

# Sampling design optimization for EROD measurements in fish

B. Beliaeff, T. Burgeot\*

IFREMER, BP 21105, F-44311 Nantes Cedex 3, France

**ABSTRACT:** Activity of 7-ethoxyresorufin-O-deethylase (EROD) in fish is certainly the best-studied biomarker of exposure applied in the field to evaluate biological effects of contamination in the marine environment. Since 1991, a feasibility study for a monitoring network using this biomarker of exposure has been conducted along French coasts. Using data obtained during several cruises, this study aims to determine the number of fish required to detect a given difference between 2 mean EROD activities, i.e. to achieve an *a priori* fixed statistical power ( $1-\beta$ ) given significance level ( $\alpha$ ), variance estimations and projected ratio of unequal sample sizes ( $k$ ). Mean EROD activity and standard error were estimated at each of 82 sampling stations. The inter-individual variance component was dominant in estimating the variance of mean EROD activity. Influences of  $\alpha$ ,  $\beta$ ,  $k$  and variability on sample sizes are illustrated and discussed in terms of costs. In particular, sample sizes do not have to be equal, especially if such a requirement would lead to a significant cost in sampling extra material. Finally, the feasibility of long-term monitoring is discussed.

**KEY WORDS:** Biomonitoring · Estimation · EROD activity · Statistical power · Sample size · Sampling design

## INTRODUCTION

A number of studies performed in the early 1970s indicated the reliability of cytochrome P450 enzymes in fish liver as an indicator of environmental contamination (e.g. Payne 1976). The cytochrome P4501A is the main form which can be induced by polyaromatic hydrocarbons (PAH), planar polychlorinated biphenyls (PCB) and dioxins. Changes in the activity of enzymes such as 7-ethoxyresorufin-O-deethylase (EROD) in marine species can also be used as biochemical markers of exposure to such chemicals in the ocean environment (Addison & Payne 1986, Payne et al. 1987). Since 1991, a long-term biomonitoring program using EROD activity measurement in fish has been developed along the French coasts. The development and validation of this biomarker have been achieved on French coasts (Galgani et al. 1991, 1992, Burgeot et al. 1994a, b, c) and in the North Sea (Eggens & Galgani 1992) for monitoring applications. The method

described by Payne & Galgani (1991) was used for EROD assays on target species fished along the French coasts.

There are 2 well-known types of error in environmental monitoring. A type II error takes place if absence of impact is reported when impact really occurred. It is apparent that failure to take such an error into account when designing an environmental monitoring program may have disastrous consequences. The probability that this error does not occur is defined as the statistical power  $1-\beta$ . Conversely, detecting an impact when in fact it does not exist is the so-called type I error, whose probability of occurrence is the significance level ( $\alpha$ ) of the appropriate statistical test.

The number of fish required to detect a significant difference between means (for instance) is an increasing function of the statistical power  $1-\beta$ . It also depends on  $\alpha$ , on the difference we want to be able to detect, and on the variability inherent in the data (Faireweather 1991, Nicholson & Fryer 1992).

The 3 objectives of this study are (1) to give correct estimates of EROD activity mean and standard error,

\*Addressee for correspondence. E-mail: tburgeot@ifremer.fr

(2) to give numbers of fish required to detect a difference between 2 mean values for different combinations of  $\alpha$ ,  $\beta$ , ratio of possible unequal sample sizes depending on trawl-fish condition and inter-individual variability, and (3) to show that these results can be beneficial in the context of biological effects field studies.

## MATERIAL AND METHODS

**Data collection.** Initially, a pilot study aimed at identifying sampling stations in the most polluted French estuaries, as well as the species available at those sites was conducted (Burgeot et al. 1994a). In accordance with ICES recommendations (ICES 1992), we chose very differently marketable species; non-commercial species offering *a priori* greater resource stability were also selected. In the Seine Bay, the dab *Limanda limanda* is a marketable reference species used for North Sea monitoring, although little information is available concerning its spatial distribution and stock status in this area. The sole *Solea solea*, another marketable species, proved suitable for monitoring of biological effects in the southern part of the Bay of Biscay, the only area where it was collected during our field surveys. The dragonet *Callionymus lyra*, a non-commercial species (a reference species in the OSPAR area), is available along the entire English Channel and Atlantic coasts of France. In the Mediterranean, the red mullet *Mullus barbatus*, a marketable species, is largely distributed along the coast. The combers *Serranus hepatus* and *Serranus cabrilla*, less frequent commercial species, are of biological interest because they are hermaphrodites, and thus could take into account possible EROD variations due to sex.

Eight oceanographic cruises were carried out: 2 within the 30 mile limit in the Seine Bay (Eastern Channel) during March and September 1992 and again in 1993, 1 in the Bay of Biscay (Gironde estuary) in November 1992, 2 off Marseille/Fos sur Mer (Mediterranean) during May and October 1993 and 1 on the south coast of Brittany (Britton Abers) in 1993 (Fig. 1). Large vertical-opening bottom trawl net fishing was performed during the day for 30 min at each station. When available, 10 trawl-fished specimens were measured and sexed. All fish were sacrificed directly after trawling. Their livers were removed immediately, washed with buffer (Tris 50 mM, KCl 150 mM, pH 7.4), frozen in liquid nitrogen and analyzed in the laboratory.

**Analytical protocol. Extract preparation:** Liver was washed in buffer (Tris 50 mM, pH 7.4; KCl 150 mM; EDTA 1 mM; glycerol 20% vol. and 1 mM dithiothreitol) at 4°C, minced (5 ml of buffer per g of tissue)

for 5 to 10 s in a Potter-Elvehjem tube and then centrifuged at  $9000 \times g$  for 15 min at 4°C. Supernatant was used as the source of enzyme.

**Protein analysis:** Proteins were measured according to the method of Bradford (1976), using mad bovine serum albumin as protein standard. Measurements were performed using a plate-reading spectrophotometer at 595 nm and expressed as  $\text{mg ml}^{-1}$ . Then, the protein quantity,  $Y$  in mg, was calculated as:

$$Y = \frac{1}{10^3} \cdot \frac{V}{V_p} \cdot f \cdot \frac{d - a_p}{b_p} \quad (1)$$

where  $d$  is the optical density for a given well,  $V$  is the well volume (in  $\mu\text{l}$ ),  $f$  the dilution factor,  $V_p$  the extract volume (in  $\mu\text{l}$ ) for protein analysis,  $b_p$  the slope of the optical density calibration curve as a function of protein quantity (in  $\mu\text{g}^{-1}$ ), and  $a_p$  is the intercept.

**Analysis of EROD activity:** A rapid microplate EROD enzymatic assay method (Grzebyk & Galgani 1991) was used. Four replicates per fish allowed analysis of 10 individuals from the same station on the same microplate. With the fluorimetric microplate method we estimated that the analytical repeatability was equal to 7%. In the following calculations the 4 measurements corresponding to 4 wells were averaged by fish. Laboratories using this method were inter-compared during an exercise organized by the North Sea Task Force in 1991 (Stagg & Addison 1995) and the method was adopted by the International Council for the Exploration of the Sea (ICES) for monitoring in the North Sea (Payne & Galgani 1991, Stagg & McIntosh 1997).

The EROD activity during  $\tau$  minutes ( $\text{pmol min}^{-1}$ ) is given by

$$X_\tau = \frac{F_{t+\tau} - F_t}{\tau \cdot V_e \cdot b_e} \quad (2)$$

The EROD activity ( $A^i$ ) relative to protein quantity ( $\text{pmol min}^{-1} \text{mg}^{-1}$  of protein) for the  $i$ th fish is given by

$$A^i = \frac{X_\tau^i}{Y^i} = \frac{10^3 \cdot V_p \cdot b_p}{\tau \cdot V_e \cdot b_e \cdot V \cdot f} \cdot \frac{F_{t+\tau}^i - F_t^i}{d^i - a_p} \quad (3)$$

where  $F_t$  is the fluorescence at time  $t$  for a given well [in fluorescence units (FU)],  $\tau$  is the time-delay between 2 measurements in the same well (in min),  $V_e$  is the extract volume for the enzymatic activity assay (in  $\mu\text{l}$ ), and  $b_e$  is the slope of the fluorescence calibration curve as a function of enzyme concentration (in  $\text{FU } \mu\text{l} \text{pmol}^{-1}$ ).

**Statistical analysis. Comparison of 2 EROD activities:** Enzymatic activity is assumed to follow a log-normal distribution (Clarke & Green 1988, Krawczak-Krogulecka et al. 1995). Therefore a logarithmic (base 10) transformation of EROD activity data was performed to assure an approximate normal distribution.

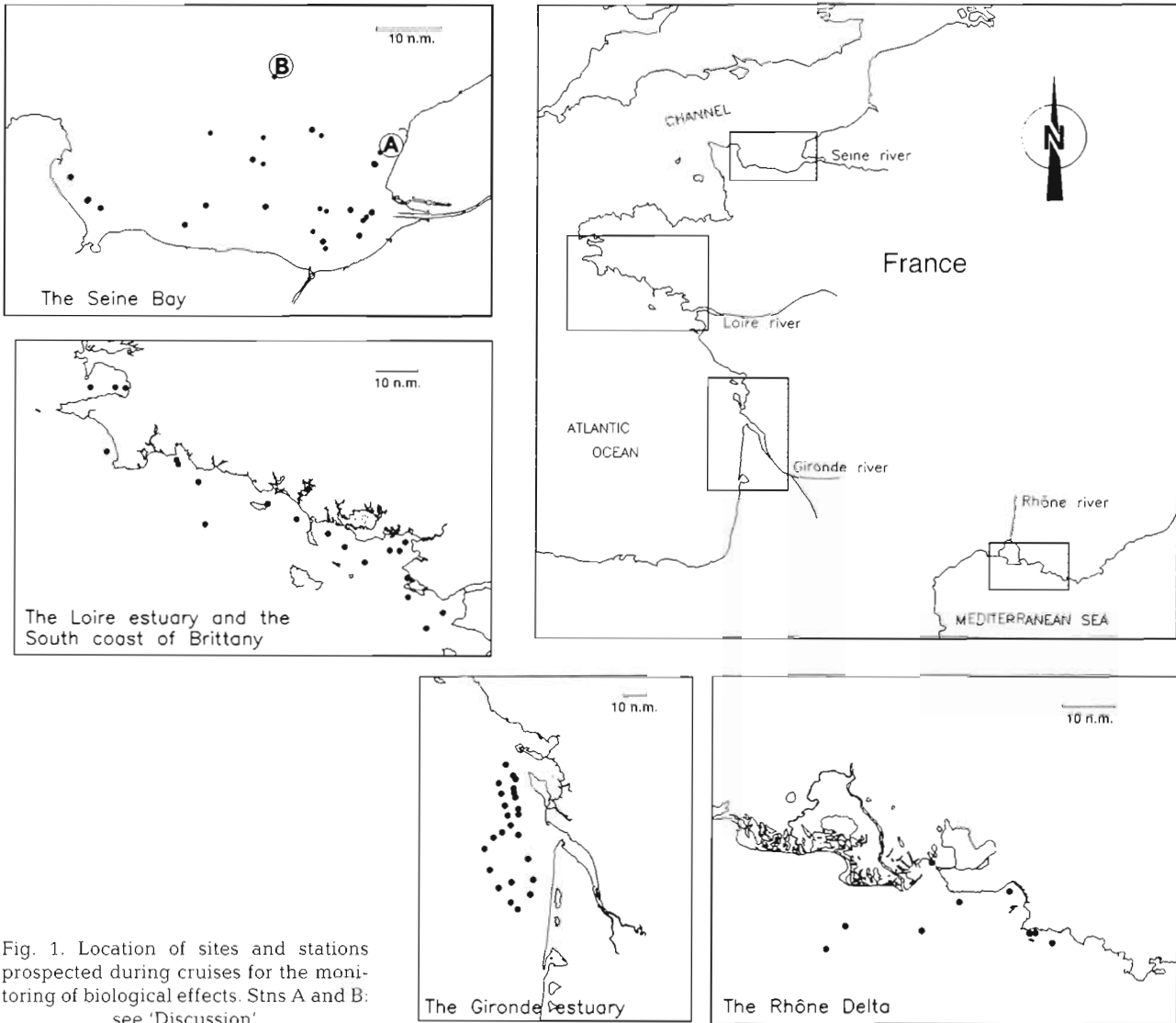


Fig. 1. Location of sites and stations prospected during cruises for the monitoring of biological effects. Stns A and B: see 'Discussion'

As  $A$  is assumed to be distributed normally, a  $t$ -test can be used to compare EROD activity means, say  $m_1$  and  $m_2$ , obtained independently at 2 sampling stations, in the case of equal variances. The number of degrees of freedom (df)  $v = n_1 + n_2 - 2$ , where  $n_1$  and  $n_2$  are the numbers of fish used respectively at Stns 1 and 2.

Assuming that the 2 unknown true means are equal,  $t_v$  follows a  $t$ -distribution with  $v$  df, and is compared to the observed  $t$  value at a significance level of  $\alpha$ . This procedure also holds in the case where the unknown population variances are unequal. Sokal & Rohlf (1981) give the approximate number of degrees of freedom when the variances are unequal (i.e. in the well-known case called 'Behrens-Fisher problem').

**Sample size required:** Suppose we want to detect a difference  $\delta$  ( $\delta = \mu_1 - \mu_2$ ) in EROD activity between 2 sample means stations, at an  $\alpha$  significance level (where  $\mu_1$

and  $\mu_2$  are the respective means). If this difference exists we want to detect it with a probability  $1 - \beta$ . Then the required number of fish at each station is (Sokal & Rohlf 1981):

$$n \geq 2 \left( \frac{s^{*2}}{\delta} \right)^2 (t_{v, 1-\alpha/2} + t_{v, 1-\beta})^2 \quad (4)$$

with

$$s^{*2} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

where  $s^{*2}$  is the pooled sample variance,  $s_1^2$  and  $s_2^2$  are respectively the sample variance for sample 1 and 2, and  $t_{v, 1-\alpha/2}$  and  $t_{v, 1-\beta}$ , respectively, are the values of the  $t$ -distribution for the corresponding number of degrees of freedom and probability values (subscripts). We can adapt Eq. (4) to the case where different sizes of samples can be projected. For example, Let  $n_2 = kn_1$ . In this

Table 1. Ranges of estimates of EROD activity mean and standard error (Min, Max) by species of fish (pmol min<sup>-1</sup> mg<sup>-1</sup> of protein; log<sub>10</sub> transformed values) from *N* sampling stations

	Mean	Standard error	<i>N</i>
Dragonet	0.912, 2.056	0.025, 0.312	48
Dab	0.721, 1.928	0.045, 0.469	13
Comber	-1.238, 0.985	0.043, 0.106	5
Mullet	0.821, 1.771	0.036, 0.078	9
Sole	1.319, 1.800	0.058, 0.145	7

case it can easily be shown that the required size for sample 2 is:

$$n_2 \geq (1+k) \left( \frac{s^*}{\delta} \right)^2 (t_{v,1-\alpha/2} + t_{v,1-\beta})^2 \quad (5)$$

and  $n_1 = n_2/k$ . It is straightforward to note that Eq. (5) holds when  $k = 1$ , i.e. for equal sample sizes.

A FORTRAN program was developed for the different calculation needs. The 'Numerical Algorithm Group'<sup>®</sup> numerical library was used to calculate the optimal number of fish  $n_2$  in Eq. (5). To this end, an iterative procedure was used as the *t* statistic is a function of the sample size.

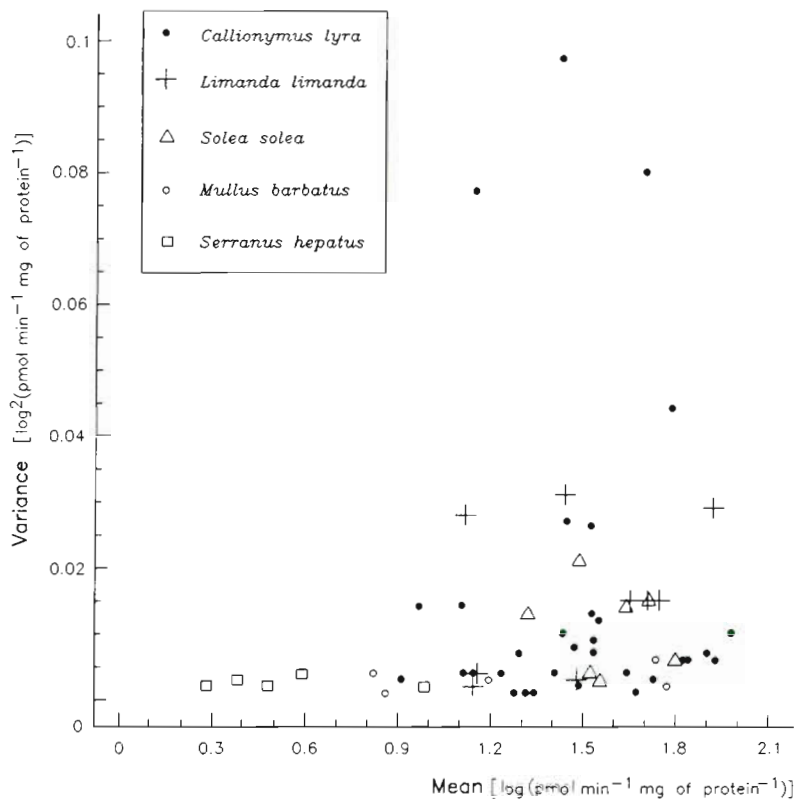


Fig. 2. Plot of EROD activity variance of the estimator function against the corresponding mean at one sampling station

## RESULTS

Results from 82 sampling stations were used for all target species. Table 1 provides the ranges of means and standard error for all the target species. The highest standard error values for samples of 10 fish were obtained for the dragonet and the dab. Lower variance values were noted for Mediterranean species, with particularly low EROD activity values for the comber. No relationships could be found between the mean of log-transformed EROD activity at one sampling station and the corresponding variance of the estimator (Fig. 2). This would justify the use of the logarithmic transformation for variance homogeneity.

Table 2 gives the number of fish required to detect a given significant difference  $\delta$  between 2 independent EROD samples, for different combinations of  $\alpha$ ,  $\beta$ , standard errors  $SE_1$  and  $SE_2$ , and projected ratios  $k = n_2/n_1$ . Table 2 should be seen as an operational tool, giving practical information based on realistic parameter values. We choose  $\delta$  to take the values 0.3, 0.5, 0.7 and 1.0, roughly corresponding, after back-transformation, to factor values 2, 3, 5 and 10 (most commonly obtained in field studies) between 2 EROD activities on an arithmetic scale.  $\alpha$  took the values 0.05 and 0.20, while values of  $\beta$  were 0.1 and 0.2. Values of standard errors were chosen according to Fig. 1. 0.08 and 0.16 can be considered respectively as low and high standard errors, while 0.32 might correspond to a limit for exceptionally high values, on the basis of our observations. To compute  $s^{*2}$ ,  $s_1^2$  and  $s_2^2$  were calculated, respectively, from  $SE_1$  and  $SE_2$  with  $n = 10$  fish, as generally used for our observations.  $n_1$  and  $n_2$  were rounded to the upper integer with  $k = 1, 1.25, 1.5, 1.75$  and 2.

Not every possible result was listed in Table 2, which illustrates impacts of parameter changes on the required number(s) of fish per sampling station. Cases 1 to 5 show influence of variance. Cases 5 to 8 show influence of  $\delta$ . Cases 7 and 16, on the one hand, and Cases 7 and 9, on the other hand, show the sensitivity of results to, respectively, power and significance level. More on the interaction between  $\alpha$  and  $\beta$  is shown by comparing Case 9 with Case 10, and Case 10 with Case 16. Cases 7 and 11 to 14 and Cases 6 and 15 illustrate the influence of  $k$ . Finally, Cases 16 to 25 correspond to all possible combinations of parameters leading to  $n_1 = n_2 = 10$  fish.

In each case, the hypothesis of variance homogeneity was tested and could never be rejected at a 5% risk level.

Table 2. Number of fish ( $n_1$ ,  $n_2$  with  $n_2 \approx kn_1$ ; the number of fish  $n_1$  is calculated through an iterative algorithm,  $n_1$  and  $n_2$  were rounded to the upper integer) required to detect with power  $1-\beta$  and significance level  $\alpha$  a given difference  $\delta$  (values are  $\log_{10}$ -transformed; an approximate back-transformed value is given in parentheses) between 2 EROD activity means estimated with respective standard errors  $SE_1$  and  $SE_2$ . (Each case gives practical information based on realistic parameter values)

Case	$SE_1$	$SE_2$	$1-\beta$	$\alpha$	$\delta$	$k$	$n_1$	$n_2$
1	0.32	0.32	0.8	0.20	1.0 (10)	1	11	11
2	0.16	0.32	0.8	0.20	1.0	1	8	8
3	0.08	0.32	0.8	0.20	1.0	1	7	7
4	0.16	0.16	0.8	0.20	1.0	1	4	4
5	0.08	0.16	0.8	0.20	1.0	1	3	3
6	0.08	0.16	0.8	0.20	0.7 (5)	1	5	5
7	0.08	0.16	0.8	0.20	0.5 (3)	1	8	8
8	0.08	0.16	0.8	0.20	0.3 (2)	1	18	18
9	0.08	0.16	0.8	0.05	0.5	1	12	12
10	0.08	0.16	0.9	0.05	0.5	1	16	16
11	0.08	0.16	0.8	0.20	0.5	1.25	7	9
12	0.08	0.16	0.8	0.20	0.5	1.5	7	10
13	0.08	0.16	0.8	0.20	0.5	1.75	7	12
14	0.08	0.16	0.8	0.20	0.5	2	7	13
15	0.08	0.16	0.8	0.20	0.7	2	4	8
16	0.08	0.16	0.9	0.20	0.5	1	10	10
17	0.08	0.08	0.9	0.05	0.4 (2.5)	1	10	10
18	0.08	0.16	0.8	0.10	0.5	1	10	10
19	0.08	0.16	0.9	0.10	0.6 (4)	1	10	10
20	0.08	0.32	0.8	0.20	0.8 (6.3)	1	10	10
21	0.08	0.32	0.8	0.10	0.9 (8)	1	10	10
22	0.16	0.16	0.8	0.05	0.7	1	10	10
23	0.16	0.16	0.9	0.05	0.8	1	10	10
24	0.16	0.32	0.8	0.10	1.0	1	10	10
25	0.16	0.32	0.9	0.20	1.0	1	10	10

## DISCUSSION

It is a classical statistical result that the sample size required to achieve a significant  $t$ -statistic value when comparing 2 means increases with noise level and the desired value of power, and decreases with significance level and magnitude of the impact to be detected. Our results (Table 2) allow a quantitative assessment of sample size sensitivity to these parameters.

Though the present work may be considered as a first phase in sampling optimization for monitoring of biological effects in the marine environment, it is not intended to prejudice the environmental significance of observations. According to the National Research Council (1990), 'whether changes in the environment are statistically significant has no bearing on the extent to which the changes may be either meaningful or important (i.e. have ecological or human consequences)'. Thus, a basic question concerns the magnitude of the effect to be detected, say  $\delta$ . It should be identified *a priori* for purposes of power studies (Camacho & Vascotto 1991). Here species- and even site-specificity are likely to make this identification very difficult. Again, from our field experience we have to decide whether a factor of 3 as here for the dragonet between reference site and impacted site could correspond to a biological

effect. Similar values were obtained in the North Sea studies which showed a pollution gradient and lower activities at the coastal sites (Stebbing et al. 1992). Factors of 1.5 to 2 and of 5 in EROD activity were observed between coastal sampling stations and offshore locations, respectively, by Eggens (1996) for flounder *Platichthys flesus* and dab *Limanda limanda* and by Sleiderink et al. (1995) for dab.

There is a difference between the number of fish used in a  $t$ -test for comparing 2 observed EROD activity means, as is usually done in field situations, and the number of fish required to detect an *a priori* given impact, the magnitude of which would be the difference between the 2 preceding means, using a  $t$ -test. It corresponds to the degree of confidence we have in non-significant results. To illustrate this point we present an actual ecotoxicological example.

The objective of our survey in the Bay of Seine was the identification of the impact of contaminants, through measurement of exposure bio-effects, between a suspected impacted station (A) located in the petroleum harbour of Antifer (Seine Maritime, France) subject to PAH contamination, and an *a priori* reference station (B) offshore (Fig. 1). This can be seen as the simplest form of a gradient study, where we expected an impact of high magnitude. A difference between the 2 log-transformed means of 0.55 was

found to be significant ( $\alpha = 0.05$ ). This difference corresponds to approximately a factor 3.5, after back-transformation, between Stns A and B. Standard errors were 0.06 for Stn B and  $0.17 \text{ pmol min}^{-1} \text{ mg}^{-1} \text{ protein}$  for Stn A, a high value according to our empirical variability scale, and respective sample sizes were 3 and 9. These numbers were not defined *a priori* from a sampling design, but resulted from a certain number of trawling operations, integrating oceanographic constraints. Power studies can help in designing similar surveys. Let us consider  $0.5 \text{ pmol min}^{-1} \text{ mg}^{-1} \text{ protein}$  (approximately a factor 3 after back-transformation) to be our impact target. With standard errors close to those in the above example Case 11 shows that 9 and 7 fish, with  $k = 1.25$  and  $\alpha = \beta = 0.2$ , are required to detect this difference, in an average of 80% of the cases. The difference in the sample sizes between the 2 points of view (comparison between 2 means and power study) correspond to the degree of confidence we have in finding no significant difference between Stns A and B. This is the price to pay if we want to be wrong only 20% of the time, when saying there is no significant impact on Stn A relative to Stn B. Of course, in the first case, a difference could be detected but without any control. Taking past observations into account, power studies provide reasonable sample size guesses, after having defined acceptable error levels. These error levels are also a function of an *a priori* knowledge of site contamination. For instance, suppose we want to be able to detect a significant impact in a known highly contaminated area, like an industrialized estuary. In this case, power can be increased to e.g. 0.9 and Table 2 might give an idea of sample sizes required. Consequently this approach allows us to modulate sampling effort, using all *a priori* information.

Values of  $\alpha$  and  $\beta$  could ideally be pre-defined if the costs relative to these 2 errors were known. In certain cases, costs relative to  $\alpha$  could be estimated: e.g. the cost of reducing inputs for pollutants, assuming that the biomarker level is clearly associated with an exposure. However, estimating the 'environmental' cost due to some organic contamination seems difficult. In the context of environmental preservation we should be more conservative, on  $\beta$ , but again without knowing the actual degree of conservatism. On the other hand, it is our choice to be more liberal with respect to  $\alpha$ . More concretely, the combination allows detection of a factor of 10, after back-transformation, between 2 EROD activity means with 11 fish in a case of extreme variability (Case 1), 8 fish with high standard error values (Case 2), and only 3 fish for still realistic variability values (Case 5). Comparison of Cases 7 and 9 shows the impact of being conservative on significance level. It is our field experience, shared with other research groups, that catching 12 fish instead of 8 at least might

require extra trawlings, and at worst is impossible with regard to fish availability. Case 10 shows the high amount of effort required in monitoring applications, if a too high  $\beta$  value is targeted.

We understand recommendations to give some order of magnitude to number of fish, e.g. 10 fish (ICES 1996). For practical reasons, sample sizes were given around this value. To get samples of 10 fish should not be a necessary target for at least 2 reasons. Firstly, using unequal sample sizes, although not optimal, allows a certain amount of flexibility (see examples in Table 2). Again, it is better to have a rough idea *a priori* of the order of magnitude to be detected. Case 15 illustrates this point where a difference  $\delta = 0.7$  (factor 5 after back-transformation) can still be detected at  $\alpha = 0.2$  significance level, with sample sizes of 4 and 8 fish. Secondly, it is a realistic way of taking into account fish availability and sampling constraints. It would be a mistake to spend a great deal of energy, and consequently money, trying to obtain extra material to reach some mythical number of samples, especially when power studies show that it might not be necessary.

When material is not abundant, considerable fishing effort, i.e. several runs, can be required to obtain the number of fish specified by the survey protocol. Therefore the total catch may be used with the risk of a very heterogeneous sample (e.g. 9 very close sizes but with much smaller individuals). The analytical cost of an individual EROD activity measurement is easily quantifiable. However, this is not the case for an estimation of the cost related to increased fishing effort at a given station, inasmuch as there is a daily inescapable fixed cost connected with the use of a ship by a crew. Nevertheless, a trawl was estimated to be 50 times more expensive than the cost of analysis (microplate method for EROD measurement) for 10 fish. However, the fishing effort not devoted to obtaining additional fish could, for example, be reallocated to intensifying the spatial sampling grid.

We have seen that some variables could be controlled *a priori* before sampling. Nevertheless, the variability inherent in the data remains the most hypothetical one. Therefore, close attention should be addressed to this sampling stage. Indeed, inter-individual variability was the major variance component of EROD activity mean. In this study, analytical variability was estimated through subsampling, and had little influence on standard error values, using simulated numbers of subsamples per fish. Highest standard error values were estimated for dab and dragonet (Fig. 2) However the number of sampling stations represents 75% of the total for these 2 species (Table 1). Thus, it is more likely, especially for dragonet, that large or even extreme standard error values will occur, which in our

case were often related to the presence of an atypical fish presenting a much lower EROD activity value than the other fish in the same sample. Consequently, it is easier to detect a significant difference between 2 EROD activities when the batch of fish used is not too heterogeneous. Like several other authors in field studies (Cooreman et al. 1993, Sleiderink et al. 1995), we could not find significant influence of adult fish size or sex on EROD activity for dragonet and dab.

As inter-individual variation was the major component of total variability for EROD activity measured at a given station, it is advisable to use a higher number of fish if no increase in fishing effort is required. Although it is advisable to reduce sample heterogeneity, i.e. inter-individual variability, this is subject to a strong availability constraint for long-term monitoring. In this context, evaluating the spatio-temporal representativeness of EROD activity measurement at a station is an essential phase in the study of the feasibility of long-term monitoring with respect to exposure levels and trends (Kwiatkowski 1991). In particular, the idea of representativeness of a prospected zone raises some questions with respect to the fishing effort performed, the migratory capacities of the species considered and estimations of exposure levels through measurements of specific contaminants in water or sediment.

## CONCLUSION

Proposition of significance level and statistical power values is essential to sampling design, and a help for decisions in the field with respect to surveying an area in order to obtain some predetermined number of fish. We showed that the effort to obtain extra fish might not be justified, considering the impact to be detected, with a more flexible sampling design allowing unequal sample sizes, and close attention paid to fish batch homogeneity.

Exploratory studies still need to be carried out in the context of the feasibility of long-term monitoring of biological effects. In particular, the biannual data (dates chosen outside reproduction periods) available at certain sites are inadequate for seasonality estimation. A higher frequency would be required for experimental purposes to obtain a reasonable estimation of intra-annual variability in relation to the seasonal cycles of biological material. Ideally, environmental and biological interactions should be assessed before an exposure effect is attributed in any precise way to pollutants. The influence of factors such as temperature, food and the reproductive period on mixed-function oxidases, a complex multi-functional enzymatic system, has been demonstrated (McMaster et al. 1991, Cooreman et al. 1993).

Achievement of a more efficient diagnosis of environmental conditions could be provided by a long-term monitoring scheme involving the combined use of several biological markers and measurement of chemicals. In this case, the optimal number of fish will be a compromise generated by power studies defined for each biomarker

*Acknowledgements.* The authors gratefully acknowledge Philippe Gros for valuable comments on the manuscript, M. J. Thébaud for secretarial assistance and Louis Giboire for drawing the figures.

## LITERATURE CITED

- Addison RF, Payne JF (1986) Assessment of hepatic mixed function oxidase induction in winter flounder (*Pseupleuronectes americanus*) as a marine petroleum pollution technique. *Can Tech Rep Fish Aquat Sci* 1505
- Bradford M (1976) A rapid method for quantification of protein utilizing the principle of dye binding. *Anal Biochem* 72:248–264
- Burgeot T, Bocquené G, Pingray G, Clodefroy D, Legrand J, Dimeet J, Marco F, Vincent F, Henocque Y, Oger-Jeanerret H, Galgani F (1994a) Monitoring biological effects of contamination in marine fish along French coasts by measurement of Ethoxyresorufin-*O*-deethylase activity. *Ecotoxicol Environ Safe* 29:131–147
- Burgeot T, Bocquené G, Truquet P, Le Dean L, Galgani F (1994b) Induction of EROD activity in red mullet (*Mullus barbatus*) along the French Mediterranean coasts. *Sci Total Environ* 142:213–220
- Burgeot T, Bocquené G, Truquet P, Le Dean L, Poulard JC, Dorel D, Souplet A, Galgani F (1994c) The dragonet *Callionymus lyra*, a target species used for evaluation of biological effects of chemical contaminants on the French coasts. *Mar Ecol Prog Ser* 97:309–316
- Camacho F, Vascotto GL (1991) Framework for enhancing the statistical design of aquatic environmental studies. *Environ Monit Assess* 17:303–314
- Clarke KR, Green RH (1988) Statistical design and analysis for a 'biological effects' study. *Mar Ecol Prog Ser* 46:213–226
- Cooreman K, Roose P, Vyncke W (1993) EROD monitoring in dab from the Belgian continental shelf. *Comm Meet Int Counc Explor Sea, ICES-CM* 1993/E:14
- Eggens ML (1996) Cytochrome P4501A induction in North sea flatfish as biomarker of exposure to dioxin-type compounds. PhD thesis, University of Utrecht
- Eggens ML, Galgani F (1992) Ethoxyresorufin-*O*-deethylase (EROD) activity in flatfish: fast determination with a fluorescence plate-reader. *Mar Environ Res* 33:213–221
- Faireweather PG (1991) Statistical power and design requirements for environmental monitoring. *Aust J Mar Freshwater Res* 42:555–567
- Galgani F, Bocquené G, Luçon M, Grzebyk D, Letrouit F, Claisse D (1991) EROD measurement in fish from the northwest part of France. *Mar Pollut Bull* 22:494–500
- Galgani F, Bocquené G, Truquet P, Burgeot T, Chiffolleau JF, Claisse D (1992) Monitoring of pollutant biochemical effects on manne organisms of the French coasts. *Oceanol Acta* 15:355–364
- Grzebyk D, Galgani F (1991) Measurement of the effect of organic pollution on marine organisms: rapid determination of EROD induction using plate readers. *Aquat Living Resour* 4:53–59

- ICES (1992) Report of the working group on biological effects of contaminants. 30 April–8 May. Copenhagen, Denmark. Comm Meet Int Counc Explor Sea, ICES-CM 1992/E:5
- ICES (1996) Report of the working group on biological effects of contaminants. 4–7 March. Ostend, Belgium. Comm Meet Int Counc Explor Sea, ICES-CM 1996/E:5
- Krawczak-Krogulecka W, Draganik B, Bykowski A, Kuczynski P (1995) Seasonal variation of the EROD activity measured in the flounder collected from the shallow waters of Poland. Comm Meet Int Counc Explor Sea, ICES-CM 1995/E:17. Marine Environmental Quality Committee
- Kwiatkowski RE (1991) Statistical needs in national water quality monitoring programs. Environ Monit Assess 17: 253–271
- McMaster ME, Van der Kraak GJ, Portt CB, Munkittrick KR, Sibley PK, Smith IR, Dixon DG (1991) Changes in hepatic mixed-function oxygenase (MFO) activity, plasma steroid levels and age at maturity of white sucker (*Catostomus commersoni*) population exposed to bleached kraft pulp mill effluent. Aquat Toxicol 21:199–218
- National Research Council (1990) Managing troubled waters. The role of environmental monitoring. National Academy Press, Washington
- Nicholson MD, Fryer RJ (1992) The statistical power of monitoring programmes. Mar Pollut Bull 24:146–149
- Payne JF (1976) Field evaluation of benzopyrene hydroxylase induction as a monitor for marine pollution. Science 191: 945–946
- Payne JF, Fancey A, Rahimtula D, Porter EL (1987) Review and perspectives on the use of mixed function oxygenase enzymes in biological monitoring. Comp Biochem Physiol 86C:233–245
- Payne JF, Galgani F (1991) Biological effects of contaminants: microplate method for measurement of ethoxyresorufin-O-deethylase (EROD) in fish. ICES. Technique in Marine Environmental Sciences, No. 13
- Sleiderink HM, Oostingh I, Goksoyr A, Boon JP (1995) Sensitivity of cytochrome P450 1A induction in Dab (*Limanda limanda*) of different age and sex as a biomarker for environmental contaminants in the Southern North Sea. Arch Environ Contam Toxicol 28:423–430
- Sokal RR, Rohlf FJ (1981) Biometry. Freeman Company, New York
- Stagg R, Addison R (1995) An interlaboratory comparison of measurements of ethoxy-resorufin-O-deethylase activity in dab (*Limanda limanda*) liver. Marine Environ Res 10: 93–108
- Stagg R, McIntosh A (1997) Determination of CYP1A-dependent monooxygenase activity in the liver of the dab (*Limanda limanda*) by the fluorometric measurement of 7-ethoxyresorufin-O-deethylase (EROD) activity. Technique in marine environmental science. ICES, Copenhagen (in press)
- Stebbing ARD, Dethlefsen V, Carr M (eds) (1992) Biological effects of contaminants in the North Sea. Results of the ICES/IOC Bremerhaven workshop. Mar Ecol Prog Ser 91

This article was submitted to the editor

Manuscript first received: September 4, 1996

Revised version accepted: April 15, 1997