Ann. Ist. Super. Sanità, vol. 29, n. 1 (1993), pp. 3-14

Reproductive toxicology guidelines: comparison and application

Diana BELTRAME and Guy MAZUÉ

Farmitalia Carlo Erba, Nerviano (Milan), Italy

Summary. - Reproductive toxicity studies currently recommended by the three principal regulatory agencies, the United States Food and Drug Administration (FDA), the Committee for Proprietary Medicinal Products (CPMP) of the European Economic Community and the Japanese Ministry of Health and Welfare (MHW), have a three-segment design, with essentially similar objectives in identifying any possible adverse effects of medicinal products on all stages of the reproductive process in animals, in order to evaluate the potential risk in man. However, differences exist between the various guidelines which give rise to considerable difficulties in amalgamating experimental designs to comply with all three agencies. The main differences are between Western and Japanese recommendations and can be identified in two points which are cause for debate and form an obstacle to the mutual acceptance of studies: a) the treatment periods during pregnancy and b) the extent of studies on the progeny reared to maturity. Both points concern solely studies in rodents and are based on a different approach to the subject. Advantages and disadvantages of the differences in each study segment are considered in relation to practical applications. In comparing recommendations from different agencies, shortcomings in the instructions and nebulous or questionable requirements, but also valuable directives, are highlighted, in the hope that regulatory authorities can be encouraged to provide exhaustive information and instructions and more explicit policies in the new coordinated guidelines which are expected as a result of international harmonization. To this end, the need for greater flexibility is stressed, since the conventional designs of the segments often prove inapplicable or are deemed inadequate or unnecessary in the case of drugs whose pharmacological activity interferes with the reproductive process or which are intended for particular therapeutic modalities or purposes. In particular, regulatory authorities are urged to provide specific, coordinated guidelines for antitumor agents, taking into account that their technical application in animals should reflect treatment modalities and therapeutic uses in humans.

Key words: reproductive toxicity, current guidelines, pharmaceuticals.

Riassunto *(Direttive per la tossicità della riproduzione: comparazione e applicazione).* - Gli studi di tossicità della riproduzione attualmente raccomandati dalle tre principali agenzie regolatorie, la Food and Drug Administration (FDA) per gli Stati Uniti, il Comitato per la Valutazione delle Specialità Medicinali (CPMP) per le Comunità Europee e il Ministry of Health and Welfare (MHW) per il Giappone, consistono in un disegno sperimentale a tre segmenti il cui obiettivo, essenzialmente simile, è quello di identificare ogni eventuale effetto nocivo dei prodotti medicinali su tutti gli stadi del processo riproduttivo negli animali, per valutare il potenziale rischio nell'uomo. Tuttavia, le differenze esistenti tra le direttive non consentono di mettere insieme disegni sperimentali che siano in reciproca conformità. Le principali discrepanze si riscontrano tra le direttive occidentali e quelle giapponesi e sono imperniate fondamentalmente su due punti, fonte di discussione ed ostacolo per una mutua accettabilità degli studi: a) i periodi di trattamento durante la gravidanza e b) l'estensione degli studi sulla progenie allevata fino a maturità. Entrambi i punti riguardano soltanto gli studi nei roditori e sono motivati dal differente approccio con cui è affrontato il tema. Vantaggi e svantaggi di queste differenze sono presi in esame, nell'ambito di ciascun segmento di studio e in rapporto con le applicazioni pratiche. Inoltre, nel confronto tra le direttive delle varie agenzie regolatorie, si sottolineano carenze, richieste poco chiare od equivoche, ma anche direttive apprezzabili, per sensibilizzare le autorità regolatorie a delineare nuove coordinate direttive, attese nell'armonizzazione internazionale, che riportino informazioni ed istruzioni esaurienti e chiare linee di condotta. A tale proposito, si evidenzia la necessità di una maggiore flessibilità, dato che i disegni convenzionali dei segmenti si verificano sperimentalmente inapplicabili o si giudicano inadeguati o inutili nel caso di farmaci con attività farmacologica interferente con i processi riproduttivi, o deputati a peculiari modalità e scopi terapeutici. In particolare, si invitano le autorità regolatorie a redigere in comune accordo direttive specifiche per gli agenti antitumorali, considerando che l'applicazione tecnica nell'animale dovrebbe rispecchiare modalità ed impieghi terapeutici nell'uomo.

Parole chiave: tossicità della riproduzione, direttive attuali, farmaci.

Introduction

Reproductive toxicity studies represent a predominant part of the pre-clinical safety evaluation of candidate compounds for human therapy. A primary concern of all involved in developing new drugs is to avoid a repetition of the not so distant thalidomide disaster which, though having darkened a far from brief period of our recent history, has paradoxically been defined by some authors (Hottinger, 1964; Tuchmann-Duplessis, 1964 cited by Vichi) [1] as a "fortunate event" for its repercussions on safety evaluation procedures. It not only clarified a basic principle of teratogenesis, namely that the variability of teratogenic effects is species-dependent, but it also promoted a systematic approach to experimental teratology, making it a mandatory requirement that all

drugs be routinely tested for their safety of use during pregnancy and in women of childbearing potential before being placed on the market.

However, this part of the safety evaluation of drugs cannot be limited to an appraisal of their effects on pregnancy and particularly their potential for inducing malformations, although this is the most striking aspect from an ethical viewpoint. The entire sexual-reproductive sphere must be covered, from the possible impairment of fertility in both sexes to direct or indirect transmission of toxic effects to later generations.

To address all these problems, in 1966 the US Food and Drug Administration (FDA) published the first official regulatory guidelines for *in vivo* reproductive studies for safety evaluation of drugs for human use [2]. These guidelines indicate an experimental design that fits the purpose, with studies divided into three segments: I) Study of Fertility and General Reproductive Performance; II) Teratological Study; and III) Perinatal and Postnatal Study. The experimental procedures recommended in this document were to satisfy international regulatory authorities for a decade, serving as the principal reference for reproductive toxicity study protocols, despite the fact that from the time they were issued to 1980 many industrialized countries [3] and international organizations [4, 5] produced their own documents containing guidelines for reproduction studies with pharmaceutical products.

The need arose to harmonize the different European regulatory requirements and in 1983 the European Economic Community (EEC), on the recommendation of the Committee for Proprietary Medicinal Products (CPMP), published the first official guidelines for reproduction studies on pharmaceuticals [6]. In 1984 the Japanese Ministry of Health and Welfare (MHW) published its own official guidelines for pharmaceuticals [7], after a number of notifications [8] issued from 1965 onwards. The Japanese guidelines are followed by the majority of Asian countries.

The European and Japanese regulatory guidelines do not modify the existing US guidelines in principle, but they do introduce some innovations of fundamental importance for their contribution to the information provided by reproductive toxicology studies and for their economic implications, in that they make the studies longer and more costly. It has been realized, however, that the differences between the reproductive toxicology guidelines of the United States, Europe and Japan give rise to considerable difficulties in amalgamating experimental designs to comply with all these agencies. In recent years this has prompted regulatory authorities and industrial associations to move towards an international harmonization of guidelines in an effort to save time and resources, to the advantage of more rapid development of new drugs.

Some comparative aspects of the guidelines issued by the three most authoritative international regulatory agencies (FDA, CPMP and MHW) are discussed below in the light of their practical application.

Comparison of current guidelines for reproduction studies

The reproductive toxicity studies currently recommended by the three principal regulatory agencies (FDA, CPMP and MHW) which supervise world drug marketing have a three-segment design aimed at identifying possible adverse effects on all stages of reproductive performance, from changes in fertility and damage to the male or female gametes to late effects on the progeny. These can be summarized as follows (Fig. 1):

I) *Fertility and general reproductive studies* are performed in at least one animal species (normally the rat) in which males and females of the parental generation (F_0) are treated for a sufficiently long period prior to mating, during mating and up to sacrifice (with the exception of females for Japan). Treatment of the females continues through pregnancy (only early pregnancy for Japan) and up to weaning. All the females MHW or half of them are sacrificed at mid-pregnancy (FDA) or just before parturition (CPMP, MHW), at which time pregnancy parameters are determined and the fetuses are examined. The remainder are allowed to deliver naturally and rear their young to weaning. In special cases (FDA) or routinely (CPMP) selected offspring (F_1) from each litter at each dose level are reared (untreated) to maturity and examined for late effects of the drug in terms of auditory, visual or behavioural impairment and reproductive function.

II) *Teratology/embryotoxicity studies* are conducted in at least two animal species (usually rats and rabbits). The females (F_0) are treated during the period of fetal organogenesis and sacrificed shortly before the expected date of parturition. Pregnancy parameters are determined and the fetuses are examined. Japanese guidelines recommend additionally that for rodents one-third of the dams should normally be allowed to deliver and nurse their young. Selected offspring (F_1) from each litter and group are reared (untreated) to maturity for observation of growth and development, morphological, functional and behavioural examinations, and investigation of reproductive performance.

III) *Peri- and postnatal studies* are conducted in at least one animal species (usually the rat), in which the period of drug administration to the dam (F_0) should cover the final one-third of gestation from the end of organogenesis to parturition and continue throughout lactation to weaning. In certain cases (FDA, CPMP) or routinely (MHW) selected offspring (F_1) are reared to

maturity (untreated) so that their reproductive capacity and other late effects of the drug on growth and development as well as any signs of morphological, functional and behavioural impairment can be assessed, as in studies in the previous segments.

It is clear from the above that the three segments recommended by the three principal regulatory agencies are similar but not identical, making it impossible in practice to prepare an internationally acceptable study package. This means that if all regulatory requirements are to be fulfilled, studies must inevitably be repeated or the segments must be amalgamated into protocols satisfying the requirements of each individual regulatory agency in order to cover all markets [9]. This is time-consuming and leads to a surfeit of duplicated data which more often than not can prove equivocal or give rise to confusion when similar studies do not generate identical data.

Whilst the US and Europe have attempted to adapt their experimental procedures to mutually acceptable study protocols, the Japanese authorities are very reluctant to accept reproductive study packages conducted in other countries [9] and only recently have they shown a certain willingness to consider studies conducted according to other guidelines on a case by case basis. This is with a view to the expected international harmonization of guidelines and should serve to reduce duplications of studies between Japan and other countries in the case of toxicological studies expected to lead to the same conclusions, though conducted according to methods other than those specified by Japanese guidelines [8].

The main differences between the general directives of US and EEC guidelines on the one hand and Japanese guidelines on the other centre essentially on two points which are a constant cause for debate and uncertainty in planning reproduction studies and represent a major obstacle to the mutual acceptance of existing studies:

a) the treatment periods during pregnancy, and

b) the extent of studies on the progeny reared to maturity.

Both points concern solely studies in rodents. To understand the significance of these differences we must consider the different approach of the regulatory authorities to studies of drug effects on reproductive and developmental toxicity. Western guidelines stress the importance of identifying any kind of impairment at any moment in the reproductive process (Segment I), then going on to in-depth studies of the critical phases of pregnancy with investigation of embryofetotoxicity and teratogenic potential (Segment II) and of the effects on parturition (Segment III). The sequential nature of the segments is only theoretical, however, since priorities are actually established according to other criteria (e.g.

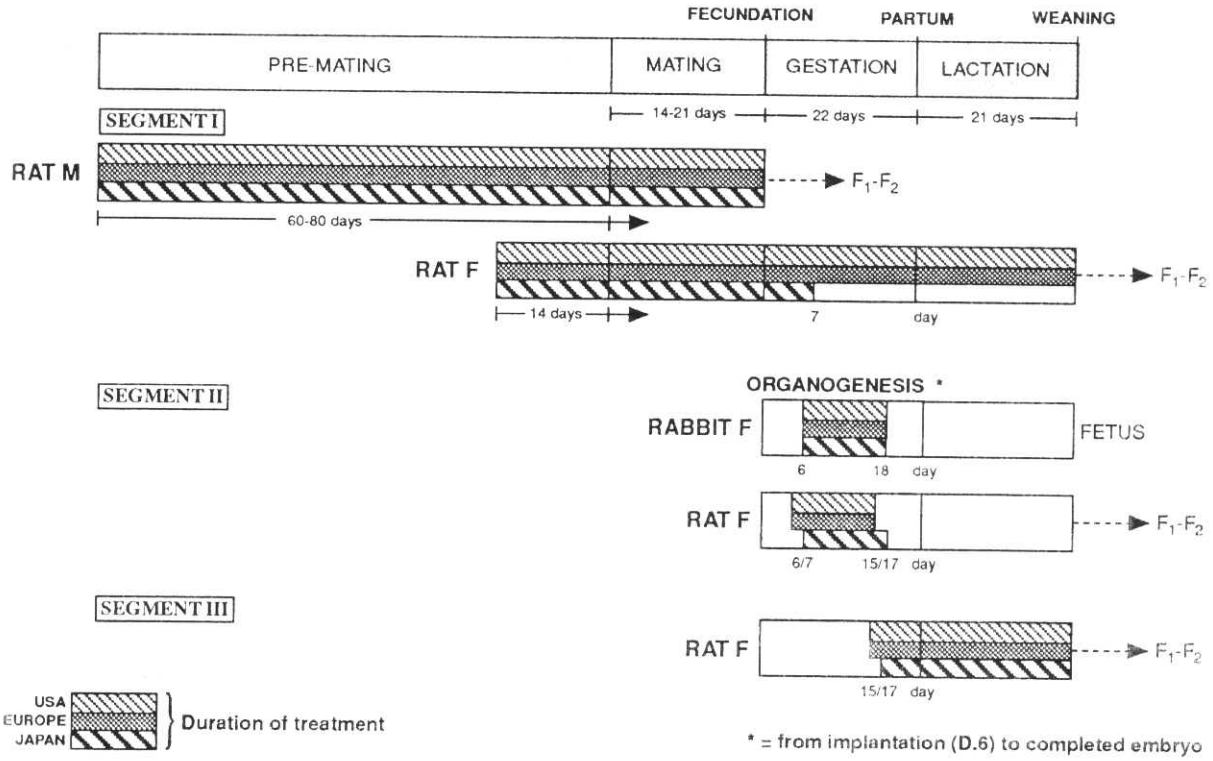

Fig. 1. - Reproductive toxicity tests recommended by the three principal regulatory agencies: FDA (USA), CPMP (EEC) and MHW (Japan).

to comply with specific government requirements or company policies, or to support chemical development plans).

Japanese guidelines are designed to give special consideration to accurately understanding the adverse effects on reproduction by dividing the time from pregnancy to weaning into three periods, each of which is employed as the administration period: from successful copulation until the beginning of fetal organogenesis (Segment I), the period of fetal organogenesis (Segment II), after the estimated period of fetal organogenesis ends (Segment III). This approach is based on the hypothesis that a long period of consecutive administration may possibly cause excessive effects on the fetuses to be examined, thus leading to improper interpretation of the results.

As regards the treatment period, the main point of divergence between the guidelines is Segment I (Fig. 1). Whilst the treatment period for males is comparable, treatment of females ends at implantation for Japan (Day 7 in the rat, Day 6 in the mouse, Day 0 being the day of copulation confirmed by sperm-positive smear or plug) and continues until termination (date of interim sacrifice or at weaning) for the US and EEC.

However, the primary objective of Japanese Segment I studies, in addition to exploring the effects on parental fertility, is to investigate developmental abnormalities in the progeny (fetuses and/or offspring delivered naturally) [8] which may be caused precisely by administering the drug prior to and in the early stages of pregnancy, whereas US and EEC studies are designed to explore all aspects of fertility and general reproductive performance of the parental generation, including developmental abnormalities in the progeny (fetuses and/or offspring delivered naturally), and should serve as a guide for subsequent in-depth studies [2]. This explains the importance of a long treatment period covering all the stages investigated.

Which of the two tests is the most valid has long been a matter for debate, on both scientific and economic grounds. The possibility of distinguishing right from the start exactly which phase of the reproductive process is affected, rather than producing a compendium of indiscriminate effects, might be a more valid and immediate aid in extrapolating the data to man. This view is supported by the fact that in practice findings that have emerged from US or EEC design Segment I studies frequently require subsequent investigation according to a non-classical experimental design and with treatment divided into shorter sub-periods or using special studies in order to identify cumulative effects and reach logical conclusions that can aid clinical evaluation of the results. From this standpoint, Japanese Segment I studies might be more appropriate, saving time and resources, if we consider that Segments II and III can provide all necessary complementary information without cumulative effects and that the sequence of treatment periods in the three segments does not leave any phase of the reproductive process uncovered. In addition, the overall shortening of the treatment period, which during pregnancy is limited to the very early stages, may be an advantage in that it often permits the use of higher dosages that are better tolerated in animals, affording a wider safety margin. However, in practice even this type of treatment, terminating immediately after implantation in the species, is still not sufficient in some cases to distinguish effects occurring early in pregnancy, for example a pre-implantation embryolethal effect (i.e. embryotoxicity), from an anti-implantation effect (e.g. due to the drug's specific pharmacological activity).

On the other hand, when Segment I studies conducted according to EEC standards do not reveal toxic effects at dosages that guarantee adequate safety margins with respect to clinical use, they might be considered so reassuring that Segments II and III would be a pointless repetition. EEC guidelines have in fact introduced two important changes to the US Segment I procedures, transforming them into a combined, exhaustive study of all the possible effects of a drug as regards reproductive and developmental toxicity.

Firstly, EEC guidelines postpone sacrifice of half of the treated dams from mid-pregnancy (Day 13 of gestation) to just prior to parturition (Day 20 of gestation in practice). The advantage of this modification is that all measurements and evaluations of the conceptus are carried out exactly as if it were an embryotoxicity study (Teratological Study/Segment II). Secondly, EEC guidelines request routine examination of any late effects of a drug on the progeny, as regards important indices of physical, functional and behavioural development, including studies of reproductive function, a recommendation that was mentioned only in vague terms in the FDA guidelines. The latter addition should also make the study qualify for evaluation by the Japanese authorities, as it provides the information they require for Segments II and III and which are generally lacking in Western protocols for the same studies. In practice, the data generated in a Segment I study can replace the assessment of postnatal growth and development of the offspring in Segment II or III studies [8].

Other discrepancies in the treatment periods during pregnancy are found in studies during fetal organogenesis (Segment II) and peri- and postnatal studies (Segment III), but in the rat only. The US guidelines for Segment II establish that the treatment period in this species should be from Day 6 to 15 of pregnancy (Day 0 = sperm positive smear) and this period is conventionally left unchanged in EEC protocols, whereas Japanese guidelines require treatment from Day 7 to Day 17 (Day 0 = copulation confirmed). In Segment III the start of treatment is usually taken to be Day 15 of pregnancy in US and EEC protocols [10], as it is not specified in the

guidelines, whereas Japanese guidelines specify Day 17 of pregnancy. These apparently minimal shifts take on great importance in the practical application of the guidelines when testing drugs that may inhibit implantation (e.g. anti-prolactins and anti-estrogens) or modify the physiological mechanisms of parturition (e.g. non-steroidal anti-inflammatory drugs, prostaglandins, anti-estrogens). With these drugs, therefore, it is highly likely that studies conducted according to different guidelines will yield different results. For example, the anti-implantation effect will be more evident in a US or EEC Segment II study and interferences with parturition (anticipation, delay, impaired expulsion) may be found in a Segment II study according to Japanese guidelines (parturition is not envisaged in US and EEC studies), whilst they will be more severe in a Segment III according to US and EEC protocols than in a Japanese Segment III. All these possibilities demand careful consideration on a case by case basis and in the light of existing pharmacological knowledge of the product. This is true not only when planning reproduction studies, with an eye to the expected marketing areas and the regulatory agencies that must be satisfied, but also when revising the final documents.

Post-weaning investigations (physical, functional and behavioural development and reproductive performance) on the offspring brought up to maturity are the second major point of discrepancy between Western and Japanese guidelines, and are again a matter that should be seen against the background of the different approaches to the problem of reproductive toxicity. EEC guidelines (Table 1) recommend that these investigations be included in Segment I, thus extending this requirement to all drugs, whereas in US guidelines it applied only to "special cases".

According to Western regulatory requirements and with a view to saving resources, priority is obviously given to Segment I, because the length of treatment envisaged (male and female F_0, and from premating to weaning) makes it the study that affords the most opportunity for detecting any late effects on the development of the progeny and it also serves as a guide for subsequent in-depth studies.

Examination of offspring after birth is not usually performed in the Segment I study according to Japanese guidelines, but is strongly recommended in a recent revision of these guidelines [8], especially when other information (data on existing similar drugs or results of other toxicology studies) suggests that disorders in the postnatal development of the offspring may occur as a result of Segment I treatment. Therefore, when appropriate, examinations should be considered on offspring delivered naturally from some or all of the mothers in a Segment I study, as in Segments II and III.

Western regulatory requirements also call for investigation of postweaning development in Segment III, but the recommendations are vague, stating that "Because some drugs may produce effects not detectable in early life, some of the offspring may have to be raised to adulthood" [2] or "under certain circumstances" [6] and it is not clear when and if these investigations are mandatory and under what circumstances studies lacking

Table 1. - Postweaning investigations on F_1 offspring

Study type	FDA (USA) [2]	CPMP (Europe) [6]	MHW (Japan) [7]
Segment I	±	+	±
	Reproductive function (in special cases, e.g. sex steroids)	Auditory and visual function, behavior, reproductive function	Not applicable (recommended [8] when appropriate)
Segment II	-	-	+
	Not applicable	Not applicable	Morphological and behavioral development, reproductive function
Segment III	±	±	+
	Behavior and reproductive function (when appropriate)	Auditory and visual function, behavior, reproductive function (under certain circumstances)	Morphological, functional and behavioral development, reproductive function

this section will be accepted or rejected by the regulatory authorities of Western countries. Since the onus of choice rests with the pharmaceutical companies, it is important when drawing up a study program to bear in mind that, on the basis of existing knowledge of the drug and information from other reproduction studies, peri- and postnatal studies can show up developmental defects in the progeny more distinctly than a Segment I study. All effects on the preceding phases (parental fertility, pre-implantation and organogenesis) that might be misleading or deflect judgment are eliminated and attention is focused on the most appropriate period (at least in rodents) for eliciting effects on functional differentiation that result in defects such as alterations of the immune, endocrine and central nervous systems and effects on secondary sexual characteristics, leading to reproductive deficiencies. All these effects can be detected only if the progeny is followed to adulthood.

Japanese guidelines attach great importance to the detection of abnormal functional development, requiring that this section of studies be conducted as part of Segment III for all drugs. Studies conducted in other countries with insufficient data on growth and development of the offspring are therefore not normally accepted by Japanese authorities.

Investigations of the functional and behavioural development of the offspring are not applicable in Western Segment II studies which do not include natural delivery, whereas they are required for Segment II studies in Japan, which include a proportion of dams that are allowed to deliver naturally and rear their young to weaning. The Japanese authorities attach great importance to investigation of postnatal development after maternal treatment only during the period of fetal organogenesis. The organogenetic period (morphogenesis and organogenesis) is the most critical phase of the development of the embryo and the stage at which it is most sensitive to the teratogenic action of exogenous agents, so fetal disorders (such as retarded development, malformations and death) but also disturbances in postnatal growth, development and different functions can occur and should be evaluated at precisely those dosages which give no apparent indication of teratogenicity. The literature shows that behavioural deficits can be induced even during organogenesis. Segment II studies in rodents that do not include this study phase are not normally accepted by Japanese authorities.

The guidelines provide no definite indications as to the methods and procedures to be employed for evaluation of postnatal development and of the various functions in the offspring. This aspect has long been debated by many associations and study groups and is still a matter requiring evaluation. A number of testing methods have been proposed and recommended by experts and adopted in reproduction laboratories. However, little attention seems to be paid to whether these tests are suitably chosen and correctly applied. Since no categories of tests have been defined for application to different classes of drugs or to all drugs, each laboratory has chosen its own standard battery of tests among those in current use, but this is often destined to become a set procedure to be applied indiscriminately to all drugs. The items for study are numerous and time-consuming, and a large mass of data is produced, therefore there is a tendency to simplify procedures, for example by modifying the original tests (e.g. behaviour) and concentrating observations at the theoretical times at which events or responses are expected (e.g. physical development), with the result that the findings may be invalidated and precious subtle information may be lost.

Since it is generally recommended that batteries of tests be applied in a number directions to increase the probability of identifying a possible developmental or behavioural deficit, many investigators launch into a vast number of observations, to the detriment of quality, thus generating confused, inconclusive data. In view of the great attention paid to the postnatal development of the offspring and the considerable resources invested in this study section, it is to be hoped that the regulatory authorities will soon provide coordinated guidelines for tests and procedures, in order to avoid the risk of encumbering reproductive toxicity tests with a heavy burden of questionable data, produced solely to satisfy regulatory requirements.

In the context of research into late effects on the offspring, the procedure for assessing the reproductive performance of the F_1 generation also warrants the attention of the regulatory authorities. EEC guidelines specify that reproductive function should be determined in the progeny by allowing at least one male and one female from each litter of dosed animals to breed and produce one litter (Segment I) or to assess reproductive capacity (Segment III). Japanese guidelines state that the reproductive performance of the offspring should be examined on the basis of successful pregnancy. Since no indication is given as to how to proceed, once the pregnant F_1 animals have been obtained (in fact at the time the protocol is prepared), some doubt still remains as to when and how the conceptus (F_2) should be evaluated, in order to obtain adequate information while saving resources and at the same time satisfying all regulatory requirements. In practice, by general consensus, the choice is between the two procedures routinely applied to the F_0 generation: a) Day 20 sacrifice and b) natural delivery and rearing up to weaning, but if possible reduced to a minimum to save time and money (for example, the fetuses are examined only externally and preserved against the contingency of visceral and skeletal examination, which generally does not prove necessary; the neonates are observed for survival, weight gain and some indices of physical development, but whether they have functional deficits will never be known because they are sacrificed at weaning).

Thus if the primary objective were simply to evaluate the "successful pregnancy" and "reproductive capacity" of the F_1 progeny, an interim sacrifice (e.g. at Day 13 of pregnancy) could in fact be considered sufficient as a first routine approach and would also be less costly.

The choice of doses is another point of the guidelines that warrants consideration, as it may be a cause for rejection by Japanese authorities in the case of studies conducted in Western countries. Developmental toxicity studies should normally be conducted at (or at least at) three dose-levels. The high dose should evidence minimal maternal toxicity according to all guidelines. The intermediate dose should, in principle, be the geometric mean of the highest and the lowest doses. The lowest dose should be a low multiple of the proposed therapeutic dose (FDA), should be sufficient to produce in the species a pharmacodynamic effect similar to the desired therapeutic effect (EEC), or should not cause any adverse effect in parental animals, fetuses or young (MHW).

Determination of the no-observable-effect-level (NOEL) is a fundamental requirement of Japanese guidelines and an essential condition for studies conducted in other countries to qualify for consideration. This concept of the NOEL, based on consideration of the risk/benefit ratio in therapeutic use, denotes an entirely different approach from that of Western guidelines as regards the choice of the low dose and doubles the study objectives (observation of toxic effects and assessment of the safe dose). In the view of the MHW it is not sufficient to know that a drug damages reproduction in the parental generation and development of the derived generation, but it is also essential to establish the no-effect dose level for each generation. This means that even a dose that causes only effects due to the drug's pharmacological action cannot be considered the NOEL. In practice, when using a species in which a given drug shows marked pharmacological activity, submultiples of the expected therapeutic dose may have to be used to determine the NOEL and satisfy Japanese regulatory requirements.

Another matter that requires standardization is the question of controls. US guidelines state that a negative control is essential and it is often advantageous to include a positive control group. EEC guidelines make no mention of control groups, perhaps because they accept the US recommendations as valid. Japanese authorities place great emphasis on this aspect, recommending use within each study segment of a negative control group (whose data are directly subject to analysis of study results, it being desirable for the results to be supplemented by reference to background data obtained in the past) and also a positive control group (i.e. dosed with a substance known to have a potent reproductive toxicity/teratogenicity) or a comparative control group (i.e. dosed with an available drug with a similar chemical structure or pharmacological effects), if necessary. However, these specifications merely serve to clarify the recommendations already contained in US guidelines.

The routine use of these additional control groups is certainly impractical if not pointless. Positive or comparative controls (reference compound) serve a purpose only when it is presumed that the test substance may cause similar effects to those of known drugs with similar chemical structure and effects. However, Japanese guidelines also recommend the use of two negative controls when the solvents or emulsifiers used in the formulations contain additives or excipients that might induce adverse effects on reproduction. In these cases a negative control should receive such vehicles and additives alone and the other control (the true negative control) should receive the inert vehicle, without excipients, or could even be completely untreated. This indication, though generally applied in practice, is lacking in Western guidelines.

Another factor which gives rise to considerable uncertainty when it comes to planning studies of drugs intended for all markets is the size of the dose groups, i.e. the recommended number of animals for developmental toxicity studies (Table 2), which carries a considerable weight in the assessment of the safety of drugs and in the economic aspects of a study. The recommendations of the major regulatory agencies do not differ substantially but are somewhat nebulous. General recommendations to use a given number of animals, without any clear indication, lend themselves to subjective interpretation, so that if a company chooses to cut costs by using the minimum required number of animals, it may risk invalidating the studies (obviously this danger can be avoided by using a number of animals far in excess of requirements). In Segment I studies according to US and EEC guidelines, for example, it is not clear whether the number of animals/dose is that required to start a study (pre-mating) or refers to the number of animals to be used for mating (i.e. Japan) and evaluation of fertility and reproductive performance, in which case a larger number of animals would seem to be necessary to avoid common problems (e.g. fluctuations of the normal fertility rates in the strain used, drug-related or accidental deaths which would reduce the population examined).

Another question is whether the number of rodents in US Segment II and III studies refers to sperm-positive females or females with confirmation of pregnancy. In US and EEC Segment II studies the term "pregnant" is not sufficiently explicit. It could refer to all evaluable pregnancies (including those with maternal death, abortions and total resorptions) or only to pregnancies with fetuses at term, in which case the number of animals per dose for evaluation might prove lower than that recommended because of marked toxicity or pharmacological activity. The same considerations apply to Segment III.

Table 2. - Minimun number of animals per dose group recommended in developmental toxicity studies

Study type	Animal species	FDA (USA) [2]		CPMP (Europe) [6]		MHW (Japan) [7]	
		M	F	M	F	M	F
Segment I	Rodents	10	20	24	24	20 (for matings)	20
Segment II	Rodents		20		20 pregnant		30 with successful pregnancy (1/3 allowed to deliver spontaneously)
	Non-rodents		10 pregnant (rabbits)		12 pregnant		12 with successful pregnancy (rabbits)
Segment III	Rodents		20		12 pregnant		20 with successful pregnancy

Japanese guidelines are a little more explicit in that they state that the number of animals indicated means those with successful pregnancy in Segments II and III: "the term animals with a positive indication of pregnancy or pregnant animals denotes those in which implantation sites are demonstrated in the uterus at necropsy" [8]. It can therefore be assumed that the recommended number includes maternal deaths, abortions and total resorptions.

As regards the number of animals for studies in other species, all guidelines provide only general suggestions, stating that an adequate/sufficient number of animals should be used to allow a clear conclusion to be drawn from the study or to permit a meaningful assessment of safety or to yield appraisable data. In practice, the use of less well-known species than the mouse, rat, hamster or rabbit (e.g. monkey, guinea pig) in reproductive toxicity studies would lead to considerable difficulties in establishing a suitable number of animals to allow a meaningful assessment of safety.

There are also some differences in the recommendations for certain procedures (e.g. structural examination of rodent fetuses, incubation of rabbit fetuses, culling of neonates in the first week of life, etc.) which are not a cause for debate as regards the mutual acceptability of reproduction studies but fall within the operating choices of individual laboratories.

Application

The developmental toxicity study segments outlined by the various guidelines represent the conventional tests required by regulatory authorities as the basis for screening of new drugs, in order to evaluate the potential risk in man. Fertility and teratogenicity studies must be carried out before extending clinical trials to pregnant women and individuals of reproductive age. Peri- and postnatal studies must be performed at least prior to the new drug application. As a first approach Segment II studies (fetal organogenesis/teratogenesis) are generally done in the two conventional species, rat and rabbit, to verify the drug's embryotoxic and teratogenic potential and establish whether or not it can be used in pregnant women. The tests are then extended as quickly as possible to fertility and general reproductive performance (Segment I) and peri- and postnatal studies (Segment III) in at least one species.

The suitability of the species should be demonstrated prior to these tests in general pharmacology studies and pharmacokinetic studies in pregnant animals, as recommended by all the guidelines. Species and strains that metabolize the product in a similar fashion to man should normally be chosen. Before starting tests that include maternal treatment during lactation it is extremely useful to discover whether the drug is excreted in the milk. In practice, however, these data are not always available prior to the start of reproduction studies, therefore the usual species recommended by the guidelines are employed (generally the rat and the rabbit, more rarely the mouse) as their general metabolic mechanisms are known. Obviously provision is made for given studies to be repeated in another species if necessary.

Regulatory authorities generally require that the same species be used in all three segments and that the species used for reproductive toxicity studies be one of those already selected for long-term toxicity studies. Obviously

full background data on the species and strain must be available and should serve as reference parameters for interpretation of the results (background pregnancy and litter data, background data for spontaneous malformations and susceptibility to teratogens).

The doses for a main study are best chosen in a range of doses tested preliminarily in pregnant animals of the same species. This is because a drug's toxicity may be expressed differently in pregnant or lactating animals with respect to non-pregnant animals. The doses for the preliminary studies can be obtained from general toxicity studies (3 and 6 months) or else from dose-range finding studies in non-pregnant animals if general toxicity studies are not available (e.g. in the rabbit). If preliminary studies are to provide correct indications for selection of the doses, the protocols must scrupulously include at least all the same records with the same frequency as in the main study, though on a smaller number of animals, and must generally comply with the guidelines, with the exception of those minimal, rational modifications required to save time.

Thus the duration of pre-mating treatment of the males in a preliminary Segment I study should cover at least the duration of a spermatogenic cycle (ideally males from three-month general toxicity studies could be used), whereas treatment of lactating dams and postnatal studies on the progeny could cover only the first week of lactation. Visceral and skeletal examination of the fetuses in preliminary Segment II studies should be considered necessary at least to investigate or confirm any indication of teratogenicity emerging from the external examination. If this effect is confirmed, however, the utility of performing a main study becomes questionable, since the drug can already be declared teratogenic on the basis of the preliminary study alone.

With a view to an exchange of documents between Western countries and Japan, and until the international guidelines have been unified, an attempt should be made to amalgamate experimental designs to satisfy all regulatory requirements whenever possible (Segment I studies are not compatible in any case). For example, a Segment II study in the rat and the mouse can be made compatible by extending the administration period in the rat to cover days 6 to 17 (in the mouse the treatment periods already coincide) and by adding a sub-group of animals to be brought to parturition so that the progeny can be followed to maturity. In the rabbit the experimental design is already the same. Segment III studies should be extended to include reproduction in the F_1 generation.

A reasonable number of animals must be used, in compliance with regulatory requirements, but the number should be sufficient to provide an adequate amount of data for correct analysis and a clear interpretation of the results, in the context of the type of segment applied.

Since the guidelines provide general directives but few technical details, it is up to each pharmaceutical company to decide upon the quality of the methods and procedures used. It should be borne in mind, however, that if the procedures, observations and records are too few, inaccurate or superficial they may not permit correct appraisal of the adverse effects of a drug on the reproductive process.

It is also the responsibility of the pharmaceutical companies to investigate any particular problems that may arise and clarify equivocal results that might emerge from application of the classical segments. For example, Segment II must be repeated in one of the species already used if uncertain or equivocal results are obtained (e.g. a suspected teratogenic effect) or sometimes to use a more appropriate dose range. On the other hand, a segment may be repeated in an alternative species to verify specific adverse effects (e.g. alterations of prenatal and postnatal development, effects on specific reproductive phases, etc.), especially if the species already used is not judged suitable for study of a particular drug.

Special studies should be conducted to identify the prenatal or postnatal stage responsible for adverse postnatal effects on the progeny (e.g. cross-fostering), or to assess a drug's toxic effects on the neonate (e.g. by direct treatment) when it is not possible to study the effects by transmission in the mother's milk.

In the most difficult cases, the conventional design of the segments is found to be inapplicable and modified designs or supportive studies must be used. There are some drugs which, because of their specific pharmacological action, require more flexible experimental designs or alternative strategies to ascertain their toxicological effects whilst avoiding interference by the pharmacological effect. This is the case, for example, when the study segments must be completed in a species that has already provided the background data in previous studies (e.g. general toxicity, Segment II) and on which the drug exerts its specific activity at doses very close to those foreseen for clinical use. It is very difficult, for example, to apply conventional Segment I (US, EEC) and Segment III (US, EEC, Japan) studies in their unmodified form to drugs which interfere with parturition in the tested species, inducing dystocia, abortion or premature delivery, and thus elevated perinatal mortality of the offspring. It is obvious that in these cases the raw material (progeny) on which to assess postnatal effects is lacking. With all the more reason, a classical Segment I (US, EEC, Japan) cannot be applied at high enough doses in the case of drugs that modify estrus and inhibit implantation, since no conceptus is obtained. Even in these special cases, however, the regulatory authorities are inflexible and require the maximum documentation on which to judge the safety of a drug. In cases where

such a choice can be justified, a classical segment can be performed in an alternative species that meets the metabolic requirements for the drug under study but is less sensitive to the specific pharmacological effect. If there is not sufficient background information on the species, however, the study risks remaining an isolated item of information with validity only as a supportive study, unless the entire package of reproduction studies is repeated in the species in question to satisfy specific regulatory requirements.

If an alternative species cannot be used, the possible modifications to the treatment schemes of the segments involved (generally Segments I and III) must be decided on a case by case basis, in order to avoid the pharmacological effect that masks possible toxic effects or halts certain developmental stages. By withdrawing treatment at critical times in the reproductive process when the drug exerts its pharmacological action, it is often possible to employ doses that provide adequate safety margins with respect to the therapeutic dose and allow the prenatal and postnatal effects on the progeny to be assessed.

These studies are generally accepted by the regulatory authorities of the Western countries not as studies replacing the classical segments but as exploratory studies to support the conventional studies, provided that there are existing studies demonstrating that the conventional schemes cannot be employed. They are not generally taken into consideration by the Japanese authorities, whose main concern is to determine the NOEL (independently of a drug's pharmacological activity) for all types of drug, but only in the conventional experimental design. To meet Japanese regulatory requirements, therefore, it is essential to produce studies performed according to a classical design, using low doses that may even be submultiples of the expected therapeutic dose. In a recent revision of the guidelines [8], however, the Japanese authorities consider the importance of additional studies in the case of Segment II, involving a short duration of treatment if high embryonic/fetal mortalities are observed. This is in view of the possibility that embryonic/fetal losses caused by treatment at an early stage of organogenesis may hinder expression of abnormalities that would have been induced by continued, subsequent treatment, or that abnormalities due to treatment at an early stage of organogenesis may go undiscovered because of embryonic/fetal deaths caused by continued subsequent treatment.

This concept of breaking down the treatment period into sub-periods is very important for Segment II, but should be considered for all study segments when highly toxic substances such as cytotoxic anti-cancer agents are used. These anti-cancer agents should not be included in general guidelines for medicinal products, but should have their own specific guidelines in view of their particular therapeutic modalities (cyclic or intermittent treatment). There are as yet no formal guidelines for anti-cancer agents. EEC authorities tend to require only Segment II studies for cytotoxic agents and US authorities suggest that teratology and fertility/reproduction studies be performed for cytotoxic agents, whilst Japanese authorities require all segments to be applied as for other drugs, with determination of the NOEL. Whether the conventional segments should be applied to these drugs is questionable on two accounts:

1) the high toxicity of these drugs means that experimental species generally do not tolerate long-term continuous treatment (including the 10-13 days required for Segment II) at dosages high enough to guarantee the required safety margins for clinical purposes;

2) human anti-cancer chemotherapy is cyclic and not continuous. Experimental data acquired with continuous treatment may therefore lead to incorrect extrapolation to humans either because the dosages tested are too low or because of the cumulative effects of continuous treatment.

In addition, modern chemotherapy uses combined-drug protocols with various drug combinations and mixtures of different cytotoxic and hormonal drugs, giving rise to possible synergistic and additional effects. The effects of combined therapy may not be comparable to the effects of individual drugs in animals. Drug combinations are not ordinarily tested in animals, but the long-term effects of these protocols should be predicted for children and young people achieving remission who may later wish to have children.

In the case of anti-cancer drugs the design of reproductive toxicity studies should be reconsidered as regards the treatment modalities and should concentrate on the phases that are relevant to therapeutic use. A fertility/reproductive study with cyclic treatment only during the pre-mating period (or also including the mating period) would mimic the human situation more realistically for prediction of long-term effects. Mating of treated animals of one sex with untreated partners would be more useful in identifying the origin of possible genetic damage in later generations. Recovery observations should be considered if necessary. A gestational study with one or more short treatment periods during gestation, including early and late pregnancy, might be more appropriate than the continuous treatment specified by conventional guidelines in that it would mimic possible real-life situations in women (treatment before confirmation of pregnancy and during the second and third trimesters of pregnancy). Short treatment periods would also serve to identify the time of peak sensitivity

to embryolethal and teratogenic effects, making it possible to indicate the stages at which there is a higher or lower risk, as an aid to estimation of the risk/benefit ratio for pregnant women. These treatment schemes would allow the use of higher dosages, and in some instances the cumulative effects of maternal toxicity due to continuous treatment could be avoided. It should also be possible to test simple combination therapies. Studies of the effects on later generations should not be necessary if there are clear indications of teratogenicity in dams killed pre-term after treatment up to and including the period of fetal organogenesis because if the drug is teratogenic in at least one species it is obviously contraindicated in women at least during the first trimester of pregnancy, even as regards risk/benefit estimation. On the other hand, late effects on the offspring should always be studied when treatment is carried out from the end of organogenesis to parturition to discover whether the drug can be used in the last three months of pregnancy.

Studies with treatment during lactation should be considered unnecessary as these drugs are contraindicated in breast-feeding women, and peri- and postnatal effects can be explored in gestational studies which include treatment during the latter period of gestation.

The new endocrine agents for the treatment of cancer also warrant special consideration by the regulatory agencies, not to introduce changes in the tests but to guide regulatory requirements in consideration of the precise therapeutic aim, as already proposed by some scientific committees [11].

In practice these drugs are among the most difficult to investigate according to conventional study designs because they can exert a pharmacological activity at dosages lower than the clinical dose, which means that the possible toxicological effects are masked and adequate safety margins are not guaranteed. Therefore, considering that conventional tests are often applicable only after considerable modification, reproductive toxicity studies are appropriate when therapeutic use is envisaged in individuals of reproductive age (e.g. in women of child-bearing age), but become a waste of time and money when the intended clinical use is in post-menopausal women, for example. Likewise, it is common sense not to require fertility studies in female animals and studies during organogenesis and the peri- and post-natal period in the case of new endocrine agents intended for therapeutic use only in men (e.g. in prostatic cancer).

Thus, although current guidelines can be satisfactorily applied to the majority of drugs, special consideration should be given to developmental toxicity studies with particular drugs that require more flexible guidelines in view of the therapeutic use for which they are intended.

Conclusions

Regulatory guidelines for reproductive toxicity studies in the pre-clinical evaluation of drugs have been in existence for at least 25 years, but differences between the Japanese guidelines and those of Western countries cause considerable problems as far as the mutual acceptability of studies is concerned. Regulatory agencies and scientists are therefore working towards international standardization in this field. It is to be hoped that the regulatory authorities involved in this harmonization process will be prepared to consider that different approaches are needed for different classes of drugs and especially for new drugs.

Practical application of the current guidelines has shown that rigid designs cannot always be applied to all drugs and they can even be considered an unnecessary waste of resources when applied routinely merely to satisfy regulatory requirements, without taking into consideration the specific conditions of therapeutic use. For some drugs, more flexible experimental designs with the possibility of in-depth studies of certain aspects would be much more appropriate than conventional designs to support extrapolation of the experimental data to man.

The need is therefore felt for the international regulatory agencies not only to agree on common general guidelines, but to be unanimously prepared to accept flexible study designs, on a case by case basis, and to introduce special guidelines for particular drugs, such as anti-cancer agents, that do not lend themselves to application of the basic guidelines.

Submitted on invitation.
Accepted on 25 September 1992.

REFERENCES

1. VICHI, F. 1969. Neuroleptic drugs in experimental teratogenesis. In: *Teratology. Proceedings of a Symposium organized by the Italian Society of Experimental Teratology*. Como, Italy, 21-22 October 1967. A. Bertelli & L. Donati (Eds). Excerpta Medica Foundation, Amsterdam. pp. 87-101.

2. U.S. FOOD AND DRUG ADMINISTRATION. 1966. *Guidelines for reproduction studies for safety evaluation of drugs for human use.*

3. ALDER, C.J. & ZBINDEN, G. 1981. Summaries of preclinical guidelines of various countries and international organizations. In: *Preclinical safety requirements in 1980*. C.J. Alder & G. Zbinden (Eds). Institute of Toxicology, Swiss Federal Institute of Technology and University of Zurich, Zurich.

4. WORLD HEALTH ORGANIZATION SCIENTIFIC GROUP. 1967. *Principles for the testing of drugs for teratogenicity*, World Health Organization, Geneva. (Technical Report No. 364).

5. WORLD HEALTH ORGANIZATION. 1975. *Guidelines for evaluation of drugs for use in man*. World Health Organization, Geneva. (Technical Report No. 563).

6. EUROPEAN ECONOMIC COMMUNITY. 1983. Council recommendation of 26 October 1983 concerning tests relating to the placing on the market of proprietary medicinal products. *Official Journal of the European Communities* No. L332 (83/571/EEC).

7. JAPANESE MINISTRY OF HEALTH AND WELFARE. 1984. *Toxicity test guideline*. Notification No. 118, Pharmaceutical Affairs Bureau, Ministry of Health and Welfare, Japan. February 15, 1984, and Notification No. 24, September 11, 1989.

8. JAPANESE MINISTRY OF HEALTH AND WELFARE. 1991. *Guidelines for toxicity studies of drugs manual*. First Edition. Yakuji Nippo (Ltd.), Tokyo.

9. LUMLEY, C.E., PARKINSON, C. & WALKER, S.R. 1991. International approaches to reproductive toxicity testing. *Teratology* **44**: 21A.

10. PALMER, A.K. 1981. Regulatory requirements for reproductive toxicology: theory and practice. In: *Developmental toxicology*. C.A. Kimmel & J. Buelke-Sam (Eds). Raven Press, New York. pp. 259-287.

11. JOINT COMMITTEE, CANCER RESEARCH CAMPAIGN HEADQUARTERS. 1989. Operation manual for control of selection, production, preclinical toxicology and phase I trials of endocrine agents for patients with cancer. *Br. J. Cancer* **60**: 265-269.