**Biost/Epi 536, Homework # 3 ID # 3318**

1. Table 1 presents descriptive statistics of the 50 patients enrolled in the PSA study. Statistics are presented for patients dichotomized based on lowest values of prostate specific antigen observed post treatment (PSA Nadir) above or below 4.0 ng/ml, as normally healthy men tend to have PSA values below 4.0 ng/ml.

Table 1: Descriptive statistics of PSA study population

|  |  |  |
| --- | --- | --- |
|  | PSA nadir >= 4.0 (n=17) | PSA nadir < 4.0 (n=33) |
|  | N | Mean | SD | Range | N | Mean | SD | Range |
| Pre-treatment PSA | 15 | 938.40 | 1597.99 | 25, 4797 | 28 | 527.37 | 1092.68 | 4.8, 3946 |
| Performance score | 16 | 77.50 | 9.31 | 60, 90 | 32 | 82.50 | 11.64 | 50, 100 |
| Age | 17 | 68.71 | 6.19 | 61, 86 | 33 | 66.79 | 5.53 | 58, 81 |
|  | N | % |  |  | N | % |  |  |
| Bone scan score |  |  |  |  |  |  |  |  |
| 1 | 0 | 0.0% |  |  | 5 | 15.6% |  |  |
| 2 | 3 | 18.8% |  |  | 10 | 31.3% |  |  |
| 3 | 13 | 81.3% |  |  | 17 | 53.1% |  |  |
| Tumor grade |  |  |  |  |  |  |  |  |
| 1 | 3 | 25.0% |  |  | 7 | 24.1% |  |  |
| 2 | 5 | 41.7% |  |  | 10 | 34.5% |  |  |
| 3 | 4 | 33.3% |  |  | 12 | 41.4% |  |  |
| Relapse in 24 months | 15 | 88.2% |  |  | 7 | 21.2% |  |  |

1. **A)** When comparing two groups with the same performance and bone scan scores but different PSA nadir levels, the odds of relapse within 24 months is estimated to be 3.3% higher (odds ratio 1.033) for each 1 ng/ml difference in PSA Nadir, with the group having the higher PSA nadir level tending toward an increased odds of relapse within 24 months. This observed difference is not statistically different from an odds ratio of 1 (P =0.445), with a 95% confidence interval suggesting that the observed odds ratio is what might be typically observed if the true odds of relapse within 24 months was anywhere between 5.0% lower and 12.5% higher for each 1 ng/dl higher PSA nadir level. We therefore lack the evidence to claim an association between PSA nadir and relapse within 24 months.

This estimate was obtained from a robust logistic regression model of relapse within 24 month on PSA nadir values adjusting for performance score and bone scan score (Table 2). Because a BSS of one predicted relapse perfectly (i.e., all patients with a BSS = 1 relapsed), these 5 patients were omitted from the model, along with 2 other patients with other missing data. The omission of all patients with a BSS=1 may result in an estimate biased toward the square of the true OR.

Table 2: Robust logistic regression model estimates (n=43)

|  |  |  |
| --- | --- | --- |
| Variable | Odds Ratio (95% CI) | P value |
| PSA nadir | 1.033 (0.950, 1.125) | 0.445 |
| Performance Score | 0.956 (0.888, 1.029) | 0.228 |
| Bone Scan Score of 2 (reference category is 3) | 0.524 (0.137, 1.999) | 0.344 |
| Intercept | 26.241 (0.059, 11613.750) | 0.293 |

**B)** When comparing two groups with the same performance and bone scan scores but different PSA Nadir levels, the odds of relapse within 24 months is estimated to be 8.4% higher (odds ratio 1.084) for each 10% difference in PSA nadir, with the group having the higher PSA Nadir level tending toward an increased odds of relapse within 24 months. This observed difference is statistically different from an odds ratio of 1 (P =0.008), with a 95% confidence interval suggesting that the observed odds ratio is what might be typically observed if the true odds of relapse within 24 months was anywhere between 2.1% and 15.0% higher for each 10% higher PSA nadir level. We therefore reject the null hypothesis of no association between PSA nadir and relapse within 24 months.

This estimate was obtained from a robust logistic regression model of relapse within 24 month on log-transformed PSA nadir values adjusting for performance score and bone scan score (Table 3). Because a BSS of one predicted relapse perfectly (i.e., all patients with a BSS = 1 relapsed), these 5 patients with a BSS=1 were omitted from the model, along with 2 other patients with other missing data. The omission of all patients with a BSS=1 may result in an estimate biased toward the square of the true OR.

Table 3: Robust logistic regression model estimates (n=43)

|  |  |  |
| --- | --- | --- |
| Variable | Odds Ratio (95% CI) | P value |
| Log PSA nadir | 2.325 (1.244, 4.348) | 0.008 |
| Performance Score | 0.950 (0.881, 1.024) | 0.182 |
| Bone Scan Score of 2 (reference category is 3) | 0.478 (0.102, 2.237) | 0.348 |
| Intercept | 35.415 (0.130, 9624.005) | 0.212 |

**C)** When comparing two groups with the same performance and bone scan scores but different PSA nadir levels, the odds of relapse within 24 months is estimated to be 2589.3% higher (95% CI: 17.8% higher, 61306.9% higher) for each 1 ng/dl increase in PSA nadir between the values of 0 and 1; after which the odds of relapse within 24 months is estimated to be 9.2% lower (95% CI: 66.9% lower, 148.9% higher) for each 1 ng/dl increase in PSA nadir between the values of 1 and 4; then 37.9% higher (95% CI: 5.5% lower, 101.1% higher) for each 1 ng/dl increase in PSA nadir between the values of 4 and 16; and finally 1.8% lower (95% CI: 14.2% lower, 2.5% higher) for each 1 ng/dl increase in PSA nadir for values greater than 16. This observed difference is statistically different from equal odds (P =0.019). We therefore reject the null hypothesis of no association between PSA nadir and relapse within 24 months.

This estimate was obtained from a robust logistic regression model of relapse within 24 month on linear spline PSA nadir values adjusting for performance score and bone scan score (Table 4). Because a BSS of one predicted relapse perfectly (i.e., all patients with a BSS = 1 relapsed), the 5 patients with a BSS of one were omitted from the model, along with 2 other patients with other missing data. The omission of all patients with a BSS=1 may result in an estimate biased toward the square of the true OR.

Table 4: Robust logistic regression model estimates (n=43)

|  |  |  |
| --- | --- | --- |
| Variable | Odds Ratio (95% CI) | P value |
| PSA nadir, linear splines |  | 0.019 |
| 0-1 | 26.893 (1.178, 614.069) |  |
| 1-4 | 0.908 (0.331, 2.489) |  |
| 4-16 | 1.379 (0.945, 2.011) |  |
| 16+ | 0.982 (0.965, 1.000) |  |
| Performance Score | 0.938 (0.858, 1.025) | 0.159 |
| Bone Scan Score of 2 (reference category is 3) | 0.423 (0.092, 1.940) | 0.268 |
| Intercept | 7.880 (0.021, 2900.115) | 0.493 |

**D)** For the model in part A (table 2), the intercept is the estimated odds of relapse in 24 months for a group of patients with 0 ng/dl PSA nadir, a 0 performance score, and a BSS=3.This group of patients is possible, but outside the range of our data. For this group the odds translates into a probability of 52.5% relapse within 24 months (26.241/1+26.241=52.482).

For the model in part B (table 3), the intercept is the estimated odds of relapse in 24 months for a group of patients with a log PSA nadir of 0, a 0 performance score, and a BSS=3.This group of patients is not possible, as a log PSA nadir of 0 would mean a patient would have to have a negative infinity PSA nadir value. Nonetheless, the model estimates that for this group the probability of relapse within 24 months is 70.8% (35.415/1+35.415=70.830).

For the model in part C (table 4), the intercept is the estimated odds of relapse in 24 months for a group of patients with 0 ng/dl PSA nadir, a 0 performance score, and a BSS=3.This group of patients is possible, but outside the range of our data. For this group the odds translates into a probability of 15.8% relapse within 24 months (7.880/1+7.880=15.760).

1. **A)** When comparing two groups that did and did not relapse within 24 months but had the same performance and bone scan scores, the estimated mean PSA nadir differs by 23.361 ng/dl, with the group that did relapse tending toward higher mean PSA nadir. This observed difference is barely not statistically different from zero (P =0.050). We therefore just barely lack the evidence to claim an association between PSA nadir and relapse within 24 months. The 95% confidence interval suggests that the observed difference is what might be typical if the true difference in PSA nadir were anywhere between 0.020 and 46.702 ng/dl, with the group that relapsed tending to have the higher mean PSA nadir.

This estimate was obtained from a robust linear regression model of PSA nadir values on relapse within 24 months adjusting for performance score and bone scan score (Table 5). Two patients with missing data were omitted from the model.

Table 5: Robust linear regression model estimates (n=48)

|  |  |  |
| --- | --- | --- |
| Variable | Coefficient (95% CI) | P value |
| Relapse in 24 months | 23.361 (0.020, 46.702) | 0.050 |
| Performance Score | -0.587 (-1.929, 0.755) | 0.383 |
| Bone Scan Score (reference category is 3) |  | 0.252 |
| 1 | -6.126 (-20.109, 7.857) |  |
| 2 | -14.552 (-31.991, 2.886) |  |
| Intercept | 59.150 (-56.957, 175.258) | 0.310 |

**B)** When comparing two groups that did and did not relapse within 24 months but had the same performance and bone scan scores, the estimated geometric mean PSA nadir differed by 1272.806% ng/dl on average, with the group that did relapse tending toward higher geometric mean PSA nadir. This result is statistically different from zero (P <0.001), with a 95% confidence interval suggests that the observed difference is typical if the true difference in geometric mean PSA nadir were anywhere between 312.9519% and 4463.719% different, with the group that relapsed tending to have the higher geometric mean PSA nadir.

This estimate was obtained from a robust linear regression model of log-transformed PSA nadir values on relapse within 24 months adjusting for performance score and bone scan score (Table 6). Two patients with missing data were omitted from the model.

Table 6: Robust linear regression model estimates (n=48)

|  |  |  |
| --- | --- | --- |
| Variable | Coefficient (95% CI) | P value |
| Relapse in 24 months | 2.619 (1.418, 3.821) | <0.001 |
| Performance Score | -0.005 (-0.062, 0.052) | 0.870 |
| Bone Scan Score (reference category is 3) |  | 0.020 |
| 1 | -1.216 (-2.184, -0.247) |  |
| 2 | -0.225 (-1.267, 0.817) |  |
| Intercept | 0.026 (-4.860, 4.912) | 0.991 |

1. **A)** All models control for performance and bone scan scores, allowing us to assess the merits of PSA nadir above and beyond these existing indicators of potential relapse within 24 months. Unfortunately, the models in question 3 include a non-informative strata of BSS=1, which can result in an estimate biased toward the square of the true OR.

The logistic regression model with PSA nadir modeled as a continuous, untransformed variable (part 2a) is the easiest to interpret: it is the forward question (i.e., does PSA nadir predict relapse) and the coefficient of the independent variable is presented in the same units in which the independent variable was measured (changes in ng/dl). The unfortunate thing about this model is that it does not result in a statistically significant relationship with relapse, which we might expect to see from a scientific perspective.

The logistic regression model with log-transformed PSA nadir (part 2b) is probably the most scientifically accurate model to test for an association. I would prefer it a priori. Since PSA is a function of cancer cells, which reproduce and grow exponentially, I would expect to see a relationship on a log scale. The statistically significant relationship between PSA nadir and relapse within 24 months is evidence of this being a more accurate relationship than non-transformed values of PSA nadir.

The logistic regression model with linear splines of PSA nadir (part 2c) is probably the best model to explore the date. It allows us to observe a very strong association up to 1 ng/dl, then not much going on afterward. This could be used to identify a threshold at which PSA nadir might be highly predictive of relapse within 24 months.

The linear regression model where relapse within 24 months is used to predict mean PSA nadir values (part 3a) is an easy-to-interpret model that would be useful to identify a threshold at which PSA predicts relapse within 24 months. The modeled mean PSA of the relapse groups could be used as a PSA nadir level that is indicative of relapse. However, because of the many high positive values, the arithmetic mean may be too high and not very informative. Therefore the linear regression model where relapse within 24 months is used to predict geometric mean PSA nadir values (part 3b) would be the best to identify a threshold effect, since geometric means are less subject to influence from outliers and PSA is scientifically expected to increase on the log scale due to the nature in which cancer cells grow.

**B)** If a high level of PSA is the result of cancer cells, then the predictor of interest (PSA nadir) is really a surrogate for the response variable (relapse to cancer within 24 months). It still seems worthwhile to identify a threshold at which PSA levels are indicative of cancer, but it is only natural that a sign of cancer, such as high PSA, would predict cancer.