1. Abstract

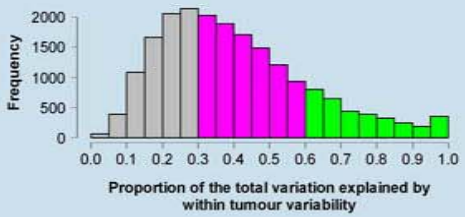
The recent advances in biology and specifically in cancer biology suggest that at least some biomarkers are heterogeneously distributed within a tumour. Our analysis of the heterogeneity of several markers (interstitial fluid pressure, level of oxygenation, carbonic anhydrase IX) analyzed in cervix cancer tissue suggested that the variability within tumour was almost as large as the variability between tumours. For example the variability within a tumour of carbonic anhydrase IX accounted for 40% of the total variability. The ultimate goal is to find associations between markers and outcome. Due to their heterogeneous distribution the association sought may appear weak or non-existent. Thus, in the literature, there is a wide range of conclusions on the presence or lack of association between a marker and an outcome. This paper presents an approach to quantifying the heterogeneity and examines how it influences the estimated effect size and the ability to detect it (power of the study). It is shown that the estimated effect size and the power of the study are diminished when the marker is measured with error. Therefore, it is essential that the first step in investigating a biomarker is to examine its variability within a tumour prior to designing the study with enough replicates to give a representative value for a tumour.

2. Are tumours heterogeneous ?

The tumour heterogeneity in gene expression profile was explored using samples from 33 biopsies taken from 11 patients (1). The genes' expressions were evaluated by means of the Affimetrix U133 Plus 2.0 oligonucleotide chip. For each gene the total variance was partitioned between variance within tumour and variance between tumours. For a perfectly homogeneous tumour, the expression of a gene in the different biopsies of the same tumour should be the same. In other words, the variability within tumour is zero. Thus a small ratio of the variability within tumour to the total variability indicates homogeneity in the expression of a gene. A larger value of this ratio is suggestive of high heterogeneity. The distribution of this ratio over the 20,000 genes was:

- 17% with ratio > 0.6
- 46% with ratio > 0.3 and ≤ 0.6
- 37% with ratio ≤ 0.3

The histogram of the ratio of variance within to the total variance.



3. Estimating the heterogeneity

A biomarker is measured in *n* tumors with *k* replicates for each tumour. The total variance can be partitioned (2) into variance within tumour (σ_W^2) and variance between tumours (σ_B^2). A measure of heterogeneity is the ratio of the within tumour variance and the total tumour variance.

If one measurement per tumour is taken, then the measure of heterogeneity is:

$$\varepsilon_1 = \frac{\sigma_W^2}{\sigma_W^2 + \sigma_B^2}$$

If, for each tumour, several measurements (say *k*) are taken, then the average value for one tumour has a smaller variance :

$$\sigma_W^2 / k$$

In this case, the measure of heterogeneity is:

$$\varepsilon_k = \frac{\sigma_W^2 / k}{\sigma_W^2 / k + \sigma_B^2}$$

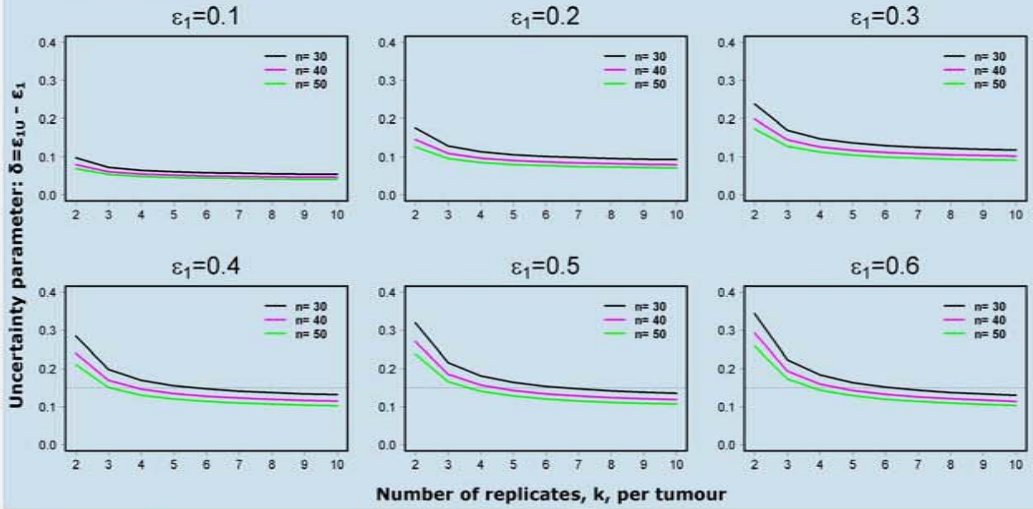
- ε_k represents the proportion of the variance due to the heterogeneity of the tumour.
  - k is chosen such that ε_k is smaller than a certain given value. For a large value of ε_1 the number of replicates, *k*, may be impractically large.
- ε_1 >0.6 is of less interest.
  - In this case, the variability within is larger then the variability between patients.
  - The number of replicates could be very large.
- Large values of ε_1 are of concern. Since ε_1 is estimated, one can calculate its 95% confidence interval: (ε_{1L} , ε_{1U}). The upper limit of the confidence interval (ε_{1U}) is of interest.

Definition: *Uncertainty parameter (δ)* as the difference between the upper limit of the 95% confidence interval (3) and the estimate.

$$\delta = \varepsilon_{1U} - \varepsilon_1$$

4. The change of the uncertainty parameter (δ) with ε_1 , n (the number of patients)

- If ε_1 =0.1, then δ <0.1 and δ does not change much with *n* and *k*
- If ε_1 >0.1, then δ decreases dramatically for lower levels of *K*: 2 to 5.
- If ε_1 =0.6, then δ > 0.1 even for n=50.
- For ε_1 =0.6 and δ =0.15, one could sample either any one of:
  - 40 tumours with 5 replicates
  - 50 tumours with 3 replicates



5. Calculating the sample size of the study when the outcome is survival

n_d = number of observed deaths
N=number of patients necessary to observe n_d during the study duration
p_d = probability to observe deaths during thr duration of the study
a=accrual time
f=the added follow-up after the end of accrual
HR = hazard ratio (the fold increase of the hazard for one unit increase in the explanatory variable)
σ = the standard deviation for the explanatory variable
$Z_{1-\beta}$ and $z_{1-\alpha/2}$ = the quantiles of the standard normal distribution.
For example for α=0.05, $z_{1-\alpha/2}$ =1.96 and for β=0.2, $Z_{1-\beta}$ =0.8
1-β = power of the study

The number of deaths necessary to be observed are:

$$\sqrt{n_d} = \frac{z_{1-\alpha/2} + z_{1-\beta}}{\sigma \log(HR)}$$

The probability for deaths to be observed during the study is:

$$p_d = 1 - \frac{e^{-\lambda f} - e^{-\lambda(a+f)}}{\lambda a}$$

The number of patients necessary to detect an effect size of HR, at the level of significance α with the power 1-β is:

$$N = \frac{n_d}{p_d}$$

6. The sample size calculation when the explanatory variable is measured with error

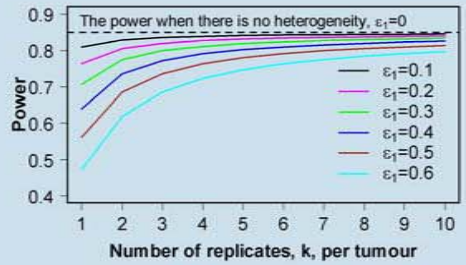
When the explanatory variable is measured with error and the measure of heterogeneity is ε_k the number of deaths necessary to observe during study (4) is given by:

$$\sqrt{n_d} = \frac{z_{1-\alpha/2} + z_{1-\beta}}{\sqrt{1-\varepsilon_k}\,\sigma_B \log(HR)}$$

Note:

- The sample size needs to be increased by a factor of $1/\sqrt{1-\varepsilon_k}$
- The standard deviation is obtained from the variance between patients.

The change in the power of the study with different ε_1 and different number of replicates



7. Example—Design of study of association between a heterogeneous biomarker and outcome

Suppose that a study is designed which seeks to investigate the importance of the interstitial fluid pressure (IFP) as a biomarker for cervix tumours. There is evidence in the literature that IFP is heterogeneous in the tumour.

Step 1. The pilot study (4).

Aim: to study the feasibility of measuring IFP in cervix tumours multiple times.

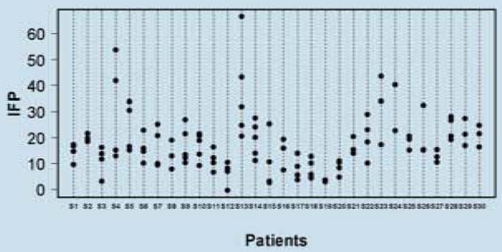
30 patients were accrued. IFP was measured as follows:

- 2 patients with 2 replicates
- 8 patients with 3 replicates
- 13 patients with 4 replicates
- 7 patients with 5 replicates

Conclusion:

- 3-4 is a reasonable number of replicates
- The accruing rate is 20 patients/year
- Based on the analysis of the heterogeneity ε_1 = 0.59

The within patient variability of IFP values for the pilot study



Step 2. Planning the association study.

Assumptions for calculating the sample size

- Patients will be accrued over 5 year period
- 2 more years of follow-up will be added at the end of the study

8. Conclusions

- Many biomarkers are heterogeneous
- The heterogeneity needs to be measured before designing the study for the association with outcome.
- The measurement needs to be taken in replicates to reduce the variability within an individual
- The ratio of the variability within individual needs to be accounted for when the association study is designed.