Eur J Clin Pharmacol (1992) 43: 235–244

European Journal of Clinical Pharmacology

Surrogate endpoints: a basis for a rational approach

J.-P. Boissel, J.-P. Collet, P. Moleur, and M. Haugh

Unité de Pharmacologie Clinique, Hôpital Neurocardiologique, Lyons, France

Received: June 27, 1991/Accepted in revised form: February 12, 1992

Summary. In clinical trials, the clinical endpoint is often replaced by an intermediate endpoint, known in some instances as a "surrogate" endpoint. The reasons for the substitution are often both practical and financial. At present, no theoretical basis or practical guidelines exist to help in the choice of surrogate endpoints.

An approach is proposed here, based on three provisos which can be verified using one of a series of equations, if sufficient data on the pathophysiology and epidemiology of the disease are available. It is shown that even a strong statistical correlation is not a sufficient criterion for the definition of a surrogate endpoint.

It is apparent that results obtained with the commonly used "surrogate" endpoints should be cautiously considered, and that the assessment of treatments should, when possible, be based on clinical rather than intermediate endpoints.

Key words: Surrogate endpoints, Clinical trials; risk/benefit ratio, assessment

Endpoints in clinical trials

In clinical trials the use of clinical endpoints is often not feasible, and intermediate or surrogate endpoints [1] have to be chosen to replace the "real", clinical endpoint [2]. For many years this practice has been widespread and has enabled the pharmaceutical industry to save both time and money in their drug development programmes [3]. However, many questions have recently been posed concerning the suitability of currently used surrogate endpoints [1].

An intermediate endpoint can be defined as a response variable which is statistically correlated with the clinical endpoint. It can qualify as a surrogate endpoint if it can be used as an appropriate alternative to a clinical endpoint in a clinical trial. This definition is somewhat more restrictive than that proposed by others [4–6]; for example, Ellenberg and Hamilton [4], used the amount of analgesic consumed by cancer patients as an example of a surrogate endpoint for pain. However, the reduction in analgesic consumption itself could be considered as a clinical endpoint. They also gave the example of the use of skinfold thickness as a substitute for nitrogen, potassium and water level measurements in the assessment of muscle mass, none of those being clinical endpoints.

The following examples can be used to illustrate the proposed definition of a surrogate endpoint.

1. The assessment of blood lipid-lowering drugs is commonly based on their ability to decrease blood lipid concentrations, whereas these drugs are prescribed for patients with hypercholesterolaemia for the prevention of atherosclerosis, which, in turn, is thought to be responsible for disabling or fatal events, such as myocardial infarction and sudden cardiac death.
2. Phase III trials of new drugs for the treatment of hypertension are designed to record effects on blood pressure, viz. the extent and duration of the decrease in the systolic and diastolic blood pressure. In practice, these drugs are prescribed to prevent cerebral haemorrhage, ischaemic stroke, congestive heart failure and various coronary events.
3. Treatments for the prevention of post-operative venous thrombosis in the lower limbs are assessed using the reduction in the incidence of positive ^{125}I-labelled fibrinogen uptake test (FUT) results, with or without a positive phlebogram. These treatments are, in practice, used for the prevention of fatal pulmonary embolism.

The intermediate endpoints used in these examples, blood cholesterol level, blood pressures and FUT results, are more readily measured than the corresponding clinical endpoints, thus making the clinical trials easier to perform. They are statistically correlated with the clinical endpoints, although the clinical event is rarer than the intermediate event. In all three examples, the supporting evidence for the role of the intermediate endpoint in the pathophysiological model of the underlying condition or disease differs substantially, but in all three instances a

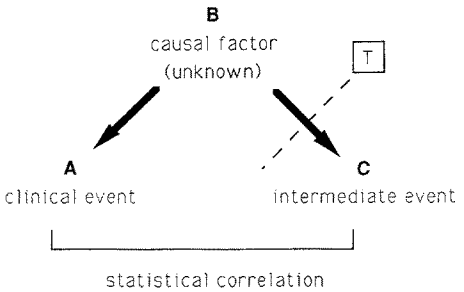


Fig. 1. An example of a pathophysiological and effect model, which shows that event C is an inappropriate surrogate endpoint. Treatment T has an effect on the occurrence of the intermediate event C, without affecting the occurrence of the clinical event (see text)

cause – effect relationship is assumed (not proven) between the intermediate and clinical endpoints.

Certain differences can be noted in the examples. Deep venous thrombosis (DVT), which occurs prior to pulmonary embolism, is detected by a positive FUT result and can be considered as a clinical endpoint in its own right, and, while pulmonary embolism is assumed to be caused by DVT, it may also be caused by other, rare conditions, such as right atrial myxoma and sepsis. High blood cholesterol levels, hypertension and smoking have been identified as risk factors for coronary heart disease (CHD), and, although in subjects presenting these risk factors, CHD is more likely to occur, symptomatic CHD may also develop in subjects with no known risk factors. Moreover, whilst it is true that the occurrence of high blood pressure and cholesterol levels has been statistically correlated with CHD, no formal proof exists to show that they cause CHD.

Amongst the many advantages offered by surrogate endpoints [1, 4, 5] is the fact that they are generally easier to standards than their corresponding clinical endpoints; for example, blood cholesterol levels can be standardized by performing all determinations in a central laboratory, whereas determination of the cause of death is more difficult to standardise. Also, surrogate endpoints, such as blood pressure and serum cholesterol level, are quantitative variables, whereas the corresponding clinical endpoints are binary variables. In a clinical trial using the former, fewer subjects will be required to give equivalent power to the statistical tests and the trials can be performed in a shorter period of time (e.g. over a few weeks), rather than the several years it would take for the clinical outcomes to appear.

However, the fact that not all intermediate endpoints are appropriate surrogate endpoints can be illustrated by the history of antiarrhythmic drugs [7–9]. Until recently, it was accepted that drugs which prevented the occurrence or recurrence of ventricular arrhythmias (e.g. ectopic beats, tachycardia) would also prevent sudden cardiac death in patients suffering from these arrhythmias. Ventricular arrhythmia was, therefore, taken to be a surrogate endpoint for sudden death. However, in a recent trial [9], it was shown that patients treated with this type of drug had a higher risk of sudden death than patients treated with placebo, indicating that, in this case, reducing ventricular arrhythmias did not reduce the risk of sudden death. The explanation for the observation is thought to involve the toxicity of Class Ic antiarrhythmic drugs, which may outweigh their possible therapeutic benefit. It is apparent that the choice of a suitable surrogate endpoint is of prime importance, so several points should be carefully considered when the choice is being made.

One major point is the confusion between statistical correlation and proof of a cause-effect relationship. This is shown schematically in Fig. 1, where a clinical event, A, is caused by an unknown factor, B, which also causes a second event, C. Although the mechanisms of the two events are completely unrelated, there is a statistical relationship between the events, so that C may be considered as a risk factor for A. The treatment T, may modify the event C, without having a direct effect on B, and it will, consequently, have no effect on A. Thus, although they are correlated, C is not a suitable surrogate endpoint for the clinical event A. This demonstrates how important an understanding of the pathophysiology of the disease of interest is when a surrogate endpoint is to be selected.

A second point is related to the long-term safety of the study treatment. Although a suitable surrogate endpoint may be identified, short-term studies on restricted numbers of patients, will not enable any long-term or rare toxicity to be demonstrated. This problem, unlike the first, is

Table 1. Abbreviations used and some basic assumptions made

Abbreviation:	Significance:
E	occurrence of a clinical event (CE), e.g. myocardial infarction
I	occurrence of an intermediate endpoint (IEP) e.g. positive FUT result, or a decrease in blood pressure
nI	absence of an intermediate end point
P(E/I)	the probability of E, given I (I has occurred)
P(E/nI)	the probability of E given not I (I has not occurred)
[I-E]	relationship or interaction between an IEP and a CE
L	likelihood ratio for E, given by P(E/I)/P(E/nI)
∧	the Boolean operator "and"

- The effects of a treatment, T, are denoted by t as a subscript, e.g. P(E/I) becomes $P_t(E/I)$.
- The rate differences for I and E under treatment are denoted as $\delta_t(I) = P(I)-P_t(I)$ and $\delta_t(E) = P(E)-P_t(E)$, respectively. If $P_t(I) = k\,P(I)$ with $k \geq 0$, and if the treatment works, i.e. $0 \leq k < 1$, then $P_t(E/I) = rP(E/I)$ and $P_t(E/nI) = r'P(E/nI)$.
- The likelihood ratio for E is given by $L = P(E/I)/P(E/nI)$, when the occurrence of I has been assessed in a given population; thus $L_t = r.r'^{-1}L$.
- The probabilities P(E/I), P(E/nI), $P_t(E/I)$, and $P_t(E/nI)$, are assumed to be constant for a population which is not constant, i.e. P(I) varies. This assumption is often made in the medical field and is the basis for the Bayesian analysis of medical diagnoses. Estimates of these probabilities can be obtained from relevant epidemiological studies. The ideal situation, when P(E/I) = 1 and P(E/nI) = 0, is almost a theoretical notion, and the surrogate endpoint should be chosen to approach this situation as closely as possible. For example, if AIDS is considered as a global diagnosis, then HIV seropositivity can be considered as a surrogate end point for this clinical endpoint when considering prevention of contamination.

Table 2. Verification of the third proviso: four possible ways of estimating $\delta_t(E)$ and $\delta_t(I)$

	T does not affect [I-E]	T affects [I-E]
P(E/nI) = 0	$\delta_t(E) = P(E/I).\delta_t(I)$	$\delta_t(E) = P(E/I).[(1\text{-}kr^{-1}).P(I)]$
	3rd proviso satisfied if P(E/I) is known	r and P(E/I) should be known, r should be constant
P(E/nI) > 0	$\delta_t(E) = P(E/nI).[L\text{-}1]\delta_t(I)$	$\delta_t(E) = P(E/nI).\{P(I).[L.(1\text{-}rk)\text{-}1 + r'k] + 1\text{-}r'\}$
	3rd proviso satisfied if P(E/I) and P(E/nI) are known	3rd proviso satisfied if T is isotropic for [I-E] and r is constant

See Table 1 for an explanation of the abbreviations used

treatment-specific, and as such, has no direct bearing on the choice of a suitable surrogate endpoint, but, nonetheless it should be considered.

From this brief presentation, the need for a rigorous approach to the choice of a surrogate endpoint, especially in terms of its relationship to the clinical endpoint, is evident. A formal framework is now proposed, which may be used to define the requirements for a surrogate endpoint and to validate the choice. By combining two different approaches, one mathematical and the other heuristic, how the requirements can be satisfied is examined, using two examples to illustrate the process. Since the mathematical derivations are not essential to the discussion, the majority are presented as appendices.

This approach has been chosen for two reasons; first, the causes of many human diseases are either remote or unknown, or they may be multiple, and in such a complex situation, the best solution consists of developing a model which fits the available data, the latter often being quantitative in nature. And second, the assessment of the efficacy and safety of a new therapy, which is based on randomised clinical trials, is often expressed in quantitative terms. Under certain conditions, randomised clinical trials may provide two, equally important pieces of information, namely a standardised test of the hypothesis of a clinical benefit, and an estimate of the magnitude of the true benefit. The latter is fundamental for public health policy. This quantitative aspect of new therapy assessment forms the cornerstone of the approach.

Surrogate endpoints may either be binary or quantitative variables, or, to be more precise, a quantitative change in a parameter, such as blood pressure. In the following discussion the simplest case, the binary variables, first is considered since, a quantitative variable may be considered as a special form of a binary variable.

The abbreviations and certain basic assumptions used are summarised in Table 1.

Provisos

To be considered as a suitable substitute for a clinical endpoint, a surrogate endpoint should satisfy the following three provisos:

First proviso: convenience. Although the surrogate endpoint should be easier to assess than the corresponding clinical endpoint, the most important characteristic is its frequency, i.e. it should occur more often than the corresponding clinical endpoint, so that –

$$P(I) > P(E)$$

Second proviso: relationship. The relationship between the surrogate endpoint and the clinical endpoint should be well established, both qualitatively and quantitatively, through relevant epidemiological studies. Mathematically, there should be a formal relationship between the two probabilities, i.e.

$$P(I) = f[P(E)]$$
$$\text{and } P(E) = f^{-1}[P(I)]$$

Third proviso: estimate of clinical benefit. An estimate of the expected clinical benefit, should be derivable from the estimate of the reduction in the incidence of the surrogate endpoint, which can be obtained from randomised clinical trial data. In mathematical terms, if P(I) has been estimated for a given disease, then P(E) can be derived. Also –

$$\delta_t(E) = f'\{\delta_t(I)\}$$

where δ_t represents the effect of the treatment.

Comments

1. If these provisos and their properties described are satisfied, the intermediate endpoint may be considered as an appropriate surrogate endpoint. The second and third provisos are related to [I-E], and can be considered from either a mathematical or a pathophysiological approach, the choice being dependent only on practical considerations. This may not apply in the rare instances where the intermediate endpoint is predictive of a subclass of clinical endpoints (e.g. polyposis and colon cancer). In the following discussion it will be assumed that the first proviso is satisfied without reducing the generalisability.
2. The third proviso is essential, since δ_t is necessary for the assessment of the risk/benefit ratio, and so it should be derived from the observed effect on the surrogate endpoint.

Four possible situations

Once an estimate of $\delta_t(I)$ has been obtained from a randomised clinical trial, the estimation of $\delta_t(E)$ from this is a central step in the assessment of the risk/benefit ratio of a given treatment. The probabilities, P(E), P(I), $P_t(E)$ and $P_t(E)$ are linked by two basic equations (see Appendix 1),

which are correct, irrespective of [I-E], and of whether the intermediate endpoint is a true surrogate endpoint or not.

Since it is assumed that P(E/I), P(E/nI), P_t(E/I), and P_t(E/nI) are constant, Equations 1′ and 3′ in Appendix 1 are linear functions of P(I) and P_t(I), respectively. Although it is true in most cases that P_t(E/I) = P(E/I) and P_t(E/nI) = P(E/nI), it is important not to exclude cases where the treatment may modify [I-E][1], in which case P_t(E/I) ≠ P(E/I) and P_t(E/nI) ≠ P(E/nI).

The resulting Equation 5′ confirms that the effect on the clinical endpoint depends, on the effects of the treatment on the intermediate endpoint, and thus on [I-E], as shown by the presence of L_t and P_t(E/nI) in the equation. There are four different possibilities for the estimations δ_t(E) and δ_t(I) (i.e. the third proviso), depending on whether:

- the clinical endpoint can occur when the intermediate endpoint has not occurred
- the treatment has an effect on the relationship [I-E].

These particular cases, which considered mathematically in Appendix 2, and are summarised in Table 2, are briefly described here.

Case I, P(E/nI) = 0, and the treatment does not affect [I-E]. This is shown by Equation 8′ in Appendix 2. When P(E/I) is known from previous epidemiological studies, i.e. if the second proviso is satisfied, then the third will also be satisfied. In order to prove that an intermediate endpoint is a true surrogate endpoint, it is necessary to show that P(E/nI) = 0, and that [I-E] is unaffected by the treatment. This is possible if either the mechanism of [I-E] or the mechanism of action of the treatment are known.

Case II, P(E/nI) = 0, and the treatment modifies [I-E]. The effect on the clinical endpoint is given by Equation 9′ in Appendix 2. In this case it is generally not possible to verify the third proviso, unless each treatment has the same effect on [I-E], in which case r is constant, and if an estimate of P_t(E/I) has previously been obtained in a randomised clinical trial with the appropriate clinical event as the endpoint. The first of these conditions can only be verified if several randomised clinical trials have been performed with the given clinical event as the endpoint.

Case III, P(E/nI) > 0, and the treatment does not modify [I-E], i.e. the clinical event may occur in patients who do not present the intermediate endpoint, e.g. myocardial infarction may occur in patients with a normal blood pressure and cholesterol level. The effect on the clinical event is given by Equation 10′ in Appendix 2. The third proviso can be satisfied if estimates of P(E/I) and P(E/nI) are available from previous epidemiological studies. Previous randomized clinical trials with the clinical event as the endpoint are not necessary for the assessment of the intermediate endpoint as a surrogate endpoint, although this is often the only method of demonstrating that the treatment has no effect on [I-E]. The example of blood pressure – stroke, given below illustrates this point.

Case IV, P(E/nI) > 0, and the treatment modifies [I-E]. In this case δ_t(E) is no longer a simple function of δ_t(I), and the equation becomes 11′ in Appendix 2. In this case the third proviso can be verified if k, r, r′, P(E/I), and P(E/nI) are known, which is only possible in the unlikely instance when r and r′ are independent of the type of treatment being tested.

A special case may arise when the effects of a treatment on the intermediate endpoint and the clinical event are based on the same pharmacodynamic mechanisms, although there is no causal relationship between the intermediate endpoint and the clinical event, but they do share a common remote cause, as can be seen in Model no. 2 which is presented below (see also Fig. 2). In this case the effect of the treatment on [I-E] is said to be *isotropic,* since r = r′. If r and k are greater than 0, δ_t(E) is always positive. When r is independent of treatment, r = r′ and is constant. In this situation the second and third provisos can be verified if there is sufficient data from epidemiological studies and randomized clinical trials using the clinical event as the endpoint.

A quantitative model

The application of the above probabilistic model may appear to be difficult when the intermediate endpoint in question is a quantitative variable, e.g. high blood pressure, or high blood cholesterol level. In the clinical situation, only patients with hypertension or hypercholesterolemia are treated, and instead of representing a difference in frequency, δ_t(I) represents a decrease in blood pressure or cholesterol level. If a decrease to below a certain level is taken as 'success', S, then (I) can be replaced by (S) in the model. Alternatively, it is possible to establish a quantitative model directly, and thus both models can be used for quantitative endpoints.

Several prospective epidemiological surveys have shown that the risk of CHD is dependent both on blood pressure and serum cholesterol level [10, 11]. Mathematically [I-E] can be represented by:

$$R(E) = f(x) \quad (1)$$

where x is the blood pressure or cholesterol level, and R is the risk. Thus, the first and second provisos are satisfied and the third proviso becomes:

$$\delta[R(E)] = f\{\delta f(x)\} \quad (2)$$

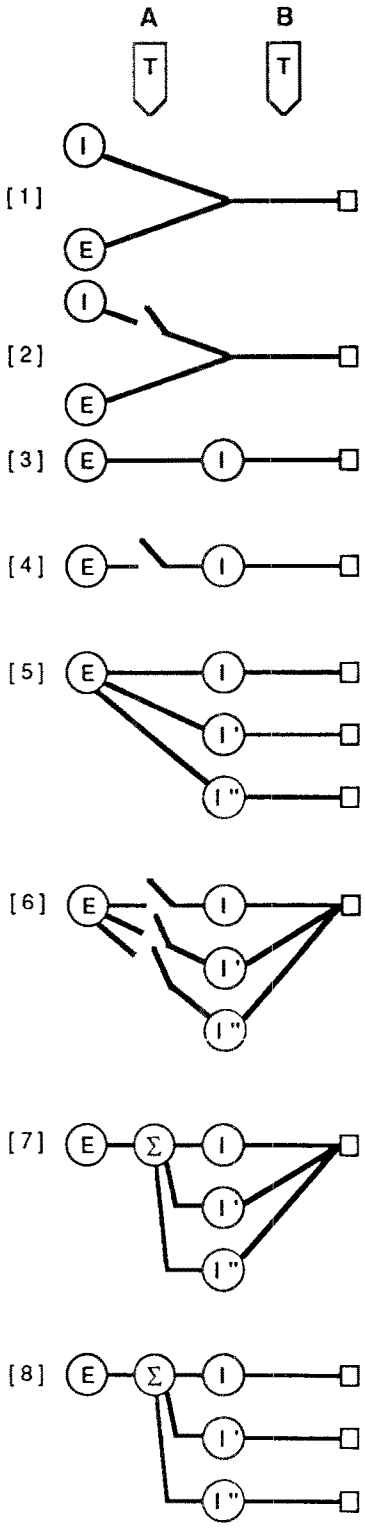
and

$$\delta_t [R(E)] = f_t\{\delta_t f(x)\} \quad (3)$$

with both δf(x) and δ[R(E)] being positive and δ_tf(x) and δ_t[R(E)] being negative. The right-hand-side of Equation 10 is f′ (x). δx, where f′ (x) is the derivative of f(x). This gives:

$$\delta[R(E)] = f'(x).\, \delta x \quad (4)$$

[1] For example: to explain the inconsistent relationship between the rate of reperfusion after administration of a thrombolytic agent during the acute phase of myocardial infarction and the reduction in mortality rate, it has been claimed that there is a facilitating effect on distal coronary blood flow as a result of the reduction in the blood fibrinogen level



		2nd and 3rd provisos
I and E share a common remote cause. However, I is not a causal link in the mechanism resulting in E, although E cannot occur without I.	$P(E/I) = 1$ $P(E/nI) = 0$	are always met if T acts before the branch (B) or on both pathways (A)
I and E share a common remote cause, whereby they are statistically correlated. However, I is not a causal link in the mechanism resulting in E, and E can occur without I.	$P(E/I) \neq 1$ $P(E/nI) \neq 0$	are always met if T acts before branch (B).
I is a mandatory link in the mechanism leading to the occurrence of E.	$P(E/I) = 1$ $P(E/nI) = 0$	are always met.
I is a mandatory link in the mechanism leading to the occurrence of E. However, if the circuit-break is open, E does not occur, even if I has occured.	$P(E/I) \neq 1$ $P(E/nI) = 0$	are always met.
E has more than one, independent remote cause, each having a different pathway, only one of which involves I.	$P(E/nI) \neq 0$	are met if the other Is are not concurrent causal factors (i.e. do not take over if I disappears), or if T acts on all pathways.
E can occur through different pathways all originating from the same remote cause. Any pathway can lead to E, but only one involves I.	$P(E/nI) \neq 0$	are met if the other Is are not concurrent causal factors (i.e. do not take over if I has disappeared) or if T acts on all pathways or on the common cause.
The occurrence of E requires the addition of inputs from different pathways, one of which involves I, arising from a common remote cause. No single pathway is sufficient to cause E, but a combination not involving all of them, can be sufficient.	$P(E/nI) \neq 0$	are met, depending on the weight of I in Σ; are always met if T acts on all pathways or on the common cause.
The occurrence of E requires the addition of inputs from different pathways, one of which involves I, arising from different remote causes. No single pathway is sufficient to cause E, but a combination, not involving all of them, can be sufficient.	$P(E/nI) \neq 0$	are met, depending on the weight of I in Σ; are always met if T acts on all pathways (B) or after Σ.

Fig. 2. Some possible schematic pathophysiological and treatment effect models. Treatment T can act either at point A or point B.

When $\delta_t x = \delta x$, generally, $\delta_t[R(E)] \neq \delta[R(E)]$. However, if the treatment does not modify [I-E], Equations 2 and 3 are identical, and

$$\delta_t[R(E)] = f'(x).\ \delta x \quad (5)$$

The third proviso may be verified by fitting Equation 4 to the results of many randomised clinical trials using different treatments[2]. If evidence from other sources supports the hypothesis that the treatment does not modify [I-E], then Equation 5 can be used to estimate $\delta_t(E)$.

[2] Better verification that Equation 5 is true for a large number of subgroups can be obtained by using the values of δx obtained from each individual trial

Table 3. Expected and observed risk reduction for an absolute reduction of diastolic blood pressure of 6–5 mm Hg (data taken from references 12, 13)

Fatal and non-fatal events	Expected $\delta_t(E)$[a] %	Observed $\delta_t(E)$[b] % (95% confidence interval)
Stroke	35–40	42 (33–50)
CHD	20–25	14 (4–22)

[a] from 7 (stroke) or 9 (CHD) epidemiological studies, mostly fatal events, over 10 years
[b] from 14 RCTs, fatal and non-fatal events over 5 years

Mechanisms of [I-E]

This mathematical approach, based on probability theory and using data from both epidemiological studies and randomized clinical trials, is valuable in the development of a general solution. However, for investigating the relationship [I-E], it is useful only in a few cases, e.g. the blood pressure – stroke example given later. Thus, the mechanism of [I-E] should be more thoroughly examined by means of a basic pathophysiological investigation of the underlying disease.

When the clinical event is the final outcome in a linear sequence of events, and the intermediate endpoint has been proven to be a unique causal link in the sequence, it is obvious that:

$$P(E/I) = 1$$

and

$$P(E/nI) = 0$$

and that the intermediate endpoint is suitable as a surrogate endpoint. Consideration of the mechanism of [I-E], offers a further, complementary approach to this problem. Some of the possible mechanisms for [I-E] which involve both the intermediate endpoint and the clinical event are given schematically in Fig. 2. From these, it can be seen that I may be related to E through a parallel link (Models 1 and 2), or through a sequential link (Models 3 to 8), and that between the remote cause and the occurrence of E, there may be one (Models 1 to 3) or several pathways (Models 4 to 8).

It is also interesting to note that, for example, in Models 3 and 4, although the intermediate endpoint may be an appropriate surrogate endpoint, the treatment can be efficacious for E without affecting I, depending on whether it acts in the system before or after the intermediate endpoint (see the vertical arrows representing the treatment in Fig. 2). This is another example of an interaction between the treatment and [I-E]. As shown in Models 7 and 8, the selected intermediate endpoint may not be an appropriate surrogate endpoint, even if it has been shown to be a causal factor for the clinical event, because the threshold above which E will occur is always, or almost always, surpassed independent of the occurrence of I. It is also possible that the intermediate endpoint may be an appropriate surrogate endpoint, even when it is not a causal factor, as shown in Models 1 and 2. Once again, it can be seen that the manner in which the treatment interacts with [I-E], and the [I-E] relationship itself, are key issues in this problem.

Examples

Decrease in blood pressure as a surrogate endpoint for the risk of stroke or coronary heart disease

In the assessment of new antihypertensive treatments, blood pressure is a standard surrogate endpoint for clinical events for which the patients involved present higher risks. It is assumed that a reduction in blood pressure will be associated with a reduced risk of either stroke or CHD following the results of numerous prospective cohort studies showing a positive correlation between them [12]. A linear logarithmic relationship is thought to exist between the baseline diastolic blood pressure and the risk of both fatal and non-fatal coronary heart disease and stroke [12, 13]. Given that:

$R(E) = f(x; \tau)$: the risk in untreated patients (or controls in clinical trials) with a given diastolic blood pressure (x) at baseline, over a period τ.

$R_t(E) = f_t(x; \delta x; \tau)$: the risk in treated patients with a diastolic blood pressure (x) at baseline, and a treatment effect, δx, over a period τ.

It was stated above that log R (or ln R) is linearly related to the diastolic blood pressure, therefore:

$$\ln R = k.x + b \quad (6)$$

which is equivalent to:

$$R = e^{[k.x + b]} \quad (7)$$

and from this:

$$\frac{\delta R}{R} = k.\delta x \quad (8)$$

If it is assumed that any absolute reduction in risk or blood pressure will be small enough to be replaced by δR and δx, respectively, this model implies that the relative reduction (or increase) in risk is proportional to the absolute reduction (or increase) in the diastolic blood pressure, irrespective of the magnitude of the change. Using the data obtained from a cohort survey and several randomised primary prevention clinical trials of various hypertensive treatments (diuretics, propranolol and reserpine, given separately or concomitantly), Collins et al. [14] calculated the overall reduction in the risk of both fatal and non-fatal stroke and CHD. The corresponding observed, relative reductions $\delta_t[R(E)]$, were compared with the expected reductions, calculated using the data from the cohort studies, $\delta[R(E)]$, for $\delta_t x = \delta x$ (see Table 3).

The experimental relative risk reduction is fairly close to the expected range for stroke. Thus, Equation 5 holds and it can be concluded that the treatments tested did not modify [I-E]. Since $P(E/nI) > 0$, the diastolic blood pressure and stroke fall into the case where only P(E/nI) and P(E/I) are known (Case III above; Table 3). Blood pressure can be used as a surrogate endpoint for stroke, since it has been shown to be isotropic in several randomised clinical trials with different treatments.

However, it is not possible to predict the effect of treatment on the risk of CHD from the risk function, in other words Equation 5 is not correct for CHD. Thus, $P_t(E/I) \neq P(E/I)$, and the treatments tested are likely to modify [I-E]. Blood pressure and coronary heart disease are examples of Case IV situations. Thus, Equation 11′ of Appendix 2 becomes

$$\delta_t(E) = P(E/nI).P(I).[L.(1 - rk) - 1 + k] \quad (9)$$

since it can be assumed that r' = 1 if the hypertensive treatments have no effect on the risk of CHD in normotensive patients. The above equation can be estimated directly by using the observed $\delta_t[R(E)]$, i.e.

$$\delta_t[R(E)] = \delta[R(E)] = F_t\{\delta_t f(x)\} \quad (10)$$

which is also implicit in Table 3. It is tempting, therefore, to accept blood pressure as a surrogate endpoint also for CHD, but care should be taken, since the statistical significance of a 14 % reduction in CHD with a reduction in diastolic blood pressure of 5–6 mm Hg is borderline, and as the confidence limit may not include 0 simply by chance alone (Table 3). Also, it is not possible to verify that $r = r$, i.e. that the treatment effect is independent of the treatment, due to the limited size of the observed effect, which varied from trial to trial. In addition, P(E/I) and P(E/nI) may vary, depending on the population studied.

The discrepancies between the results for stroke and CHD may be due to several reasons, including the lack of statistical power in the randomised clinical trials on CHD[3] differences in the requirements for treatment duration and/or follow-up, cardiotoxicity of the tested drugs (hazard model), inaccuracy of available data, lack of efficacy of the tested treatments for CHD and, finally, different disease models for CHD and stroke. For example, cerebral haemorrhage and ventricular hypertrophy are thought to be related to a unique alteration of the arteriolar wall, leading to increased vascular resistance in the peripheral, splanchnic and renal arterial beds, which is not arteriosclerosis. Left ventricular hypertrophy, whether or not caused by hypertension, may have a role in the relationship between coronary arteriosclerosis and the occurrence of CHD, by increasing the myocardial oxygen demand. Thus, the direct effect of an absolute reduction in blood pressure on left ventricular hypertrophy may have an indirect effect on the occurrence of CHD, and thus can explain the lack of fit. Alternatively, it could be simply that arteriosclerosis is not as closely involved in the relationship with cerebrovascular ischaemic disease as it is in the relationship with CHD, and thus the reduction in blood pressure may have a limited effect on the arteriosclerosis. This possibility is supported by the observation that spontaneously hypertensive rats die of stroke, cardiac failure and renal lesions, but not CHD. The existence of different relationships between hypertension and CHD or stroke, is further supported by the selection of stroke-prone and stroke-resistant, spontaneously hypertensive rats by Okamoto et al. [15]. Although the limited life-span in the hypertension animal models suggests that Model no. 2 (Fig. 2), is the most likely model for stroke, congestive heart failure and renal failure, the available data from human epidemiological and clinical studies support this only as a model for stroke. For CHD, Models no. 5 or 8 have been suggested as being more appropriate.

This example also illustrates other useful points, viz, the analysis necessitates the careful assessment of bias arising from potential confounding factors. R(E) may change with the duration of follow-up (and/or treatment) and the values of other risk factors, e.g. serum cholesterol, smoking habits, diabetes, age, and family history. It is possible that the patients included in clinical trials do not have the same profile as those included in cohort studies which gave the value of R(E) = f (x); this is an important point which warrants further discussion, but it will not be undertaken here. It is also often difficult to explore the effect of the treatment on [I-E] using an heuristic approach, because disease mechanisms are often only poorly understood.

Antiarrhythmic effects and mortality

In the Introduction, it was pointed out that ventricular arrhythmias, i.e. ventricular ectopic beats and tachycardia, can no longer be considered as surrogate endpoints for sudden death or total mortality, following the results from CAST [9]. The release of antiarrhythmic drugs many years ago was based on their effect on premature ventricular depolarisation, and so the data published prior to that of CAST are still useful to illustrate the theory. The reasons for their choice as surrogate endpoints have been discussed in several prospective surveys [8, 16], and they are that sudden cardiac death is very often caused by ventricular fibrillation which can be provoked by ventricular ectopic beats and tachycardia, and the fact that these arrhythmias in the ECG of ambulatory patients with clinical evidence of CHD have been correlated with an increased risk of subsequent sudden death. Patients who have suffered from myocardial infarction and who have ventricular arrhythmia have a greater risk of death, especially sudden death. A series of placebo-controlled, randomised clinical trials which use these arrythymias as endpoints have shown the efficacy of several antiarrhythmic agents, but, with the exception of the CAST, those trials were not large enough to test the treatment effect on the clinical event, i.e. sudden cardiac death. Seven of the 12 published randomised clinical trials of antiarrhythmic agents in post-myocardial infarction patients have provided data showing evidence of efficacy on both arrhythmias and mortality [16, 17]. In all the trials the follow-up for the observation of arrhythmia was short compared with that in CAST, and although there was a marked reduction in the incidence of ventricular arrhythmias in treated patients compared to control patients, irrespective of the definition of arrhythmia used, the incidence of death was not different.

From the published data, it is not possible to verify the effect of treatment on [I-E], because $P_t(E/nI)$ is not given. However, observations with long-term ECG monitoring

[3] Lack of statistical power is an unlikely explanation, since the rate of CHD is higher than that of stroke for a given level of diastolic blood pressure

in post-myocardial infarction patients have shown that sudden cardiac deaths are likely to be due either to primary ventricular fibrillation, or ventricular fibrillation secondary to a ventricular tachycardia. It can be assumed, therefore, that antiarrhythmic agents have some effect on [I-E], and because $P(E/nI) > 0$, this is an example of Case IV. Pooled data from the seven trials mentioned above showed no correlation between the reduced risk of ventricular arrhythmia and a reduction in total mortality [16], and thus the third proviso is not satisfied. Moreover, there is some indication that these agents may have a noxious effect, because of their proarrhythmogenicity, which can annul any potential benefit on the risk of sudden cardiac death [18].

Discussion

Surrogate endpoints have been defined in various ways by different authors; for example, Ellenberg and Hamilton [4] wrote "investigators use surrogate endpoints when the endpoint of interest is too difficult and/or expensive to measure routinely and when they can define some other, more readily measurable endpoint, which is sufficiently well correlated with the first to justify its use as a substitute"; Wittes et al. [5] defined them as "an endpoint measured in lieu of some other, so-called, 'true' endpoint"; Hillis and Siegel [6], using the term 'observation' instead of endpoint, defined a surrogate observation as, "an observed variable that relates in some way to the variable of primary interest, which we cannot conveniently observe directly"; and finally, Prentice [19] reviewed the definitions proposed by others and concluded that a surrogate endpoint is, "a response variable for which a test of the null hypothesis of no relationship to the groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint". He went on to propose an operational criterion for the identification of surrogate endpoints, which was based on a derivation of time-to-failure variables, but which can probably be extended to other sets of covariates, e.g. blood cholesterol and coronary events. His criterion assumes that [I-E] cannot be modified by the tested treatment and he suggests an approach to test this assumption, but he does not take into consideration an estimate of the effect on the clinical endpoint.

The definition and approach presented here are different from these essentially because there is emphasis on the concept of a clinical endpoint and also on the estimate of the treatment effect on the clinical endpoint. It is apparent, from what has been stated above, that even a comprehensive and accurate knowledge of the mechanism of a disease may not be sufficient to enable a surrogate endpoint to be chosen. In reality, even a superficial knowledge of the mechanism is often not available, and for the majority of situations it is not possible to sketch simple diagrams, such as those in Fig. 2. Thus, although the mathematical approach does not provide firm guidelines, as described above it can provide a framework, within which we can work.

This problem cannot be resolved by a single approach and, therefore, when using surrogate endpoints, it is important to bear the following three points in mind:

1. An accurate, well-documented model of the condition is a prerequisite, as has been emphasized by many authors [5, 19], but it should be remembered that a model is only a model and as such may be far from reality, and that new data may reveal its inadequacy.
2. Precise, relevant data from prior epidemiological studies is always priceless.
3. In some situations, it is obligatory to have results from a sufficient number of carefully conducted, randomised clinical trials using a clinical event as the outcome, involving a large number of different treatments (covering all possible interactions between potentially effective treatments and [I-E], i.e. all possible therapeutic models). These should have been conducted so that an intermediate endpoint can be validated as a true surrogate endpoint.

Due to the limited experience available in this area, it is not possible to give precise guidelines to validate the choice of surrogate endpoints. However, it is possible to establish a sketchy algorithm from the derivations presented above. This is based on the type of available data, and assumes that they are both reliable and adequate:

1. *If only epidemiological data are available.* The choice will not be possible since the third proviso cannot be verified.
2. *If only data concerning the pathophysiology of the disease and the mechanism of action of the treatment (which will be referred to as PP-MAT data) are available.* The choice will not be possible because the second and third provisos cannot be verified, ever in the optimal case when:
   - this corresponds to either Model 3 or 4, i.e. the intermediate endpoint is the only cause of the clinical event or is one step in the sequence,
   - this corresponds to Model 1 or 2, i.e. it is not causal but it is strongly correlated, with both having a common remote cause, and the treatment acts at B
   - or, there is strong evidence to suggest that the treatment does not modify the interaction [I-E].
3. *If epidemiological and PP-MAT data are available.* The second and third provisos may be checked if the PP-MAT data enables the three conditions in 2 (above) to be satisfied, i.e. those concerning the model and the effect of treatment on the interaction [I-E].
4. *If data from epidemiological studies and randomised clinical trials with clinical events as endpoints are available.* It is possible to check the second and third provisos by using Equations 4, 5 or 7′–10′ and the corresponding conditions, which will be dependent on the situation being investigated.
5. *If only data from randomised clinical trials with clinical events as endpoints are available.* As in 4 above.

It should be noted that verification of the first proviso is not always feasible, due to the cost. For example, FUT and phlebography results are expensive as intermediate endpoints, are subject to technical difficulties and can also be hazardous for the patients. On the other hand, assessment of death from pulmonary embolism is inexpensive and simple, if a suitable definition is available. An indirect

consequence of the use of intermediate endpoints, such as FUT or phlebography results, is the selection of subgroups of the target population, i.e. those patients who give their consent and who are suitable for the procedure.

These guidelines do not consider the potential hazards associated with the treatment when the clinical event can be adversely affected, i.e. the so-called "negative" component, or when potential adverse effects on the surrogate endpoint take longer to appear than the "positive" effects.

Conclusion

In general, those who use intermediate endpoints as surrogate endpoints for clinical events do not know if they satisfy the three basic provisos, in particular the second and third, since there are usually insufficient data available. This can result in the release and widespread use of new treatments before their risk/benefit ratio has been correctly determined. It is suggested that the assessment process for the efficacy of new treatments should also provide proof of the benefit and an estimate of it. It has been shown that even strong statistical correlation is not adequate as a criterion for the definition of a surrogate endpoint. In conclusion, since it is almost always extremely difficult, or even impossible, to prove that a given intermediate endpoint is, without doubt, an appropriate surrogate endpoint, whenever possible the assessment of new and old treatments should be based on clinical rather than intermediate endpoints.

Appendix 1

Basic equations

The derivation of the basic equations is straightforward. Given P(E) and P(I), and $P_t(E)$ and $P_t(I)$, which are linked through [I-E], irrespective of [I-E], and of whether the intermediate endpoint is a true surrogate endpoint for the clinical event or not:

$$P(E) = P(E \wedge I) + P(E \wedge nI)$$
$$P(E) = P(E/I).P(I) + P(E/nI).P(nI)$$
$$P(E) = P(E/I).P(I) + P(E/nI).[1 - P(I)]$$
$$P(E) = P(E/I).P(I) + P(E/nI) - P(E/nI).P(I)$$

which gives:

$$P(E) = [P(E/I) - (E/nI)].P(I) + P(E/nI) \quad (1')$$

If $P(E/nI) \neq 0$, since $L = P(E/I)/P(E/nI)$, equation (1') becomes:

$$P(E) = P(E/nI).\{[L - 1].P(I) + 1\} \quad (2')$$

This same derivation applies under treatment, therefore:

$$P_t(E) = [P_t(E/I) - P_t(E/nI)].P_t(I) + P_t(E/nI) \quad (3')$$

and if $P_t(E/nI) \neq 0$, equation (3') becomes:

$$P_t(E) = P_t(E/nI).\{[L_t - 1].P_t(I) + 1\} \quad (4')$$

Although it may be assumed in most cases that $P_t(E/I) = P(E/I)$ and $P_t(E/nI) = P(E/nI)$, it is important not to exclude the cases where the treatment may modify [I-E], in which case $P_t(E/I) \neq P(E/I)$ and $P_t(E/nI) \neq P(E/nI)$. The effect of treatment on the clinical endpoint, if any, is given by:

$$\delta_t(E) = P(E) - P_t(E)$$
$$= P(E/nI).\{[L - 1].P(I) + 1\} - P_t(E/nI).\{[L_t - 1].P_t(I) + 1\} \quad (5')$$

Appendix 2

Mathematical properties of [I-E]

Depending on the nature of [I-E], there are cases where an intermediate endpoint is a true surrogate endpoint:

1. P(E/nI) = 0

E cannot occur when I has not occurred. It is probable that $P_t(E/nI) = 0$, therefore:

Equation 1' becomes: $P(E) = P(E/I).P(I)$
and Equation 3' becomes: $P_t(E) = P_t(E/I).P_t(I)$

and therefore:

$$\delta_t(E) = [P(E/I) - kP_t(E/I)].P(I) \quad (6')$$

or:

$$\delta_t(E) = P(E/I).(1 - kr^{-1}).P(I) \quad (7')$$

1.1. If the treatment T does not affect [I-E]

This gives $P_t(E/I) = P(E/I)$, and if the treatment has a favourable effect on P(I), then $0 \leq k < 1$, and from Equation 6', it can be deduced that $\delta_t(E) > 0$. Also from Equation 6':

$$\delta_t(E) = P(E/I).\delta_t(I) \quad (7')$$

Therefore, since P(E/I) is known from previous epidemiological studies, the third proviso is satisfied[1].

1.2. If the treatment modifies [I-E]

This case, where $P_t(E/I) \neq P(E/I)$, cannot be excluded. In this case $\delta_t(E)$ can be derived from P(I) and $P_t(I)$, only if $P_t(E/I)$ is known. If an intermediate endpoint is used instead of a clinical event, $P_t(E/I)$ cannot be estimated. It is not possible, therefore, to verify the third proviso, except if each treatment has an identical effect on [I-E] and r is constant and can be taken to be *r*, and there is at least one randomised clinical trial with a clinical event as the endpoint which provides an estimate of $P_t(E/I)$, and therefore:

$$\delta_t(E) = P(E/I).(1 - kr^{-1}).P(I) \quad (9')$$

In general, this condition can be checked only if several randomized clinical trials have been completed with a clinical event as the endpoint.

2. P(E/nI) > 0

The clinical event may occur in patients who do not have any risk factors, e.g. patients with normal blood cholesterol levels and blood pressure can suffer from CHD.

2.1. If the treatment does not affect [I-E]

In this situation $P_t(E/I) = P(E/I)$ and $P_t(E/nI) = P(E/nI)$, and therefore $L_t = L$. From Equation 5':

$$\delta_t(E) = P(E/nI).[L-1].\delta_t(I) \quad (10')$$

The second proviso is satisfied because P(E/I) > P(E/nI), since the intermediate endpoint is a risk factor for the clinical event. The third proviso is satisfied if estimates for P(E/I) and P(E/nI) are available from the data from previous epidemiological studies.

---

[1] if P(E/I) is not constant, $\delta(E)$ can be derived from $\delta(I)$, if the function which links P(E/I) to P(I), written as H[P(I)], is known

2.2. *If the treatment modifies [I-E]*

In this case $P_t(E/I) \neq P(E/I)$, and $P_t(E/nI) \neq P(E/nI)$. Also $L_t \neq L$, and in general, $\delta_t(E)$ is no longer a simple function of $\delta_t(I)$. Equation 5′ becomes:

$$\delta_t(E) = P(E/nI).\{P(I).[L.(1\text{-}rk)\text{-}1 + r'k] + 1\text{-}r'\} \quad (11')$$

The second and third provisos can be checked only if k, r, r', P(E/I) and P(E/nI) are all known. This is possible in the unlikely case where r and r′ are independent of the type of treatment, and if previous randomized clinical trials with a clinical event as the endpoint have yielded sufficient data, and if precise estimates of P(E/I) and P(E/nI) are available from the data from previous epidemiological studies. If *r* and *r′* are taken to be the constant values of r and r′, respectively, Equation 10′ then becomes:

$$\delta_t(E) = P(E/nI).\{P(I).[L.(1 - rk)\text{-}1 + r'k] + 1 - r'\} \quad (12')$$

There is an exceptional case where the effects of the treatment on the intermediate endpoint and the clinical event are:

- based on the same pharmacodynamic mechanism (i.e; the treatment acts on the same receptor to evoke the effects), and thus they are statistically correlated, but
- the intermediate endpoint is not a causal link in the sequence of events which results in the occurrence of the clinical event, although they share a common remote cause (see model no. 2, Fig. 2).

In this case the effect of the treatment on [I-E] is said to be isotropic because r = r′ (another isotropic effect is seen when P_t (E/I) = P (E/nI). The comments made in section 2.1. (above) apply and Equation 11′ is reduced to:

$$\delta_t(E) = P(E/nI).\{P(I).[(L - 1).(1 - rk)] + 1 - r\} \quad (13')$$

It is simple to verify that $\delta_t(E)$ is always positive if is greater than I, which is always true if I is a risk factor for E, and r and k are less than, i.e. the treatment is effective on both I and E. If r is independent of the type of treatment, r becomes *r*, which is constant, and therefore the second and third provisos can be verified if sufficient data is available from epidemiological studies and randomized clinical trials with clinical events as endpoints. Equation 12′, thus, becomes:

$$\delta_t(E) = P(E/nI).\{P(I).[(L - 1).(1 - rk)] + 1 - r\} \quad (14')$$

References

1. Friedman L, Furberg CD, DeMets DL (1984) Fundamentals of clinical trials, 2nd edn. PSG Publishing, Littleton
2. Moleur P, Boissel JP (1987) Definition of a surrogate endpoint. Controlled Clinical Trials 8: 304
3. Herson J (1989) The use of surrogate endpoints in clinical trials (an introduction to a series of four papers). Stat Med 8: 403–404
4. Ellenberg SS, Hamilton JM (1989) Surrogate endpoints in clinical trials: cancer. Stat Med 8: 405–413
5. Wittes J, Lakatos L, Probstfield J (1989) Surrogate endpoints in clinical trials: cardiovascular diseases. Stat Med 8: 415–425
6. Hillis A, Seigel D (1989) Surrogate endpoints in clinical trials: ophthalmologic disorders. Stat Med 8: 427–430
7. Ruskin JN (1989) Editorial: the Cardiac Arrhythmia Suppression Trial. New Engl J Med 321: 386–388
8. Winkle RA (1980) Detection of patients at high risk for sudden death: the role of electrocardiographic monitoring. In: Kulbertus HE, Wellens HJJ (eds) Sudden death. Martinus Nijhof, The Hague
9. The CAST investigators (1989) Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. New Engl J Med 321: 406–412
10. Kannel WB, Gordon T, Schwartz MJ (1971) Systolic versus diastolic blood pressure and risk of coronary heart disease. The Framingham study. Am J Cardiol 27: 335–346
11. Pooling Project Research Group (1978) Relationship of blood pressure, serum cholesterol, smoking habit, relative weight and ECG abnormalities to incidence of major coronary events: final report of the Pooling Project. J Chron Dis 31: 201–306
12. MacMahon S, Petro R, Cutler J, Collins R, Sorlie P, Neaton J, Abbott R, Godwin J, Dyer A, Stamler J (1990) Blood pressure, stroke, and coronary heart disease. Part 1. prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. Lancet 335: 765–774
13. Boissel JP, Gueyffier F (1990) High blood pressure and clinical events linked with atherosclerosis. In: Atkinson J, Capdeville C, Zannad F (eds) Coronary and cerebrovascular effects of antihypertensive drugs. Transmedico Europe Ltd, UK, p 240–252
14. Collins R, Peto R, MacMahon S, Hebert P, Fiebach NH, Eberlein KA, Godwin J, Qizilbash N, Taylor JO, Hennekens CH (1990) Blood pressure, stroke, and coronary heart disease. Part 2. short-term reductions in blood pressure: overview of randomized drug trials in epidemiological context. Lancet 335: 827–838
15. Okamoto K, Yamori Y, Nagaoka A (1974) Establishment of the stroke-prone spontaneously hypertensive rat (SHR). Circ Res 34: I143–I153
16. Lievre M, Leizorovicz A, Boissel JP (1991) Critères intermédiaires et critères de substitution dans le développement des antiarythmiques. Arch Mal Coeur 84: 27–33
17. Hine LK, Laird NM, Hewitt P, Chalmers TC (1989) Meta analysis of empirical antiarrhythmic therapy after myocardial infarction. J Am Med Assoc 262: 3037–3040
18. Slater W, Lampert S, Podrid J, Lown B (1988) Clinical predictors of arrhythmia worsening by antiarrhythmic drugs. Am J Cardiol 61: 349–353
19. Prentice RL (1989) Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med 8: 431–440

J.-P. Boissel
Unité de Pharmacologie Clinique
Boîte Postale N° 3041
162 avenue Lacassagne
F-69394 Lyon Cedex 03
France