申請資料概要（第Ⅱ相試験における奏効率）と
ISEL 試験（第Ⅲ相試験における生存期間）の結果の関係に関する資料

# GAIYO compared to ISEL

- The Gaiyo and ISEL have looked at different primary endpoints
- The Gaiyo looked retrospectively at response rates within Iressa treated patients.
- ISEL looks at the treatment difference, Iressa compared to placebo, on survival.
- The difference seen in ISEL between Orientals and non Orientals for survival is robust, as seen in the stratified log rank test, further subset analyses and bootstrap, re-sampling procedures.

IRESSA IDEAL RR in Japanese vs Non Japanese and conflict with ISEL Oriental subset

The analysis in the Gayio examined the difference in response rates between Japanese and non Japanese patients treated with Iressa in the IDEAL study. After adjustment for confounding prognostic factors, the difference in response rate between Japanese and non Japanese was not statistically significant.

ISEL, is a large, placebo controlled trial with survival as the primary endpoint. Here the difference in survival, Iressa compared to placebo, is significantly larger in Oriental patients than non Oriental patients. This finding is not in conflict with the finding based on response rates in the non placebo controlled IDEAL trial. The survival benefit seen in Oriental patients is due to increased survival in both responders and non responders. In non Oriental patients, only responders appear to have increased survival.

Hence, based on the ISEL data, it is possible for two populations such as Oriental and non Oriental or Japanese and non Japanese, to have similar response rates but yet have different survival benefits since improvements in survival can come from two sources (responders and non responders) in the first population and only one source (responders) in the second population. This suggests that Oriental patients with stable disease have a survival improvement with Iressa. This does not not appear to be the case for non Oriental patients.

# 1839IL/0016 Multivariate Analysis of Tumour Response Rate

CONTENTS PAGE

# 1 SUMMARY

Due to a statistically significant difference being observed between Japanese and non-Japanese patients in terms of the tumour response rate endpoint, multivariate logistic analysis was performed. By employing a multivariate method of analysis, it was possible to identify baseline prognostic factors and present a more accurate comparison of the response rate seen in Japanese and non-Japanese patients.

Twenty-two baseline factors were evaluated independently to assess their value in predicting response. Using a 10% significance level, only 7 factors were found to be predictive of response (baseline lung cancer subscale, body mass index [BMI], performance status [PS], prior radiotherapy, histology, prior immuno/hormonal therapy and gender). Using only these 7 factors, all were included in one model along with the factor for ethnicity. By assessing all factors together in one model, it was possible to account for confounding factors and allow a more sensitive comparison of the apparent ethnic difference. To ensure only relevant baseline factors were retained in the multivariate model, the backward regression technique was employed at the 10% significance level. This resulted in only 4 factors being retained in the model (PS, gender, histology and prior immuno/hormonal therapy).

The final multivariate model, including all 4 significant baseline prognostic factors, and the factor for ethnicity, resulted in an odds ratio for Japanese:non-Japanese of 1.64 (p=0.2530). Although the odds ratio indicated that the estimated odds of responding was 1.64 times higher for Japanese patients compared to non-Japanese patients, the 95% confidence interval showed that the true odds ratio could lie anywhere between 0.71 and 3.93.

# 2 INTRODUCTION

Following the unadjusted analysis of the tumour response rates, further multivariate analysis was performed to identify baseline factors that could affect tumour response in this trial. This analysis was not only able to identify baseline prognostic factors, but it was also able to adjust the odds ratio when comparing ethnic groups by accounting for identified baseline imbalances. Although multivariate analysis was discussed in the clinical study report (CSR), this was based only on the factors identified at that time. However, since the initial analysis, many other baseline factors were tested for prognostic value in an attempt to gain a better understanding of the ethnic difference. Therefore, the analysis discussed in this document is based on the analysis performed after the analysis conducted for the CSR.

# 3 METHODOLOGY

As stated in the statistical analysis plan, logistic regression models were to be used to further explore a significant group difference should a difference occur. The purpose of this analysis was to learn more about the relationship between baseline factors and tumour response. This would not only allow the identification of possible prognostic factors but also allow a more sensitive comparison of groups.

Although the initial analysis using Fisher's exact test allowed us to identify the crude difference in response rates between ethnic groups it was unable to control for confounding factors. Logistic regression provided a simplified, quantitative description of the main features of the relationship between several prognostic factors and the probability of response. It enables the probability of response to be predicted even for categories in which little information is available. The logistic model derives its name from the fact that the logit transform of the response probability in each category is expressed as a linear function of regression variables whose values correspond to the levels of exposure to the baseline factors.

If $p$ is the probability of response and $(x_1, \ldots, x_k)$ are the set of baseline factors, then logit (p), or the odds of response, can be expressed as a linear combination of these baseline factors as follows:

$$\text{Logit (p)} = \log (p/(1-p)) = \alpha + \sum_{(k=1,\ldots,K)} \beta_k x_k$$

so that

$$p = e^{\alpha + \sum_{(k=1,\ldots,K)} \beta k X k} / (1 + e^{\alpha + \sum_{(k=1,\ldots,K)} \beta k 7 X k})$$

Therefore, $e^{\alpha}$ refers to the baseline probability of response. In the simple case of a two level factor $e^{\beta k}$ can be interpreted as the odds of responding for those patients exposed to factor k compared to those not exposed. More generally, $e^{\beta k}$ is the fraction by which the odds of responding is increased or decreased for every unit change in $x^k$ compared with a person for whom $x^k = 0$ and $e^{\sum(k=1,\ldots,K)\ \beta k(Xk^*-Xk)}$ is the odds of responding for a patient having baseline variables $x^*$ compared to those having baseline variables $x$.

The model parameters are estimated using the method of maximum likelihood. The likelihood of the model is the probability of seeing the observed data, and a sensible way to select the parameters is to select those which maximise the likelihood. To decide which baseline factors to exclude, a likelihood ratio test is performed. The log-likelihood test statistic is defined as – 2 times the maximised log likelihood or:

$$G = -2 \sum \{y \log p_{hat} + (1-y) \log (1-p_{hat})\}$$

Where $p_{hat}$ is the fitted $p$ obtained by putting the fitted parameters back in the model and $y$ is the response status. Comparing the difference between G from two different models to the $X^2$ distribution tells us whether or not it is sensible to include the factor in the model. A factor should only be included in the model if the difference between G for the model which includes it and G for the model which excludes it is significant at the 10% significance level with degrees of freedom equal to the difference between the degrees of freedom of the other two models.

# 4 MODEL BUILDING

When the data was analysed the group which showed a significant difference in tumour response rates was the comparison of Japanese and non-Japanese patients. To explore the reason for this apparent difference the data was analysed using logistic regression. The first analysis did not account for any baseline factors other than ethnicity and this resulted in an odds ratio of 3.27, indicating that the chances of responding was over 3 times higher for Japanese patients compared with non-Japanese patients (Table 1).

**Table 1 Unadjusted Model**

| Parameter | Odds Ratio | 95% CI | p-value | Interpretation |
|---|---|---|---|---|
| Ethnicity | 3.27 | 1.57, 7.26 | 0.0023 | The odds of responding is over 3 times higher for Japanese patients compared to non-Japanese patients. |

CI Confidence interval.

In order to account for the observed baseline imbalances seen between Japanese and non-Japanese patients further logistic modelling was performed. This allowed odds ratios to be calculated from the model parameters, but unlike simple 2 x 2 tables the odds ratios were adjusted for all other relevant factors in the model. Therefore, the methodology allows the variation in the data to be explored further, making the assessment of the ethnic difference more sensitive and accurate.

Before the modelling was performed the data was reviewed to identify clinically meaningful baseline factors that may influence tumour response. The factors were then made into binary factors (0 or 1) or continuous factors. Each of the factors were then analysed in isolation to assess whether they were predictive of response. Those factors found to be of predictive of response at the 0.10 level were then considered in the multivariate logistical analysis. Table 2 shows the p-value for each of the parameters tested in the modelling.

**Table 2 Model Building – univariate effects**

| Parameter | p-value |
|---|---|
| Duration of previous chemotherapy treatment | 0.9553 |
| Months from diagnosis to randomisation | 0.7689 |
| Number of previous chemotherapies | 0.7372 |
| Age group (<65 years vs ≥65 years) | 0.7005 |

| Parameter | p-value |
|---|---|
| Type of disease (measurable/non-measurable) | 0.5280 |
| Stage of disease (III vs IV) | 0.4530 |
| Number of evaluable lesions at entry | 0.4342 |
| Number of measurable lesions at entry | 0.4325 |
| Progressed on a previous chemotherapy | 0.3522 |
| Time from last dose of chemotherapy to randomisation | 0.3156 |
| Visceral metastases at entry | 0.1838 |
| Previously received surgery | 0.1658 |
| Tumour burden at entry | 0.1512 |
| History of lung disorder, chest pain, dyspnoea, increased cough or haemoptysis | 0.1413 |
| Previously received docetaxel | 0.1103 |
| Baseline lung cancer subscale score | 0.0923[a] |
| Body mass index at entry | 0.0887[a] |
| Performance status | 0.0619[a] |
| Previously received radiotherapy | 0.0587[a] |
| Histology | 0.0013[a] |
| Previously received other treatment[b] | 0.0004[a] |
| Gender | 0.0003[a] |

[a] $p<0.10$: significance level for inclusion in the model (as stated in protocol).
[b] Other treatments include picibanil, investigational drugs, minomycin, marimastat and NOLVADEX.

As shown in Table 2, the baseline factors found to be predictive of response in isolation were baseline lung cancer subscale score, BMI, PS, receipt of previous radiotherapy, tumour histology, gender, and receipt of previous other treatment. Although the significance level used for model building was 0.1, as stated in the protocol, a further analysis was done using a 0.15 level to assess the robustness of the model. Using the higher threshold, two more factors were included in the logistic model (see Table 2). However, when the factors were considered in further multivariate models they were rejected at the 0.15 significance level, thus resulting in the same final model as found using a 0.1 threshold level.

The next step was to fit these seven parameters in one logistical model to assess their impact on the apparent difference seen between the ethnic groups. By incorporating this information into