-control (standard therapy) of Interferon-α (IFN) combined with Cytarabine (Ara-C) in patients with newly diagnosed previously untreated Philadelphia chromosome positive (Ph+) chronic myelogenous leukemia in chronic phase (CML-CP). One thousand one hundred six (1106) patients at 177 centers in 16 countries were evenly randomized to the two arms. One thousand thirty-two (1032) was the planned number of patients. The first patient was randomized June 16, 2000 and the analysis cutoff date is January 31, 2002.

Patients in the Gleevec arm were to receive a once daily dose of 400 mg Gleevec. Patients in the control arm were to receive subcutaneously 5 mg/m^2/day of IFN and for ten days each month also 20 mg/m^2/day of Ara-C.

Patients were to receive the randomized therapy until there was no evidence of lack of response, disease progression or intolerance. Patients could have been offered the possibility of receiving the therapy of the other arm for any of the following: loss of complete hematological response (CHR), loss of major cytogenetic response (MCyR), increasing white blood cell count, intolerance of treatment, failure to achieve a CHR by 12 months, or failure to receive a MCyR by 12 months.

2.2 Statistical Issues:

Results are greatly statistically significant. There was a large difference in dose compliance between treatment arms. This submission gives positive results for one study – not two studies. There is no replicate study. Interim analysis of time to progression is an unplanned analysis. No type I error rate (level of significance) was spent for an interim analysis of time to progression at this time. There is also a statistically significant difference in the censoring distributions for time to progression.

2.3 Study CSTI571 0106

2.3.1 Background

Gleevec was approved (accelerated approval) on May 10, 2001 for the treatment of patients with chronic myeloid leukemia (CML) in blast crisis, accelerated phase, or in chronic phase after the failure of interferon-α therapy and approved (accelerated approval) on February 1, 2002 for the treatment of patients wit Kit positive unresectable and/or metastatic malignant gastrointestinal stromal tumors. Interferon-α (IFN) combined with Cytarabine (Ara-C) control therapy, is standard first-line therapy for patients with newly diagnosed previously untreated Philadelphia chromosome positive (Ph+) chronic myelogenous leukemia in chronic phase (CML-CP).

2.3.2 Data Analyzed and Sources

Data used in this review are from the electronic submission received 6/28/02 and 8/30/02. The network paths are "\\cdsesub\N21335\S_004\2002-06-28" and "\\cdsesub1\N21335\S_004\2002-08-30" in EDR. Volumes 2, 11, 12 and 33 were reviewed.

2.3.3 Study Objectives

The primary study objective is to compare the time to progression in adult patients with newly diagnosed previously untreated Ph+ CML between the experimental arms.

Secondary objectives include

- To evaluate quality of life, disease-related toxicities and treatment-related toxicities.
- To determine the rate and duration of complete hematological response (CHR).
- To determine the rate and duration of major cytogenetic response (MCyR).
- To determine the rate and duration of CHR and MCyR due crossover therapy.
- To determine overall survival.
- To evaluate healthcare resource utilization (RU).
- To determine the tolerability and safety.
- To evaluate the population pharmacokinetics.

- To perform pharmacogenomic evaluations, to study in an exploratory fashion RNA expression and SNA polymorphisms in tumor cells in the blood and bone marrow.
- To follow molecular response in patients who achieved a complete cytogenetic response (CCyR).

2.3.4 Efficacy Endpoints

The primary efficacy endpoint for approval is time to progression. Time to progression was defined as the time from randomization to the first of the following:

- Death (due to any cause when reported as the primary reason for discontinuation of treatment).
- Progression to accelerated phase (AP) or blast crisis (BC).
- Loss of MCyR.
- Loss of CHR.
- Increase in white blood cell count (if approved by the Study Management Committee).

In the case of no event, time to progression was censored at the last examination date.

An interim analysis was planned for major cytogenetic response. Cytogenetic evaluations were used in the calculation of cytogenetic responses only if:

- at least 20 metaphases were examined or
- 5-19 metaphases were examined with Ph+ ≤ 35% and either the previous or next assessment also had 5-19 metaphases examined with Ph+ ≤ 35%.

Overall cytogenetic response was defined as

- Complete cytogenetic response (CCyR) (if patients had 0% Ph+ cells at least once).
- Partial cytogenetic response (PCyR) (if patients had ≤ 35% Ph+ cells at least once).
- Minor cytogenetic response (if patients had ≤ 65% Ph+ cells at least once).
- Minimal cytogenetic response (if patients had ≤ 95% Ph+ cells at least once).
- No response (on study) (if patients never had a minimal cytogenetic response and were still on treatment).
- No response (off study) (if patients never had a minimal cytogenetic response and crossed over or discontinued for reasons other than progression or death).
- No response (progressive disease/death) (if patients never had a minimal cytogenetic response and crossed over or discontinued for progression or death).
- Ph- at baseline (patients who had no documentation of Ph+ at baseline; for ITT analyses only, these patients are excluded from per protocol analyses).

Major cytogenetic response is defined as the sum of overall CCyR and PCyR.

Other secondary endpoints include complete cytogenetic response rate, duration of major cytogenetic response, complete hematological response rate, duration of complete hematological response, observed response rates (MyCR and CHR) at 6, 9, and 12 months, time to accelerated phase or blast crisis and overall survival.

2.3.5 Sample Size Considerations

The planned sample size of 1032 was based on a target Gleevec vs. IFN + Ara-C time to progression hazard ratio of 0.75, a 5-year progression-free rate of 50% in the control arm, a 0.5 year accrual period and a median follow up time of 5.25 years. Three hundred eighty-five (385) events are required for this calculation.

For interim analyses, there is 80% power to detect a 10% increase in MCyR rate from an expected 41% MCyR rate for the IFN-Ara-C arm.

2.3.6 Interim Analysis

An interim analysis of MCyR rate was planned for 12 months after the date of the last enrolled patient.

An Independent Data Monitoring Board will evaluate safety and efficacy data on an ongoing basis and evaluate the 1-year MCyR rate.

2.3.7 Efficacy Analysis Methods

All proportions based on binary variables are presented with their corresponding exact 95% confidence interval and when appropriate (e.g., MyCR) compared using Fisher's exact test. Kaplan-Meier methods were used to determine estimates of the median times (and other quartiles) and estimated percentage of patients without an event at 6, 9 and 12 months, along with all corresponding 95% confidence intervals. P-values for comparing the two estimated distributions were determined from the log-rank and Wilcoxon tests.

Efficacy analyses are based on the intent-to-treat population – all patients as randomized.

2.3.8 Sponsor's Results and Statistical Reviewer's Findings/Comments

2.3.8.1 Baseline Characteristics

The table below gives a summary of the baseline distribution of various characteristics. The distributions between arms are similar.

Table 2.3.8.1.1: Sponsor's Summary of Baseline Characteristics

	Gleevec (n=553)	IFN+Ara-C (n=553)
Sex		
Men	342 (61.8%)	310 (56.1%)
Women	211 (38.2%)	243 (43.9%)
Age (years)		
Median	50.0	51.0
Range	18-70	18-70
< 40	141 (25.5%)	128 (23.1%)
≥ 40 - < 50	115 (20.8%)	120 (21.7%)
≥ 50 - < 60	183 (33.1%)	177 (32.0%)
≥ 60	114 (20.6%)	128 (23.1%)
Race		
Caucasian	494 (89.3%)	500 (90.4%)
Black	28 (5.1%)	24 (4.3%)
Asian	12 (2.2%)	6 (1.1%)
Other	19 (3.4%)	23 (4.2%)
Weight (kg)		
N	540	539
Median	78.7	77
Range	40.0-169.5	41.0-157.7
Body surface area (m^2)		
N	508	526
Median	1.91	1.88
Range	1.06-2.89	1.29-2.67
ECOG Performance Status		
Missing	5 (0.9%)	12 (2.2%)
Grade 0	425 (76.9%)	409 (74.0%)
Grade 1	115 (20.8%)	121 (21.9%)
Grade 2	8 (1.4%)	11 (2.0%)

The table 2.3.81.2 gives a summary of the baseline distribution of disease characteristics. The distributions between arms are similar.

Table 2.3.8.1.2: Sponsor's Summary of Disease Characteristics at Baseline

	Gleevec (n=553)	IFN+Ara-C (n=553)
Previous hydroxyurea treatment?		
No	68 (12.3%)	81 (14.6%)
Yes	485 (87.7%)	472 (85.4%)
Time since diagnosis (months)		
Median	2.14	1.77
Interquartile range	1.0-3.7	0.8-3.2
Range		

2.3.8.2 Primary Efficacy Analyses:

The following table gives the sponsor's results for unconfirmed major and complete cytogenetic response:

Table 2.3.8.2.1: Sponsor's Summary of MCyR and CCyR.

	Gleevec (n=553)	IFN+Ara-C (n=553)
MCyR and CCyR		
Number of MCyR	457 (82.6%)	220 (39.8%)
95% C.I.	0.792-0.857	0.357-0.440
p-value	< 0.001	
Number of CCyR	375 (67.8%)	110 (19.9%)
p-value	< 0.001	
MCyR and CCyR (up to crossover)		
Number of MCyR	457 (82.6%)	112 (20.3%)
95% C.I.	0.792-0.857	0.170-0.238
p-value	< 0.001	
Rate of CCyR	375 (67.8%)	41 (7.4%)
p-value	< 0.001	
MCyR and CCyR (second-line treatment)		
Patients starting second-line treatment	7	218
Number	0 (0.0%)	116 (53.2%)
95% C.I.	0-0.410	0.464-0.600

The rate of major cytogenetic response was higher for the Gleevec arm than for the IFN+Ara-C arm with the results highly statistically significant.

Reviewer Comment:

1. Valid comparisons cannot be made for second-line treatment, since those that received second-line treatment were not randomly divided between the two arms.

2. Table 2.3.8.2.2 below gives some Fisher's exact test p-values and exact 95% confidence intervals based on this reviewer's calculations for MCyR and CCyR. The p-values are quite small.

Table 2.3.8.2.2: Reviewer's Summary of some Fisher's Exact Test p-values and C.I.'s for MCyR and CCyR.

	Gleevec (n=553)	IFN+Ara-C (n=553)
MCyR and CCyR		
Number of MCyR	457	220
p-value [1]	<< 0.0000001	
Number of CCyR	375	110
95% C.I.	0.637-0.717	0.166-0.235
p-value [2]	<< 0.0000001	
MCyR and CCyR (up to crossover)		
Rate of MCyR	457	112
p-value [3]	<< 0.0000001	
Rate of CCyR	375	41
95% C.I.	0.637-0.717	0.054-0.099
p-value [4]	<< 0.0000001	

[1] p-value = 4.69×10^{-50}. [2] p-value = 3.70×10^{-60}. [3] p-value = 2.22×10^{-102}. [4] p-value = 3.95×10^{-105}.

The following table gives the sponsor's results for duration of major cytogenetic response for first-line treatment.

Table 2.3.8.2.3: Sponsor's Summary for Duration of MCyR.

	Gleevec (n=553)	IFN+Ara-C (n=553)
Number of MCyR	457	112
Number of patients who lost MCyR	7 (1.5%)	9 (8.0%)
Estimated % (95% CI) MCyR maintained after 6 months	98.8% (97, 100)	90.0% (83, 97)
Estimated % (95% CI) MCyR maintained after 9 months	98.8% (97, 100)	90.0% (83, 97)
Estimated % (95% CI) MCyR maintained after 12 months	97.7% (95, 100)	82.5% (67, 98)

The duration of major cytogenetic response is well maintained for both arms with a greater duration for the Gleevec arm.

Reviewer Comment:

The confidence intervals presented by the sponsor here are based on normal theory. This is not an appropriate way of constructing confidence intervals in this setting (with estimates far away from 0.5). More appropriate methods of constructing 95% confidence intervals here give for both treatment arms lower limits and upper limits smaller than those presented by the sponsor.

The following table gives the sponsor's results (more detailed) for major and complete cytogenetic response for those patients that received first-line IFN+Ara-C and second-line Gleevec.

Table 2.3.8.2.4: Sponsor's Summary of MCyR and CCyR for Second-line Gleevec.

Responses to first-line IFN+Ara-C	**Responses to second-line Gleevec**			
	Complete	Partial	No response	Total
Complete	2	0	1	3
Partial	3	3	4	10
No response	66	42	97	205
Total	71	45	102	218

We have that most patients (108/205 = 53%) responded to Gleevec after not responding to (crossing over from) IFN+Ara-C.

The following table gives the sponsor's results for confirmed major and complete cytogenetic response.

Table 2.3.8.2.5: Sponsor's Summary of Confirmed MCyR and CCyR (up to crossover).

	Gleevec (n=553)	**IFN+Ara-C (n=553)**
MCyR and CCyR (up to crossover)		
Number of MCyR	419 (75.8%)	67 (12.1%)
95% C.I.	0.720-0.793	0.095-0.151
p-value	< 0.001	
Number of CCyR	297 (53.7%)	15 (2.7%)

For this population, the relative frequency of a confirmed major cytogenetic response in the Gleevec arm was more than six times that in the IFN+Ara-C arm.

Reviewer Comment:

Table 2.3.8.2.6 below gives some Fisher's exact test p-values and exact 95% confidence intervals based on this reviewer's calculations for confirmed MCyR and CCyR. The results are quite statistically significant. The p-values are quite small.

Table 2.3.8.2.6: Reviewer's Summary of Fisher's Exact Test P-values for Confirmed MCyR and CCyR (up to crossover).

	Gleevec (n=553)	IFN+Ara-C (n=553)
MCyR and CCyR (up to crossover)		
Number of MCyR	419	67
p-value [1]	<< 0.0000001	
Number of CCyR	297	15
95% C.I.	0.495-0.579	0.015-0.044
p-value [2]	<< 0.0000001	

[1] p-value = $3\ 21 \times 10^{-109}$. [2] p-value = 2.56×10^{-91}.

The primary endpoint for approval is TTP. The planned cutoff date for the TTP analysis is the date of the 385th event. At the time of analysis, according to the sponsor, there were 24 and 103 events of progression respectively for the Gleevec and IFN+Ara-C arms. The sponsor reported a p-value (log-rank test or Wilcoxon test) of less than 0.001. From this reviewer's calculations the log-rank test p-value is much less than 0.0000001 (p-value = 3.76×10^{-17}) with an estimated Gleevec vs. IFN+Ara-C hazard ratio of 0.183 (95% C.I. of (0.117, 0.285)). The table below gives the break down of type of progression events.

From the FDA medical officers assessments, there were 24 and 117 events of progression respectively for the Gleevec and IFN+Ara-C arms with a corresponding log-rank test p-value (based on this reviewer's calculations) is much less than 0.0000001 (p-value = 3.62×10^{-21}) with an estimated Gleevec vs. IFN+Ara-C hazard ratio of 0.158 (95% C.I. of (0.102, 0.245)).

Table 2.3.8.2.7: Sponsor's Summary of TTP Events

	Gleevec (n=553)	IFN+Ara-C (n=553)
Total number of progression events	**24 (4.3%)**	**103 (18.6%)**
Progression to accelerated phase or blast crisis	8	32
Loss of CHR	6	39
Loss of MCyR	4	6
Increase in WBC (approved by SMC)	2	24
Death during treatment	4	2

Reviewer Comment:

1. For time to progression (sponsor's submission and FDA medical officers assessments) there was a statistically significant difference (p-values = 0.014 and 0.020, respectively) in the censoring distributions for time to progression. The censoring distribution was larger for the Gleevec arm. This difference does not seem to have a great impact on the difference between arms in time to progression.

2. The planned cutoff date for the TTP analysis is the date of the 385th event. At the time of this interim analysis, according to the sponsor's submission, there were 127 events of progression and the TTP results are quite highly statistically significant favoring the Gleevec arm. If for the remaining 258 events needed the theoretical hazard ratio is 1, then the chance of rejecting the null hypothesis that the two theoretical distributions for TTP are the same (and favoring Gleevec) is roughly 96.9%. Based on the FDA medical officers assessments, if for the remaining 244 events needed (for 385 events total) the theoretical hazard ratio is 1, then the chance of rejecting the null hypothesis that the two theoretical distributions for TTP are the same (and concluding Gleevec is better) is roughly 99.3%. At the beginning of the trial, if the theoretical hazard ratio is 1 then, after 385 events. the chance of rejecting the null hypothesis that the two theoretical distributions for TTP are the same (and favoring Gleevec) is 2.5%.

If for the remaining 258 events needed, the theoretical Gleevec vs. IFN-Ara-C hazard ratio is 0.75 (the value for which the study was powered), then the chance of failing to reject the null hypothesis that the two theoretical distributions for TTP are the same (favoring Gleevec) is roughly 0.000014. Based on the FDA medical officers assessments, if for the remaining 244 events needed (for 385 events total) the theoretical hazard ratio is 0.75, then the chance of failing to reject the null hypothesis that the two theoretical distributions for TTP are the same (and concluding Gleevec is better) is roughly 0.0000014. At the beginning of the trial, if the theoretical Gleevec vs. IFN-Ara-C hazard ratio is 0.75 then, after 385 events, the chance of rejecting the null hypothesis that the two theoretical distributions for TTP are the same (and favoring Gleevec) is roughly 80%.

So, it is rather unlikely that a non-significant result will occur at the time of final analysis of TTP.

These probabilities were calculated by this reviewer by modeling a meta-analysis of current results having 127 events with the estimator and standard error of the log-hazard ratio based on 258 events in a 1:1 randomization and assuming that the theoretical hazard ratio is one. Under this model, the calculated probabilities are independent of the particular common TTP distribution (common when the hazard ratio is 1). Determining these probabilities by using the current censored values provides different values (answers) depending on the particular common TTP distribution (common when the hazard ratio is 1).

2.3.8.3 Secondary Efficacy Analyses

The following table gives the sponsor's results for confirmed complete hematological response:

Table 2.3.8.3.1: Sponsor's Summary of Confirmed CHR.

	Gleevec (n=553)	IFN+Ara-C (n=553)
CHR (ITT principle)		
Number	523 (94.6%)	423 (76.5%)
95% C.I.	0.923-0.963	0.727-0.800
p-value	< 0.001	
CHR (up to crossover)		
Number	522 (94.4%)	302 (54.6%)
95% C.I.	0.921-0.962	0.504-0.588
Number that loss CHR	11	46
p-value	< 0.001	
CHR (second-line treatment)		
Patients starting second-line treatment	7	218
Number	3 (42.9%)	182 (83.5%)
95% C.I.	0 099-0.816	0.779-0.882

Reviewer's Comments:

1. From this reviewer's calculations for comparing confirmed CHR, the Fisher's exact test p-values are much less than 0.0000001 for the ITT principle and up to crossover analyses respectively.

2. Valid comparisons cannot be made for second-line treatment, since those that received second-line treatment were not randomly divided between the two arms.

At the time of analysis, based on the sponsor's submission, there were 10 and 36 events of progression to AP or BC respectively for the Gleevec and IFN+Ara-C arms. None of these events were due to deaths. The sponsor reported a p-value (log-rank test or Wilcoxon test) of less than 0.001 favoring Gleevec. From this reviewer's calculations the log-rank test p-value is approximately 0.000014 with an estimated Gleevec vs. IFN+Ara-C hazard ratio of 0.239 (95% C.I. of (0.119, 0.482)). Results are similar for time to progression to AP (also 10 and 36 events respectively for the Gleevec and IFN+Ara-C arms).

From the FDA medical officers assessments there were 10 and 37 events of progression to AP respectively for the Gleevec and IFN+Ara-C arms with a corresponding log-rank test p-value of approximately 0.000008 (based on this reviewer's calculations) and an estimated Gleevec vs. IFN+Ara-C hazard ratio of 0.233 (95% C.I. of (0.116, 0.469)). None of these events were due to deaths.

Reviewer Comment:
For time to progression to AP (sponsor's submission and FDA medical officers assessments) there was a statistically significant difference (p-values = 0.017 and 0.018, respectively) in the censoring distributions for time to progression to AP. The censoring distribution was larger for the Gleevec arm. This difference does not seem to have a great impact on the difference between arms in time to progression to AP.

At the time of analysis, there were 11 and 20 events of death respectively for the Gleevec and IFN+Ara-C arms. The sponsor reported that the results were not statistically significant.

The interim report on QoL consists of data from all sites except those sites in Denmark and the Flemish speaking part of Belgium. These QoL data involve 1067 patients – 533 on Gleevec and 534 on IFN+Ara-C. The Functional Assessment of Cancer Therapy – Biological Response Modifier (FACT-BRM) was used to assess quality of life. The FACT-BRM consists of a general quality of life instrument in addition to several treatment-specific modules. It consists of questions concerning the well being of physical, functional, social and emotional domains, and questions with regards to the impact of biological response modifiers on physical and emotional/cognitive functioning. The numerical codes of the responses are then summed to yield subscale or domain scale with a higher score representing better QoL. The primary QoL endpoint of TOI (Total Outcome Index) consists of 27 items – those physical and functional well-being items, and those physical and emotional/cognitive functioning treatment-specific items.

Statistical analyses of FACT-BRM TOI score used polynomial growth curves and a pattern-mixture technique was utilized to adjust for incomplete data during the 12-month period.

Statistical Comment:

1. Because of the potential for informative dropouts, any statistical analysis may not be valid and the small p-values may not be interpretable.
2. An adjustment for nominal p-values is necessary for multiple comparisons.

Table 2.3.8.3.2 by the sponsor shows the adjusted mean TOI raw scores during the first 12 months based on a mixed pattern analysis.

Table 2.3.8.3.2: Sponsor's Summary of Trial Outcome Index (TOI) Raw Score Mixed Pattern Analysis (ITT QoL population).

	Gleevec		IFN+Ara-C
Baseline Visit (Visit 1)			
Adjusted Mean ± s.e.	83.58 ±1.05		81 35 ±1 13
Diff (Gleevec-IFN)		2.23	
p-value		<0.0683	
95% CI		[-0.70, 5.15]	
Month 1 (Visit 6)			
Adjusted Mean ± s.e.	84.16 ±1.05		64.62 ±1 11
Diff (Gleevec-IFN)		19.54	
p-value		<0 0001	
95% CI		[16.61, 22.47]	
Month 2 (Visit 8)			
Adjusted Mean ± s e.	85 42 ±1 02		63.62 ±1 08
Diff (Gleevec-IFN)		21.80	
p-value		<0.0001	
95% CI		[18.88, 24.72]	
Month 3 (Visit 10)			
Adjusted Mean ± s.e.	86.42±1.04		66.08 ±1 15
Diff (Gleevec-IFN)		20.34	
p-value		<0.0001	
95% CI		[17.49, 23.19]	
Month 4 (Visit 12)			
Adjusted Mean ± s.e.	86.85±1.07		67.64 ±1.20
Diff (Gleevec-IFN)		19.20	
p-value		<0 0001	
95% CI		[16.27, 22.13]	
Month 5 (Visit 13)			
Adjusted Mean ± s.e.	86.81±1.09		68.10 ±1 25
Diff (Gleevec-IFN)		18.71	
p-value		<0.0001	
95% CI		[15.71, 21.71]	
Month 6 (Visit 14)			
Adjusted Mean ± s.e.	86.58±1.18		68.55 ±1.35
Diff (Gleevec-IFN)		18.03	
p-value		<0.0001	
95% CI		[14.97, 21.09]	
Month 9 (Visit 16)			
Adjusted Mean ± s.e.	87.03±1.28		71.84 ±1.69
Diff (Gleevec-IFN)		15.20	
p-value		<0.0001	
95% CI		[11.96, 18 44]	
Month 12 (Visit 18)			
Adjusted Mean ± s.e.	87.19±0.90		77.95 ±1.47
Diff (Gleevec-IFN)		9.24	
p-value		<0.0001	
95% CI		[5.87, 12.62]]	

Statistical Comments:

1. Using these "baseline adjusted" means and corresponding standard errors, the two-sided p-value for comparing baseline values between arms is 0.148.
2. The multipliers for the 95% confidence intervals for the difference in theoretical means are different for baseline values (1.89), month 1 visit (2.05), month 2 visit (1.96) and month 3 (1.84). Multipliers not checked for any other time point.
3. It is not clear that at each time point differences between arms in changes from baseline in TOI were analyzed. It appears that at each time point differences between arms in TOI were analyzed ("baseline adjusted" means are different). Differences from baseline should be compared.
4. It is not clear whether "baseline adjusted" refers to adjustment due to baseline covariates.
5. There was a much greater QoL "dropout" rate for the IFN+Ara-C arm.

Table 2.3.8.3.3 gives the sponsor's summary of the average QoL scores of TOI over twelve months. According to the sponsor, an intent-to-treat approach is used.

Table 2.3.8.3.3: Sponsor's Summary of Average Trial Outcome Index Raw (ITT approach).

	Gleevec		IFN+Ara-C
Trial Outcome Index			
N	486		461
Adjusted Mean	84.3		67.1
Diff (Gleevec-IFN)		17.2	
p-value		<0.0001	
95% CI		[15.5, 18.9]	

Table 2.3.8.3.4 gives the sponsor's summary of TOI raw scores at each QoL assessment up to 12 months. According to the sponsor, *"No formal statistics were performed on data in this table and results should be viewed accordingly."*

Table 2.3.8.3.4: Sponsor's Summary of Trial Outcome Index Raw Scores

QoL Parameter	Gleevec	IFN+Ara-C
Baseline Visit (Visit 1)		
N	492	484
Mean	84.0	81.8
s.d.	15.69	17.05
Median	87	84
25^{th} –75^{th} percentile	[75, 96]	[71, 95]
Min-max		
Month 1 (Visit 6)		
N	483	435
Mean	85.0	64.5
s.d.	15.98	21.56
Median	89	66

25^{th} –75^{th} percentile	[76, 97]	[49, 82]
Min-max		
Month 2 (Visit 8)		
N	472	414
Mean	86.5	64.6
s d.	15.87	20.36
Median	90	66
25^{th} –75^{th} percentile	[77, 99]	[51, 81]
Min-max		
Month 3 (Visit 10)		
N	484	416
Mean	87.0	67.3
s d.	15.69	20.73
Median	91	69
25^{th} –75^{th} percentile	[78, 91]	[54, 82]
Min-max		
Month 4 (Visit 12)		
N	472	382
Mean	87.7	67.4
s.d.	15.73	20.65
Median	91	68
25^{th} –75^{th} percentile	[78, 100]	[53, 83]
Min-max		
Month 5 (Visit 13)		
N	463	378
Mean	87.5	68.9
s.d.	16.08	20.71
Median	91	70
25^{th} –75^{th} percentile	[79, 100]	[41, 84]
Min-max		
Month 6 (Visit 14)		
N	469	381
Mean	87.3	70.8
s d.	15.93	20.67
Median	91	73
25^{th} –75^{th} percentile	[78, 100]	[57, 86]
Min-max		
Month 9 (Visit 16)		
N	425	341
Mean	87.4	74.3
s d.	16.23	20.24
Median	90	76
25^{th} –75^{th} percentile	[77, 100]	[60, 90]
Min-max		
Month 12 (Visit 18)		
N	424	313
Mean	87.5	77.3
s.d.	16.15	19.66
Median	91	81
25^{th} –75^{th} percentile	[79, 100]	[64, 93]
Min-max		

Statistical Comments:

1. Using these baseline means and corresponding standard errors, the two-sided p-value for comparing baseline values between arms is 0.036. At the 5% significance level, there is a statistically significant difference in the means of the two arms.
2. There was a much greater QoL "dropout" rate for the IFN+Ara-C arm.

2.3.9 Sponsor's Conclusion and Reviewer's Conclusion/Comments

According to the sponsor, the results of the interim analysis indicate that first-line treatment in CML patients with Gleevec significantly delays progression to AP or BC and is associated with a markedly higher rate of complete hematological response, major and complete cytogenetic response, and an improvement in progression-free rate at 12 months when compared to first-line treatment in CML patients with IFN+Ara-C. According to the sponsor, *"These differences remain large and highly statistically significant even when very conservative statistical approaches are taken such as censoring discontinuations for non-efficacy reasons in the calculation of the estimated rate of response, and the ITT approach ignoring crossover effects related to the high efficacy of second-line STI571* [Gleevec]."

The rate of major cytogenetic response was higher for the Gleevec arm than for the IFN+Ara-C arm with the results reaching statistical significance.

The primary endpoint for approval is TTP. The planned cutoff date for the TTP analysis is the date of the 385th event. At the time of this interim analysis, there were 127 events of progression and the TTP results are quite highly statistically significant favoring the Gleevec arm. If for the remaining 258 events needed the theoretical hazard ratio is 1, then the chance of rejecting the null hypothesis that the two theoretical distributions for TTP are the same (favoring Gleevec) is roughly 96.9%. So, it is unlikely that a non-significant result will occur at the time of final analysis of TTP.

There was rather poor dose compliance in the IFN+Ara-C arm. For the IFN+Ara-C arm 87.2% of the patients had a dose change from there initial dose compared to 44.8%. See the medical reviewer's review for further details. This submission gives positive results for one study – not two studies. There is no replicate study. Interim analysis of time to progression is an unplanned analysis.

Results are quite highly statistically significant.

3 Statistical Evaluation of Collective Evidence

Gleevec is proposed to be used as first-line therapy in patients with Philadelphia chromosome positive (Ph+) chronic myelogenous leukemia (CML) in chronic phase. For approval, the applicant submitted an interim report for CSTI571 0106. Study CSTI571 0106 was a randomized, open-label, multicenter, phase III study comparing the experimental treatment of Gleevec (STI571; Imatinib for injection) with the active-control (standard therapy) of Interferon-α (IFN) combined with Cytarabine (Ara-C) in patients with newly diagnosed previously untreated Philadelphia chromosome positive (Ph+) chronic myelogenous leukemia in chronic phase (CML-CP). One thousand one hundred six (1106) patients at 177 centers in 16 countries were evenly randomized to the two arms.

There was rather poor dose compliance in the IFN+Ara-C arm. For the IFN+Ara-C arm 87.2% of the patients had a dose change from there initial dose compared to 44 8%. See the medical reviewer's review for further details.

The rate of major cytogenetic response was higher for the Gleevec arm than for the IFN+Ara-C arm with the results reaching statistical significance. For those patients on the Gleevec arm 82.6% (457/553) had a major cytogenetic response compared to 39.8% (220/553) on the IFN+Ara-C arm. The results were quite highly statistically significant favoring the Gleevec arm. Many of the major cytogenetic responses in the IFN+Ara-C arm occurred after crossing over to Gleevec. Up to crossover to therapy to the other arm, there were 82.6% (457/553) major cytogenetic responses on Gleevec compared to 20.2% (112 553) major cytogenetic responses on IFN+Ara-C. Patients were to receive the randomized therapy until there was no evidence of lack of response, disease progression or intolerance. Patients could have been offered the possibility of receiving the therapy of the other arm for any of the following: loss of complete hematological response (CHR), loss of major cytogenetic response (MCyR), increasing white blood cell count, intolerance of treatment, failure to achieve a CHR by 12 months, or failure to receive a MCyR by 12 months.

The primary endpoint for approval is TTP. At the time of analysis, there were 24 and 103 events of progression respectively for the Gleevec and IFN+Ara-C arms. The results were quite highly statistically significant favoring the Gleevec arm with an estimated Gleevec vs. IFN+Ara-C hazard ratio of 0.183 (95% C.I. of (0.117, 0.285)). From the FDA medical officers assessments, there were 24 and 117 events of progression respectively for the Gleevec and IFN+Ara-C arms with an estimated Gleevec vs. IFN+Ara-C hazard ratio of 0.158 (95% C.I. of (0.102, 0.245)). It is rather unlikely that a non-significant result will occur at the time of final analysis of TTP.

Mark Rothmann, Ph.D.
Mathematical Statistician
Dated: December 19, 2002

Concur: Dr. Chen
Team Leader

Dr. Mahjoob
Deputy Division Director, DBI

Cc:
HFD-150/ Ms. Staten
HFD-150/ Dr. Bross
HFD-150/ Dr. Johnson
HFD-710/ Dr. Rothmann
HFD-710/ Dr. Chen
HFD-710/ Dr. Mahjoob
HFD-710/ Dr. Chi
HFD-700/ Dr. Anello

This review consists of 26 pages (cover page, table of contents and 24 pages of text).
C:\fda.pro\NDA_21335\June 2002\review.doc
First draft submitted 10-24-02

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

/s/

Mark Rothmann
12/19/02 04:30:19 PM
BIOMETRICS

Gang Chen
12/20/02 11:45:49 AM
BIOMETRICS

Kooros Mahjoob
12/20/02 02:15:03 PM
BIOMETRICS