

Morphological analysis of rat ovary in normal cycling

CL

Currently formed CL had attained the maximum size (Figs. 6A-C). The luteal cells had foamy eosinophilic cytoplasm (Fig. 6D). Most luteal cells were negative for PCNA (Fig. 6E), and fibrous tissue were sometimes detected in the center of CL. Previously formed CL were vacuolated and fibrous tissue infiltration was advanced (Fig. 6F).

The average numbers of follicles and newly formed CL are shown in Table 2. Whereas the numbers of all follicular types and newly formed CL varied, the counts indicated some tendencies as follows: 1) small and medium follicles were found at constant ratios in all stages; 2) large follicles of Type 8 appeared at diestrus and reached a peak in proestrus; and 3) newly formed CL were found only in the estrus stage.

DISCUSSION

The present study demonstrated that the synchronized combination of morphological findings in large follicles, which are follicle stimulating hormone (FSH)-dependent (Greenwald and Roy, 1994), and currently formed CL is allowed as the ovary reliable clarification of each estrous cycle stage. In particular, classification of CL into 2 types appears very useful for staging. There were no morphological differences in either follicles or CLs between F344 and Donryu rats.

Our results for the estrous cycle are similar to those reported previously except for the detailed description of follicular changes (OECD, 2008; Westwood, 2008). In brief, in **proestrus**, the most characteristic feature is the presence of large follicles of Type 8 (Graafian folli-

cles). In addition, cytoplasmic similarity of granulosa cells to luteal cells is characteristic of the Type 8 follicle. The increase in atretic change of Type 7 follicles indicated withdrawal of this type of follicle on the morning after ovulation. While currently formed CL at proestrus resembled those at diestrus, they appeared to be much more regressive. **Estrus** is the easiest stage to recognize due to newly formed CL. Confirmation of angiogenesis in CL in this stage is important to avoid misdiagnosing them from the edge of the Graafian follicles. A lack of large follicles of Types 7 and 8 is also a feature at estrus. Many large follicles of Type 6 might reflect the start of growing large follicles toward next ovulation (Watanabe *et al.*, 1990; Freeman, 2006). **At metestrus**, currently formed CL typically had luteal cells with basophilia and prominent nucleoli. This morphology might reflect progesterone production in rats classified into an ultrashort CL species (Freeman, 2006; Stouffer, 2006). **At diestrus**, currently formed CL were still characteristic. They reached the maximum size and contained luteal cells with eosinophilic and foamy cytoplasm. CL are not able to produce active progesterone at this stage (Watanabe *et al.*, 1990; Freeman, 2006).

CLs with cavities were encountered at metestrus in the present study. Their interpretation differs between OECD (2008) and Westwood (2008). Whereas OECD diagnoses them as ovarian luteal cysts which failed ovulation but undergoing luteinization of Graafian follicles in young adult rats (2008), Westwood described them as CL still containing occasional central fluid-filled cavities (2008). In domestic animals, CL with cavities have been accepted not to be confused with cystic follicles or luteinized cysts (Kennedy and Miller, 1993). On the other hand, the distinc-

Table 2. Follicular and newly formed CL counts per unilateral ovary (a) in F344 rats

Follicle/CL	Estrous cycle			
	Proestrus	Estrus	Metestrus	Diestrus
No. of rats examined	13	11	12	12
<i>Follicles</i>				
Small	15.9 ± 2.5 ^(b)	15.8 ± 3.3	17.0 ± 2.0	20.4 ± 4.4
Medium	2.2 ± 0.5	2.2 ± 0.5	2.0 ± 0.5	2.7 ± 0.4
Large ^(c)	16.7 ± 1.7	20.9 ± 0.1	18.0 ± 1.4	18.8 ± 2.2
Type 8	2.5 ± 0.3	0	0	0.8 ± 0.3
Atresia	10.8 ± 1.1	8.5 ± 1.4	11.9 ± 1.9	14.5 ± 1.6
<i>CL</i>				
Newly formed	0	1.6 ± 0.4	0	0

(a), the transverse section dissected to obtained maximum cut surface

(b), mean ± S.D.

(c), Including Type 8 follicles

tion between CL with cavity and luteinized cyst is sometimes difficult in rodents, multiovulated species. There are various description of similar morphology showing luteal tissue with cavity, such as CL containing cavities (Westwood, 2008), luteal cyst (OECD, 2008) or cystic CL (Yuan and Foley, 2002). In general, luteinized cysts are easy to distinguish from cystic follicles, which are very large atretic follicles and reflect impair of large follicle growth. They are caused by many chemicals and hormonal disturbances (Bogovich, 1991; Røste *et al.*, 2001; Baravalle *et al.*, 2006). In typical luteinized cyst arising from failure of ovulation and undergoing luteinization, the luteinized cells with angiogenesis formation are found in a thin cell layer or uneven distributed in crescent area in cows (Kenndey and Miller, 1993) and rats (Tamura *et al.*, 2009). The rupture of follicles at ovulation is mainly controlled by progesterone and prostaglandin, and disruptions of these hormones induce unruptured follicles with or without luteinization or formation of luteinized cysts (Davis *et al.*, 1999; Gaytán *et al.*, 2003; Shirota *et al.*, 1998; Tamura *et al.*, 2009; Tsubota *et al.*, 2009; Yuan and Foley, 2002). The CL containing small cavities seems to be normal range in normal cycling rats, because the cavity will be replaced to fibrous tissues next day, at diestrus. However, the toxicological significance of CL with cavities remains to be fully undetermined. We should pay attention an increase of CL containing cavities in treated groups. Further investigation or information is required. Sequential vaginal cytology might be very informative to check ovulation.

Detection of damage to small follicles is very important for detection of ovarian toxicity in preclinical studies. Small follicle counts using single ovarian sections are

challenging and evaluation may be misleading (Meredith *et al.*, 1999; Bolon *et al.*, 1997). However, FSH-independent small and medium follicles (Greenwald and Roy, 1994) were observed in all sections examined in the present study and at a constant rate with the estrous cycle stage. The result indicates that lack or only a few visible small follicles in a single ovarian section might give a pointer to alarm for small follicular damage. In addition, PCNA immunohistochemical staining and microscopic observation of bilateral ovaries are helpful to increase the reliability of detection. The STP Ovarian evaluation workshop group (Regan *et al.*, 2005) recommended a qualitative morphological approach for ovarian sections by toxicologic pathologists familiar with normal reproductive cycle as the first tier approach for detection of ovarian toxicity.

In conclusion, the present results demonstrate that ovarian morphology are allowed each estrous cycle stage by combination of morphologic changes in large follicles and CL undergoing development and/or regression (Fig. 7). In the classification of CL into 2 categories, morphological changes in currently formed CL appear particularly useful for staging. In addition, a single ovarian section might be sufficient for analyzing each estrous stage. Our results indicate that any morphological deviation in follicles and CL from the synchronized combination might be the first indicator of ovarian toxicity. PCNA immunohistochemical staining is useful to detect small follicles. Vaginal cytology is also informative for detection of endocrinological condition.

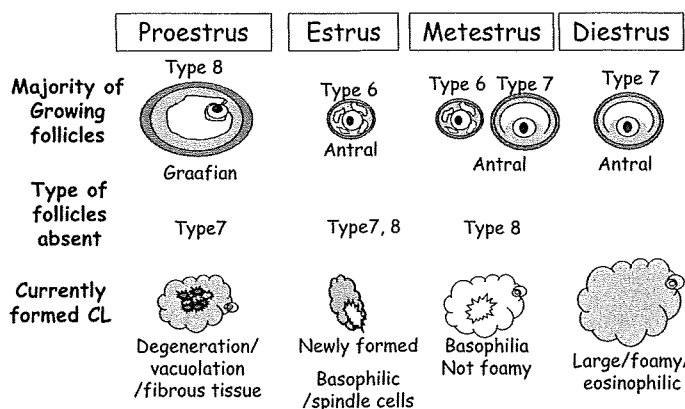


Fig. 7. A scheme for combination of morphologic features in follicles and CL at each estrous cycle in rats. The classification of follicles was referred to Pedersen and Peters (1968).

Morphological analysis of rat ovary in normal cycling

REFERENCES

- Baravalle, C., Salvetti, N.R., Mira, G.A., Pezzone, N. and Ortega, H.H. (2006): Microscopic characterization of follicular structures in letrozole-induced polycystic ovarian syndrome in the rat. *Arch. Med. Res.*, **37**, 830-839.
- Bogovich, K. (1991): Induction of ovarian follicular cysts in the pregnant rat by human chorionic gonadotropin. *Biol. Reprod.*, **45**, 34-42.
- Bolon, B., Bucci, T.J., Warbritton, A.R., Chen, J.J., Mattison, D.R. and Heindel, J.J. (1997): Differential follicle counts as a screen for chemically induced ovarian toxicity in mice: results from continuous breeding bioassays. *Fundam. Appl. Toxicol.*, **39**, 1-10.
- Davis, B.J., Lennard, D.E., Lee, C.A., Tiano, H.F., Morham, S.G., Wetsel, W.C. and Langenbach, R. (1999): Anovulation in cyclooxygenase-2-deficient mice is restored by prostaglandin E2 and interleukin-1beta. *Endocrinology*, **140**, 2685-95.
- Freeman, M.E. (2006): Neuroendocrine control of the ovarian cycle of the rat, In *The Physiology of Reproduction*, Third Edition (Knobil, E., Neill, J.D. eds.) pp.2327-2388, Raven Press Ltd., New York.
- Gaytán, F., Bellido, C., Gaytán, M., Morales, C. and Sánchez-Criado, J.E. (2003): Differential effects of RU486 and indomethacin on follicle rupture during the ovulatory process in the rat. *Biol. Reprod.*, **69**, 99-105.
- Greenwald, G.S. and Roy, S.K. (1994): Follicular development and its control. In *The Physiology of Reproduction*, Second Edition (Knobil, E., Neill, J.D. eds.) pp.629-724, Raven Press Ltd., New York.
- Kennedy, P.C. and Miller, R.B. (1993): The female genital system. In *Pathology of Domestic Animals*. Fourth Edition. (Jubb, K. V. F., Kennedy, P. C., Palmer, N. eds.) pp.349-470, Academic Press, San Diego.
- Meredith, S., Dudenhoefter, G. and Jackson, K. (1999): Single-section counting error when distinguishing between primordial and early primary follicles in sections of rat ovary of different thickness. *J. Reprod. Fertil.*, **117**, 339-343.
- Muskhelishvili, L., Wingard, S.K. and Latendresse, J.R. (2005): Proliferating cell nuclear antigen--A marker for ovarian follicle counts. *Toxicol. Pathol.*, **33**, 365-368.
- Organization of Economic Cooperation and Development (2001): Proposal for updating Guideline 416: two generation reproduction toxicity study. In *OECD Guideline for Testing of Chemicals*, pp. 1-13. OECD, Paris.
- Organization of Economic Cooperation and Development (OECD) (2008): Draft Guidance and Review Documents/Monographs, Guidance Document for Histopathologic Evaluation of Endocrine and Reproductive Tests. Part 3: Female reproductive system. http://www.oecd.org/document/12/0,3343,en_2649_34377_1898188_1_1_1_1,00.html
- Pedersen, T. and Peters, H. (1968): Proposal for a classification of oocytes and follicles in the mouse ovary. *J. Reprod. Fertil.*, **17**, 555-557.
- Picut, C.A., Swanson, C.L., Scully, K.L., Roseman, V.C., Parder, R.F. and Remick, A.K. (2008): Ovarian follicle counts using proliferating cell nuclear antigen (PCNA) and semi-automated image analysis in rats. *Toxicol. Pathol.*, **36**, 674-679.
- Regan, K.S., Cline, J.M., Creasy, D., Davis, B., Foley, G.L., Lanning, L., Latendresse, J.R., Makris, S., Morton, D., Rehm, S., Stebbins, K. and STP Ovary Evaluation Working Group (2005): STP Position Paper: Ovarian follicular counting in the assessment of rodent reproductive toxicity. *Toxicol. Pathol.*, **33**, 409-412.
- Røste, L.S., Taubøll, E., Berner, A., Isojärvi, J.I.T. and Gjerstad, L. (2001): Valproate, but not lamotrigine, induces ovarian morphological changes in Wistar rats. *Exp. Toxicol. Pathol.*, **52**, 545-552.
- Sanbuissho, A., Yoshida, M., Hisada, S., Sagami, F., Kudo, S., Kumazawa, T., Ube, M., Komatsu, S. and Ohno, Y. (2009): Collaborative work on evaluation of ovarian toxicity by repeated-dose and fertility studies in female rats. *J. Toxicol. Sci.*, **34** (Special Issue I), SP1-SP22.
- Shirota, M., Watanabe, G., Taya, K. and Sasamoto, S. (1998): Effects of indomethacin on the selective release of follicle-stimulating hormone during the period of ovulation in the rat. *J. Vet. Med. Sci.*, **60**, 1059-65.
- Smith, B.J., Plowchalk, D.R., Sipes, I.G. and Mattinson, D.R. (1991): Comparison of random and serial sections in assessment of ovarian toxicity. *Reprod. Toxicol.*, **5**, 379-383.
- Stouffer, R.L. (2006): Structure, function, and regulation of the corpus luteum. In *The Physiology of Reproduction*, Third Edition (Knobil, E., Neill, J.D. eds.), pp.475-526. Raven Press Ltd., New York.
- Tamura, T., Yokoi, R., Okuhara, Y., Harada, C., Terashima, Y., Hayashi, M., Nagasawa, T., Onozato, T., Kobayashi, K., Kuroda, J. and Kusama, H. (2009): Collaborative work to evaluate ovarian toxicity 2) Two- or four-week repeated dose studies and fertility study of mifepristone in female rats. *J. Toxicol. Sci.*, **34** (Special Issue I), SP31-SP42.
- The ICH Steering Committee to the regulatory authorities of the three ICH regions (the European Union, Japan and USA) (2008): Guidance on nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals. M3(R2). [http://www.pmda.go.jp/ich/m/Step3_m3\(r2\)_08_07_14_e.pdf](http://www.pmda.go.jp/ich/m/Step3_m3(r2)_08_07_14_e.pdf)
- Tsubota, K., Kushima, K., Yamauchi, K., Matsuo, S., Saegusa, T., Ito, S., Fujiwara, M., Matsumoto, M., Nakatsuji, S., Seki, J. and Oishi, Y. (2009): Collaborative work on evaluation of ovarian toxicity 12) Effects of 2- or 4-week repeated dose studies and fertility study of indomethacin in female rats. *J. Toxicol. Sci.*, **34** (Special Issue I), SP129-SP136.
- U.S. Environmental Protection Agency (1998): Health Effects Test Guidelines, OPPTS 870.3800, Reproduction and Fertility Effects. EPA 712-C-98-239. U.S. Environmental Protection Agency, Office of Prevention, Pesticides and Toxic Substances, Washington, DC.
- U.S. Food and Drug Administration (2000): Redbook 2000. Toxicological Principles for the Safety Assessment for Food Ingredients. IV.C.9.a. Guidelines for Reproductive Studies. U.S. Food and Drug Administration, Center for Food Safety and Applied Nutrition, Washington, DC.
- Watanabe, G., Taya, K. and Sasamoto, S. (1990): Dynamics of ovarian inhibin secretion during the oestrous cycle of the rat. *J. Endocrinol.*, **126**, 151-157.
- Westwood, F.R. (2008): The female rat reproductive cycle: A practical histological guide to staging. *Toxicol. Pathol.*, **36**, 375-384.
- Yuan, Y.D. and Foley, G.L. (2002): Female Reproductive System. In *Handbook of Toxicologic Pathology*. Second Edition (Haschek, W.M., Rousseaux, C.G., Walling, M.A. eds.), pp.847-894. Academic Press, San-Diego.

●特集／マイクロドーズ臨床試験

2. マイクロドーズ臨床試験に必要な 非臨床試験データ

国立医薬品食品衛生研究所

大野 泰雄

Non-clinical Data Needed for the Conduct of Microdose Clinical Study

Yasuo OHNO

National Institute of Health Sciences
1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan

Microdose clinical study was introduced to Japan in June 2008 for the purpose of promoting new drug development. However, the upper dose of the study is below 100 μg and the application of findings to drug development is limited. On the other hand, the ICH-M3 guidance "Non-clinical safety studies for the conduct of human clinical trials for pharmaceuticals" was revised in June 2009, which introduces other approaches to exploratory clinical studies. The guidance describes five approaches (two microdose approaches, one sub-therapeutic single dose approach, and two up to therapeutic dose range repeated-dose approaches). Because the highest doses of these studies are limited to within the anticipated non-toxic dose ranges, requirement of toxicity tests is decreased compared to the classic phase I study. The rationales for the reduction of toxicity tests are explained.

Key words : microdose clinical study, safety test, ICH, exploratory clinical study

1. 序

早期探索的臨床試験は医薬品開発の初期段階で開発候補物質の絞り込みを行うための臨床試験であり、その目的と投与量に基づき、概略1) 極めて低用量を用いて薬物動態を検討するマイクロドーズ臨床試験 (MD試験) (I型) と、2) MD試験よりは高いが、薬効用量よりは低い用量で薬物動態と薬効につながる作用を評価する準薬効用量早期探索的臨床試験 (II型)、および3) 薬効は現れるが毒性は現れないと思われる用量を用い、薬物動態や薬物相互作用、および薬効を評価する薬効用量早期探索的臨床試験 (III型) に分けられる。これらを単回あるいは反復投与の回数に応じて分けることもある。一方、ICH (International Conference on Harmonisation of the Technical Requirements for Registration of Pharmaceuticals for Human Use) では段

階的に臨床投与量を増加させることから、Table 1で示したように早期探索的臨床試験を5つのアプローチに分類した。いずれにせよ、従来の第I相試験は推定薬効用量を超え、被験者が耐えられる最高用量 (最大耐量: MTD) まで投与することが予定されているが、早期探索的臨床試験では、試験の目的を狭め、投与量を毒性の現れないと推定される用量や投与回数に限定することにより、試験実施に必要な非臨床試験の範囲を狭め、結果として、非臨床試験および臨床試験も含め、One Batchの合成でまかなえる程度までに、少量の被験物質での評価を可能とするとともに、臨床試験で要求されるGMP (Good Manufacturing Practice) 上の負荷を軽くし、臨床試験に入りやすくなるように企図したものである。

本稿ではMD試験についての厚生労働省の通知(2008)¹⁾と2009年6月にICHでStep 4の合意²⁾が達

Table 1 早期探索的臨床試験についてのICH M3 (R2) (Step 4)

臨床試験の型	内容と用量範囲	実施に必要な非臨床試験
マイクロドーズ臨床試験 (アプローチ1)	<ol style="list-style-type: none"> 1) 総投与量を100 µg以下とし、被験者1人あたり5回まで分割して投与するもの。 2) 総投与量は薬効量の1/100以下、無毒性量の1/100以下。 	<ol style="list-style-type: none"> 1) 1つの動物種(通常げっ歯類)における予定臨床投与経路または静脈内投与を用いた拡張型単回投与毒性試験(TK評価を含む)。最高用量は1000倍の暴露までで良い。 2) インビトロ標的/受容体結合プロファイルの解析。 3) 薬力学的に外挿性のあるモデルにおける薬効薬理作用についての適切な評価に基づいて、臨床投与量を設定。 4) 遺伝毒性試験は不要。構造活性相関情報があれば提出する。 5) 高放射活性物質(例えば、PETイメージング剤)の場合には、適切な薬物動態学的情報と放射線暴露量の推定が必要。
マイクロドーズ臨床試験 (アプローチ2)	<ol style="list-style-type: none"> 1) 1回あたりの最高用量が100 µg。投与は休業期間(実際もしくは予想される半減期の6倍以上)を設けて、最大5回まで(被験者あたり総投与量500 µg以下)。 2) 各投与量は薬効量の1/100以下、無毒性量の1/100以下。 	<ol style="list-style-type: none"> 1) 1つの動物種(通常、げっ歯類)における臨床投与経路を用いた7日間反復投与毒性試験(TK評価を含む)。最高用量は1000倍の暴露までで良い。他は、アプローチ1の2)-5)と同じ。
準薬効用量又は推定薬効域での単回投与試験 (アプローチ3)	<ol style="list-style-type: none"> 1) 初回用量は、最も感受性の高い動物種における毒性所見のタイプや薬効用量を考慮して選択する。ヒトでの初回用量に関して考慮すべきその他の事項については、各極のガイダンスを参考にすべき。 2) 動物での毒性がヒトにおいてモニタリングが可能であり、可逆的なものと予想される場合には、最高用量は、より感受性の高い種におけるNOAELでの暴露の1/2までが許容される。 	<ol style="list-style-type: none"> 1) げっ歯類及び非げっ歯類を用いた拡張型単回投与毒性試験(TK評価を含む)。投与経路は予定臨床経路。血液学、血液生化学、剖検及び組織病理学データが含まれること。最高用量は、MTD、MFD、又は限界量。 2) 遺伝毒性試験(Ames試験) 3) 薬理作用については、アプローチ1の2)、3)、および安全性薬理試験コアバッテリー。
反復投与臨床試験 (アプローチ4) 薬効域であるが、臨床最大耐量の評価を目的としない14日までの投与	<ol style="list-style-type: none"> 1) 両動物種で毒性がみられた場合、初回用量は適切な各極のガイダンスに従う。 2) 毒性がいずれの動物種でもみられなかった場合や毒性が1種の動物種でしかみられなかった場合、推定暴露量が、より低い暴露を示した動物種のNOAELにおけるAUCの約1/50になるように、初回用量を設定する。ヒトの初回用量に関して、例えば、薬力学活性の推定などの、考慮すべきその他の事項については、各極のガイダンスを参考にする。 3) 両動物種で毒性がみられない場合、臨床最高用量は、動物試験での最高用量におけるいずれかの種での低い方の暴露(AUC)の1/10を超えない用量。 4) 1種においてのみ毒性が示されている場合、最高臨床用量は、毒性を示した動物種から得られたNOAELでのAUC又は毒性を示さなかった動物種における最高用量でのAUCの1/2のいずれか低い方を超えない。 5) 両動物種で毒性がみられる場合、最高臨床用量は、標準的なリスク評価方法に基づく。この特殊な場合には臨床最大耐量が探索できる。 	<ol style="list-style-type: none"> 1) げっ歯類と非げっ歯類における標準的な2週間反復投与毒性試験(TK評価を含む)。用量設定は、最高用量にて予想される臨床AUCの倍数の暴露を基にする。 2) 遺伝毒性試験(Ames試験及びインビトロ又はインビボの染色体異常誘発能試験)。 3) 薬理作用については、アプローチ1の2)、3)、および安全性薬理試験コアバッテリー。
反復投与臨床試験 (アプローチ5) 薬効域であるが、臨床最大耐量の評価を目的とせず、非げっ歯類の投与期間を超えない14日までの投与。	<ol style="list-style-type: none"> 1) 初回用量は、体表面積換算で、より感受性の高い動物種における無毒性量の1/50以下。ヒトの初回用量に関して、例えば、薬力学活性の推定などの、考慮すべきその他の事項については、各極のガイダンスを参考にする。 2) ヒトにおける最大暴露は、非げっ歯類の無毒性量でのAUC又はげっ歯類での無毒性量でのAUCの1/2のいずれか低い方を超えない。 	<ol style="list-style-type: none"> 1) げっ歯類を用いた標準的な2週間反復投与毒性試験(げっ歯類が適切な種であることの理由が必要)。最高用量はMTD、MFD、又は限界量とする。 2) 非げっ歯類(n=3)における確認試験で、げっ歯類でのNOAELにおける暴露を得られると推定される用量を最短3日間かつ少なくとも予定臨床試験期間を投与する。もしくは、非げっ歯類における漸増投与試験で、げっ歯類での無毒性量における暴露を得られると推定される用量を最短3日間かつ少なくとも予定臨床試験期間を投与する。 3) 遺伝毒性試験(Ames試験及びインビトロ又はインビボの染色体異常誘発能試験)。インビボ試験を実施する場合には、げっ歯類の毒性試験の一部として組み込むことが可能な場合もある。 4) 薬理作用については、アプローチ1の2)、3)、および安全性薬理試験コアバッテリー。

成された「Guidance on Nonclinical Safety Studies for the Conduct of Human Clinical Trials and marketing authorization for Pharmaceuticals M3 (R2) : ICH M3 (R2) と略す」に基づき MD 試験に必要な非臨床試験について述べる。これ以外の早期探索的臨床試験については ICH M3 (R2) のガイダンス²⁾を参照されたい。

2. マイクロドーズ臨床試験 (MD 試験) について

1) わが国で通知された MD 試験について

平成 20 年 6 月 3 日に厚生労働省から通知された「マイクロドーズ臨床試験の実施に関するガイダンス (MD 試験指針)」は EMEA (European Medicines Agency)³⁾ および FDA (米国の Food and Drug Administration)⁴⁾ の関連する指針をベースに策定されたものである。指針では MD 試験を「ヒトにおいて薬理作用を発現すると推定される投与量 (以下「薬効発現量」という。) の 1/100 を超えない用量又は 100 μg のいずれか少ない用量の被験物質を、健康な被験者に単回投与することにより行われる臨床試験をいう」と定義し、その目的を「被験物質のヒトにおける薬物動態に関する情報を医薬品の臨床開発の初期段階に得ることである。具体的には、被験物質の吸収や血中動態、排泄特性、ヒトにおける代謝物プロファイル等を明らかにすること、分子イメージング技術を用いて被験物質の体内における局在に関する情報を得ること等である」とした。MD 試験を医薬品開発の初期に実施し、薬物動態学的に適正と思われるものを選択しておくことにより、その後の臨床開発において、体内動態が原因で開発が失敗する可能性を削減できるものと期待されている。この試験では許容される投与量が極めて低いことから Accelerator Mass Spectrometry (AMS : 加速器質量分析法) や高感度 LC/MS/MS (Liquid Chromatograph/Mass Spectrometry/Mass Spectrometry) などの高感度分析法を使用する必要がある。また、Positron Emission Tomography (PET : 陽電子放射断層撮影法) 等のイメージング技術と組み合わせることにより、ヒト体内分布を明らかにし、ヒト組織を用いた *in vitro* 試験結果と組み合わせることにより、標的部位での薬力学的作用や意図せぬ部位での作用の推定にも利用できる。

MD 試験における被験者の安全性確保のために必要な非臨床試験は、一般に、下に述べる (1) 拡張型単回投与毒性試験 (1 種動物)、(2) 局所毒性試験、および (3) 薬理作用に関する試験、および放射性同位元素を用いる場合は暴露レベルの安全性評価に限定される。

これは、MD 試験が毒性も薬効も認められないと思われる極めて低用量で行われる試験であることと矛盾しているように見えるが、被験者に思いがけない毒性が現れることを未然に防止するために行うものである。

一方、ICH M3 (R2) では、PET への応用を考慮し、反復投与できる MD 試験を導入した。それには、総投与量が 100 μg 以下のアプローチ 1 と 1 回あたりの用量が 100 μg 以下で 5 回までの反復投与で総投与量が 500 μg 以下のアプローチ 2 に分類され、それぞれについて、臨床投与量の制約と必要な非臨床試験の範囲を定めた。

2) ICH M3(R2)に記載された MD 試験 アプローチ 1

これは、前項と同様に、毒性も薬効も現れないと推定される用量で、被験物質の体内動態を検討するための試験である。反復投与が認められているが、総投与量は 100 μg 以下、かつ、拡張型単回投与毒性試験での無毒性量 (NOAEL) の 1/100 以下かつ薬効発現量の 1/100 以下に制限されている。実施に必要な非臨床試験としては、1 動物種 (通常、げっ歯類) での予定臨床経路あるいは静脈内投与での拡張型単回投与毒性試験でトキシコキネティクス (TK : 通常、血中濃度推移で評価される被験物質による全身暴露) 評価を含む試験が必要である。最高投与量は臨床投与量の 1,000 倍 (静脈内投与では体重換算、経口投与では体表面積換算) まででよい。また、薬理学的な特性を明らかにするために、*in vitro* 標的/受容体結合プロファイルの解析が必要である。さらに、薬力学的に外挿性のあるモデルにおける薬効薬理作用 (作用機序や効力) についての適切な評価に基づいて、臨床投与量を設定すべきとされている。なお、遺伝毒性試験は実施しなくてもよいが、何らかの試験あるいは構造活性相関の情報があれば治験申請時に提出すること。高放射活性物質 (たとえば、PET イメージング剤) の場合には、適切な薬物動態学的情報と放射線暴露量の推定が必要とされている。

3) ICH M3(R2)に記載された MD 試験 アプローチ 2

この場合も、毒性も薬効も現れないと推定される用量での臨床試験で、1 回あたりの臨床投与量は 100 μg 以下で、無毒性量の 1/100 以下および薬効量の 1/100 以下とされ、総投与量は 500 μg 以下に制約されている。なお、投与は休薬期間 (実際もしくは予想される半減期の 6 倍以上) を設けて、最大 5 回までとされている。一般毒性試験としては、1 種 (通常、げっ歯類) における 7 日間の反復投与毒性試験 (TK 評価を含む) を予定臨床経路あるいは静脈内投与で実施し、血液学ならびに血液生化学検査と剖検および組織病理学検査を実施する必要がある。最高投与量の設定、遺伝毒性

試験の実施, ならびに高放射活性物質への対応はアプローチ 1 の場合と同様である。

3. 拡張型単回投与毒性試験

これは従来の単回投与毒性試験に観察時期と観察項目を追加したものであり, EMEA³⁾, FDA⁴⁾, および日本の MD 試験指針¹⁾に記載されており, それらはいずれも同じ内容である。すなわち, 拡張型単回投与毒性試験においては, 投与後の観察期間は 2 週間とし, 毒性徴候の種類, 程度, 発現時期, 推移および可逆性について, 用量および時間との関連で観察し記録する。また, 適切な時期 (通常, 投与翌日および 2 週間の観察期間終了時) に血液検査, 血液生化学検査および病理組織学的検査を行う。なお, 病理組織学的検査については, 高用量群に組織学的変化がなければ, 対照群および高用量群のみ行うことで差し支えない。なお, 日本の MD 試験指針では, 意味もなく多数の臓器について, 病理組織学的検査を行うことは求められておらず, 被験物質の特性や検査結果に基づいて, 適切な臓器・組織を選択し, 実施することでよいとされている⁵⁾。なお, 投与翌日に観察するのは急性的影響が 2 週間後には回復し, 検出できなくなる可能性を考慮したためである。投与 2 週間後の観察は, 遅延毒性の有無や回復性を評価するためである。

動物は 1 種類の哺乳類の雌雄を用い, 対照群を設けたうえで実施する。当該試験の実施により, 当該被験物質の実験動物に対する最大無毒性量 (NOAEL) および最小毒性発現量を確認するか, またはマイクロドーズ臨床試験における当該被験物質の投与量に関する適切な安全域 (margin of safety; 通常, 日本の MD 試験指針では体表面積換算で 100 倍以上) を確立する。

投与経路については, MD 試験指針では当該被験物質の MD 試験における予定投与経路とされているが, ICH M3 (R2) 指針では静脈内投与も認めている。なお, 当該被験物質の MD 試験における予定用量に対し, 体表面積換算でその 1,000 倍量を用いても毒性が認められない場合, 当該 1,000 倍量を拡張型単回投与毒性試験の上限用量としても差し支えないとされている。

4. ヒトでの薬効用量推定方法

ヒト薬効用量推定方法については, MD 試験指針に以下の 2 つの方法が記載されている。

1) 経験的な方法

動物における薬効発現量をもとに体表面積換算する

ことにより, ヒトでの薬効発現量を推定するもので, FDA の初回投与量設定法のガイダンス⁶⁾や探索的 IND 研究に関するガイダンス⁴⁾においても採用されている。また, EMEA の MD 試験に関するポジションペーパー³⁾でも採用されている。しかし, MD 試験指針では, 「この予測方法はあくまでも経験則であり, 精度の高い予測法とは言い難い。動物を用いた in vivo データ及びヒト組織や細胞を用いた in vitro データ等を総合的に評価して有効血中濃度が予測可能であれば, より精度の高い方法」として, 以下の方法が推奨されている。

2) 薬物動態学的情報を用いる方法

この方法は, 最大血中濃度 (C_{max}) または血中濃度時間曲線下面積 (AUC) を基準に推定するものであり, どちらを採用するかについては, 薬効発現の作用機序等により異なる。 C_{max} を基準に推測する方法では, 「まず, 適切な動物での薬効発現量における最大血中濃度 (C_{max}) を求める。血漿タンパク結合について, 動物とヒトとの種差を補正し, ヒトでの薬効発現量における C_{max} (ヒト推定 C_{max}) を推定する。この方法は, 動物もヒトも血漿タンパクと結合していない遊離型で薬効を発現すると仮定している。なお, 遊離型, 結合同型それぞれの作用発現への寄与が不明の場合には, タンパク結合を補正した場合としない場合の両者を計算し, より低い方を採用する。更に, 動物における分布容積及び動物及びヒトにおける血漿タンパク結合情報をもとにヒトにおける分布容積 (V_d) を推定する。最後に, ヒト推定 C_{max} と V_d の積から, ヒトでの薬効発現量を計算する (MD 試験指針)」。AUC を薬効の指標として用いる場合も, 同様に行う。

5. 早期探索的臨床試験に必要な非臨床試験について

1) MD 試験に必要な非臨床試験の範囲

Table 1 に示したように, MD 試験に必要な一般毒性試験は拡張型単回投与試験のみでよいとされている。これは, 以下に述べる理由による。

著者の調査によれば, 低分子量化学物質で致死量が MD 試験の用量の約 100 倍である 0.2 mg/kg 以下の物質は少ない⁷⁾。とくに, 2 μ g/kg 以下の物は TCDD (2, 3, 7, 8-tetrachlorodibenzo-p-dioxin) および botulinum toxin のみであった。2-20 μ g/kg は abrin toxin と saxitoxin, tetrodotoxin のみであった。20-200 μ g/kg は amanitin, dinophysistoxin, okadaic acid, methylphenidate, digoxin, および digitoxin のみであった。これらによる毒性の多くは, 詳細な症状観察を含む, 通常

Table 2 経口投与での致死量の種差 (mg/kg)

	ヒト LDLo	マウス LD ₅₀	ラット LD ₅₀	イヌ LD ₅₀	動物/ヒト
Botulinum toxin A*	2.14	81.4	96		44.86
Abrin A toxin	0.007	6638			948286
Digitoxin	0.071		56	0.5	789
Cantharidin	0.28	1		60	214
Colchicine	0.4	5.886	2.5		15
Methadone HCl	0.5	70	95		190
4-Aminopyridine	0.59	658	1050	3700	6271
Alphaprodine HCl	1.4	68	90		64
Arcenic (III) oxide	1.429		40		28
Fluoroacetamide	2		5.75		2.9
Sodium metaarsenite	2		41		21
Thallium (I) sulfate	2.166	35	16	16	16
Emetine	2.941	12	12		4.1
Warfarin	6.667	323	1.6		48
Barium carbonate	11		418		38
Aconitin	28	1			0.04
Chlordane	29		283		9.8

* : Unit/kg

の単回投与毒性試験あるいは安全性薬理試験で検出できる。しかし、微生物等に由来するトキシン類等のなかには、毒性が2 µg/kg以下の用量で現れる物も多く、また、一般の症状観察では検出しがたい肝・腎・内分泌器官等にも毒性が現れることがある。このような状況に対応するためには、血液学ならびに血液生化学的検査や尿中酵素測定、さらに病理学的検査も必要である。

なお、単回投与の早期探索的臨床試験は拡張型単回投与毒性試験に基づいて行うことが許容されるが、この場合は、GLP (Good Laboratory Practice) に基づく信頼性の高いデータが必要である。なお、致死性を評価目標とすべきではない。また、バイオテクノロジーで作られた薬品については、未知のところが多いことから、MD試験指針では、「生物由来製品又は体内で如何なる受容体に関与するか十分な知見が得られていないものなど従来の医薬品とは全く異なる作用機序による薬理作用を期待した化合物を投与する場合については、個別にその安全性等についての考察が必要であり、本ガイダンスをそのまま適用することはできない」とされた。ICH M3 (R2) 指針でも「バイオテクノロジー応用医薬品については、本ガイダンスは、臨床開発に

関連する非臨床試験の実施時期に関してのみの指針となる」とされた。

2) ヒト暴露レベルの1,000倍までの用量が必要な理由

今までに蓄積された化学物質等の毒性データによれば、MD試験で使用される100 µgという用量で重篤な毒性を現す低分子化合物はほとんどないこと⁷⁾、また、あったとしてもそれらの毒性は単回投与毒性試験で検出可能なものである。一方、経口投与の単回投与試験での致死量の種差は100倍を超えることがあるが (Table 2)、多くの場合、薬理活性の1/100という用量の設定で重篤な副作用は回避できると思われる。しかし、ヒトと動物との間に1,000倍を超える種差があった fialuridine による死亡事故⁸⁾やヒトとカニクイザルとの間にリンパ球の反応性に大きな差があった CD28 に対するアゴニスト抗体 TGN1412 により重篤な副作用を発現した事故⁹⁾などもあり、必ずしも非臨床試験を臨床用量の1,000倍を超える高用量まで実施したからといって予測できない場合もある。機械的に毒性試験を実施するだけでなく、被験物質の化学構造や薬理作用の特徴を十分把握し、必要に応じて静脈内投与試験の実施やヒト由来試料を用いた検討、薬理作用の延

長としての副作用も含め、毒性学的学識を基にした考察で判断する必要がある。

しかし、ヒトでの毒性を事前に予測し、対応準備しておくためには、なるべく、現れる毒性徴候を把握しておくことが望ましい。一方、単回投与毒性試験の結果は動物種のみならず、動物の系統、実験施設等により大きくばらつくことが知られている。これらのことから、ヒトに初めて投与する前の拡張型単回投与毒性試験の最高用量はヒト試験での暴露量の1,000倍まで投与することが求められたものである。この際、経口投与での暴露量はおもに体表面積換算で換算されるが、静脈内投与の場合は体重換算で計算される。

なお、静脈内投与試験の結果は、経口投与の結果と比べ、ヒトへの予測性はより高いと考えられることから、ICH M3 (R2) 指針では静脈内投与試験で行うことも考慮された。

Fialuridine (FIAU) の経験

NIH が慢性 B 型肝炎治療薬として開発したもので、実験動物での試験では肝、膵、骨格筋、神経への毒性は認められなかったことから、1993年3月から、0.10-0.25 mg/kg/day 24 週間投与、1-2 週ごとに観察とのプロトコルで6カ月臨床試験を15名の患者で開始したところ、13週で lactic acidosis (LA) と肝不全が突然発現し、7名は進行性 LA、黄疸悪化、肝機能悪化を呈し、3名も中程度の肝毒性を起こした。膵臓炎、神経障害、筋障害も発生し、投与停止した。同年6-8月の間に、肝臓、膵臓障害で5名が死亡し、2名は肝移植で救命した⁹⁾。

組織検査では肝の脂肪蓄積が認められたが、肝細胞の壊死や構造変化はわずかであった。電子顕微鏡検査で肝細胞のミトコンドリア異常と脂肪蓄積が認められた¹⁰⁾。一方、Richardson ら (1994, 1999)^{11,12)}によれば、企業データより、経口投与での最大耐量 (MTD) はラット 70 日間投与で 510 mg/kg/日、イヌ 90 日間投与で 3 mg/kg/日、サル 30 日間投与で 25 mg/kg/日と推定している。また、ラットに 500 mg/kg/day を 90 日間、ウッドチャックに 7.5 mg/kg/day を 90 日間投与しても弱い変化のみしか認められなかった。すなわち、反復投与での致死量にヒトとラットとの比較では 2,000 倍以上、イヌとでは 10 倍以上、サルとでは 100 倍、ウッドチャックとでは 30 倍以上の差があった。

その後の研究により、FIAU は hENT1 発現細胞でミトコンドリア内に輸送されるが¹³⁾、マウス ENT1 は FIAU を輸送するが、ミトコンドリアに分布せず、マウス肝臓から調製したミトコンドリアは FIAU を輸送しないことが明らかになった¹⁴⁾。

そこで、医薬品や農薬等の承認申請の際には以前から一般薬理試験 (安全性予測の面を強調し、安全性薬理試験とも呼ぶ) の実施が求められていた。MD 試験の実施前に、すべての安全性薬理試験を終了させておくことは必ずしも必要ではないが、適切な *in vivo*/*in vitro* 試験により、治療標的に関連した薬理作用や思いがけない薬理作用などを調べ、被験物質の主たる薬理作用について明らかにしておく必要がある。また、薬効発現量を明らかにしておくことが必要である。これは MD 試験における最高用量の設定や薬理作用の延長として現れる副作用を予測するために利用される。また、この段階で薬理作用発現の種差を調べておくことは、適切な毒性試験動物種の選択にも資する。

ICH M3 指針でも、呼吸、循環、中枢神経系への作用が生命にとってとくに重要であることから、ヒトに初めて投与する前に、それらに対する被験物質の作用を安全性薬理試験として GLP 下で実施することを求めてきた。一方、TGN1412 の臨床試験で予期せぬ重篤な障害が発生したことを踏まえ、EMA (2007)¹⁵⁾ はヒト初回投与量設定のために MABEL (Minimal Anticipated Biological Effect Level) 法を推奨した。これは、すべての利用可能な *in vitro* および *in vivo* の PK/PD データをモデリングアプローチでまとめ、必要に応じて、安全係数を適用して初回投与量を決定するものである。これに必要な PK/PD データとしては、ヒトおよび適切な動物種由来の標的細胞を用いた 1) *in vitro* の標的部位への結合および受容体占有データ、2) *in vitro* 濃度-反応曲線データ、また、適切な動物種での 3) *in vivo* 用量/暴露-反応関係データ、および 4) 薬理作用発現用量での暴露データが挙げられている。TGN1412 のヒト T 細胞刺激作用の最小有効濃度 (0.1 $\mu\text{g/mL}$) を基に計算すると、体重 70 kg の男性での投与量は約 0.003 mg/kg となる。これに安全係数を考慮し、さらに低い臨床用量で行うべきであったとされている¹⁶⁾。ICH M3 検討班では、薬理作用を *in vivo* およびヒト標本を含む *in vitro* 試験系で検討し、PK/PD 関係を考慮してヒトでの薬理作用発現用量を推定し、それに安全係数を加味して初回投与量を定めるべきとの考えから、ICH M3 (R2) 指針では MD 試験においても、*in vitro* での標的部位や各種受容体への結合濃度をあらかじめ調べることによる薬理的なプロファイリングを行っておくこと、また、薬力学的に適切なモデルを用いて薬理作用の特徴を明らかにすることを推奨した。

3) 薬理試験の必要な理由

一般毒性試験では必ずしも検知されない毒性があ

Table 3 遺伝毒性不純物の臨床開発段階での許容量 ($\mu\text{g}/\text{day}$)

	暴露期間						
	単回	2日-14日	14日-1月	1月-3月	3月-6月	6月-12月	12月以上
Müller et al (2006)	120	120	120	40	20	10	1.5
FDA 案 (2008)	120	120	60	20	10	5	1.5
EMA Q&A (2008)	120	60	60	20	10	5	1.5

TGNI412の経験¹⁶⁾

TGM1412は白血病と自己免疫疾患治療を目的とするCD28アゴニスト作用を有するモノクローナル抗体であり、カニクイザルで50 mg/kgを4週間投与しても忍容性が認められたことから、2006年3月13日のFIM臨床試験で健康志願者(6名)に0.1 mg/kg infusion(投与量はMD試験の約50倍)後、全員に極めて重篤な有害作用が発現した。Cytokine release syndrome or Cytokine storm(発熱、低血圧、多臓器不全)であり、集中治療を受けた。

4) 遺伝毒性試験について

遺伝毒性試験はMD試験に要求されていない。これは、発がん物質でも Threshold of Toxicological Concern (TTC: 1.5 $\mu\text{g}/\text{human}/\text{day}$, 約30 ng/kg/day)以下の暴露については、一般に発がんの懸念は無視できるとされていることによる^{17,18)}。ただし、aflatoxin-like compounds, N-nitroso-, azoxy-, steroids, TCDD-like compoundsについては、0.15 $\mu\text{g}/\text{day}/\text{human}$ でも生涯摂取すると発がんリスクが $1/10^6$ より高いものが多い¹⁹⁾。しかし、医薬品においては、ベネフィットも勘案し、生涯リスクを $1/10^5$ で判断するのが適切である²⁰⁻²²⁾。また、TTC値は特定の条件下(短期間暴露、生命に危険な状況、既知の不純物、食物等からの暴露)では1.5 $\mu\text{g}/\text{day}$ 以上でも良いと考えられる²¹⁾。そこで、Müllerら(2006)²⁰⁾はリスクが暴露量に比例するとの考えを前提としたLinear extrapolation modelにより生涯発がんリスクを $1/10^5$ に保ったまま、短期間に外挿した(Table 3)。また、FDA²²⁾やEMA²¹⁾は上記に不確実係数を2加味した基準を作成したが、ICH M3の検討グループではFDAの案を採用し、5回までの反復投与であれば、100 μg の投与が許容されると判断したものである。

6) 局所刺激性試験について

局所刺激性については、MD試験の実施前に評価しておく必要がある。ただし、動物福祉を考慮し、無駄な動物実験を排除するため、拡張型単回投与毒性試験における投与局所の観察により局所刺激性の評価が可

能であれば、改めて試験を実施する必要はない。ICH M3 (R2) 指針でも、MD試験の前に実施することを必ずしも求めている。ただし、新規の媒体を使用する場合には、媒体の局所毒性を評価すべきとされている。

6. その他

放射性標識体を用いる場合は、放射線被曝のレベルとその安全性に関する評価を事前に終了しておく必要がある。Accelerator Mass Spectrometry (AMS: 加速器質量分析法)を用いたMD試験では、放射線暴露は放射線障害防止法で定められた基準より低くとも、有用なデータを得ることが可能である。また、国際放射線防護委員会(ICRP: International Commission on Radiological Protection)による勧告(ICRP 1990)²³⁾で示された「一般公衆の年間被ばく線量限度」以下で実施可能である。ただし、不必要な摩擦を避け、社会に受け入れてもらうためには、独立した第三者委員会において適切な暴露/安全性評価を行うことにより、試験を実際に実施する医師や被験者の安全性についての懸念を除くことが望ましい。

バイオ医薬品等への適用に関して、FDA(2006)の指針⁴⁾によれば、30 nmolまでならば、生物製剤についてもMD試験を適用可能である。しかし、この根拠は不明確であり、これ以下の用量でも強い毒性を現す毒素等は多くある⁷⁾。そこで、MD試験指針では「生物由来製品又は体内で如何なる受容体が関与するか十分な知見が得られていないものなど従来の医薬品とは全く異なる作用機序による薬理作用を期待した化合物を投与する場合については、個別にその安全性等についての考察が必要であり、本ガイダンスをそのまま適用することはできない。」としたものである。しかし、これは、必ずしもこれらの物質についてMD試験実施を否定しているものではなく、その安全性について、投与方法や候補物質の特性を十分理解したうえで、より慎重な考察を行ったうえで実施することを求めたもの。しかし、現在の医薬品開発の多くが、バイオ医薬品であることを考慮すると、ICHでの審議を踏まえ、

それらの開発の標準的手順を示す必要がある。なお、ペプチド・タンパク質のMD試験の意義については、矢野ら²⁴⁾の総説を参照されたい。

7. 終わりに

ガイドラインに従って作成された非臨床試験の結果を適切に利用すれば、志願者や患者に起こりうる多くの毒性を予測しうる。しかし、すべての物質について、ヒトでの毒性を正確に予測することはできない。また、動物を用いる非臨床安全性試験では、これ以上、動物実験を行っても新たな知見は得られないという限界もある。基本的に、臨床試験に必要な非臨床試験の範囲は、被験物質の特性や投与用量、また、被験者の特性や人数に応じたリスクとベネフィットを勘案しながら、定めるべきものである。また、事前に得られた臨床、非臨床データに基づき、安全性を確認しながら、段階的に行われるべきものである。そこで、ICH M3 (R2) の改訂に際しては、非臨床試験とのタイミングを示すにあたり、臨床試験を第I、II、III相という分類ではなく、試験期間、対象被験者数、疾病の種類により区分された。

文 献

- 厚生労働省医薬食品局審査管理課長通知(2008)「マイクロドーズ臨床試験の実施に関するガイダンス」(薬食審査発第0603001号)。
- ICH M3 (R2) (2009) Guidance on nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals. ICH consensus guideline.
- EMA (2003) Committee for Proprietary Medicinal Products. Position paper on non-clinical safety studies to support clinical trials with a single microdose. (CPMP/SWP/2599/02, 2003.1.23)
- FDA (2006) Exploratory IND Studies. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER). January, 2006.
- 厚生労働省医薬食品局審査管理課 事務連絡(2008)「マイクロドーズ臨床試験の実施に関するガイダンス(案)」に対する意見募集の結果について(2008.6.3)
- FDA (2005) Guidance for Industry. Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER)
- 大野泰雄. マイクロドーズ臨床試験の毒性学的根拠について. マイクロドーズ臨床試験:理論と実践—新たな創薬開発ツールの活用に向けて—. 杉山雄一, 栗原千絵子(編), じほう社, 2007.
- 山口照英, 石井明子. 次世代バイオ医薬品の開発にあたっての非臨床・臨床試験について —TGN1412事故が医薬品開発に与えたインパクト—. 谷本学校 毒性質問箱 第10号. 安全性評価研究会編集委員会(編), サイエンティスト社, 2007.
- Brahams D. Deaths in US fialuridine trial. *Lancet*. 1994; 343 (8911): 1494-5.
- McKenzie R, Fried MW, Sallie R, Conjeevaram H, Di Bisceglie AM, Park Y, et al. Hepatic failure and lactic acidosis due to fialuridine (FIAU), an investigational nucleoside analogue for chronic hepatitis B. *N Engl J Med*. 1995; 333: 1099-105.
- Richardson FC, Engelhardt JA, Bowsler RR. Fialuridine accumulates in DNA of dogs, monkeys, and rats following long-term oral administration. *Proc Natl Acad Sci USA*. 1994; 91: 12003-7.
- Richardson FC, Tennant BC, Meyer DJ, Richardson KA, Mann PC, McGinty GR, et al. An evaluation of the toxicities of 2'-fluorouridine and 2'-fluorocytidine-HCl in F344 rats and woodchucks (*Marmota monax*). *Toxicol Pathol*. 1999; 27: 607-17.
- Lai Y, Tse CM, Unadkat JD. Mitochondrial expression of the human equilibrative nucleoside transporter 1 (hENT1) results in enhanced mitochondrial toxicity of antiviral drugs. *J Biol Chem*. 2004; 279: 4490-7.
- Lee EW, Lai Y, Zhang H, Unadkat JD. Identification of the mitochondrial targeting signal of the human equilibrative nucleoside transporter 1 (hENT1): implications for interspecies differences in mitochondrial toxicity of fialuridine. *J Biol Chem*. 2006; 281: 16700-6.
- EMA (2007) Guideline on strategies to identify and mitigate risks for first-in-human clinical trials with investigational medicinal products.
- Duff, Gordon W. Expert Group on Phase One Clinical Trials: Final Report. 2006, Dec 7.
- JECFA (1999), WHO Technical Report Series 868, 49th report of the Joint FAO/WHO expert Committee on Food Additives.
- Munro IC, Ford RA, Kennepohl E, Sprenger JG. Correlation of structural class with no-observed-effect levels: a proposal for establishing a threshold of concern. *Food Chem Toxicol*. 1996; 34: 829-67.
- Kroes R, Renwick AG, Cheeseman M, Kleiner J, Mangelsdorf I, Piersma A, et al.; European branch of the International Life Sciences Institute. Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. *Food Chem Toxicol*. 2004; 42: 65-83.
- Müller L, Mauthe RJ, Riley CM, Andino MM, Antonis DD, Beels C, et al. A rationale for determining, testing, and controlling specific impurities in pharmaceuticals that possess potential for genotoxicity. *Regul Toxicol Pharmacol*. 2006; 44: 198-211.
- EMA (2006) Guideline on the limits of genotoxic impurities.
- FDA (2008) Guidance for Industry. Safety Testing of Drug Metabolites.
- ICRP (1990) Recommendations of the International Commission on Radiological Protection. *ICRP Publication* 60.
- 矢野恒夫, 杉山雄一, 加藤基浩, 馬屋原宏, 清原孝雄, 残華淳彦ほか. ペプチド・タンパク質のマイクロドーズ臨床試験の意義・実施基盤. *臨床評価* 2008; 35: 651-65.

An experimental design for judging synergism on consideration to endocrine disruptor animal experiments

Nobuhito Matsunaga^{1*,†}, Jun Kanno², Chikuma Hamada³ and Isao Yoshimura³

¹*Kyowa Pharmaceutical, Inc., 212 Carnegie Center, Suite 101, Princeton, NJ 08540, USA*

²*National Institute of Health Sciences, Tokyo, Japan*

³*Tokyo University of Science, Tokyo, Japan*

SUMMARY

This paper investigates an appropriate statistical design for an animal experiment to evaluate synergism of two test chemicals. It assumes a certain number of animals are divided into groups, each of which is treated with a combination of dose levels of two chemicals. A design is identified by the set of group size for each combination of doses, including the case where the dose of either one chemical is zero. The power of *t*-test to detect synergism by positive surplus of response on a simultaneous administration group from the additivity plane composed of the responses on single administration groups is adopted as the criterion for the appropriate design. The applicable design is investigated for the application to real cases of endocrine disrupter study conducted at the National Institute of Health Sciences of Japan.

It revealed that the dose level of the simultaneous administration group should be located inside or on the boundary of a triangular region and that the total number of animals should be the same as those for single administration groups. Copyright © 2008 John Wiley & Sons, Ltd.

KEY WORDS: additivity; animal experiment; experimental design; endocrine disruptor; synergism; triangular region

1. INTRODUCTION

In the past, environmental pollutants were regulated according to individual effects. However, recently, there has arisen the problems of combinations of complex pollutants, and regulations that address synergism have become necessary. As a result, experimental researches have been conducted on pollutant synergism. The investigation by Kanno, one of the authors (Kanno *et al.*, 2001) on the synergism of endocrine disruptors, using the rodent uterotrophic assay, is an example of such researches.

In our experiments using multiple test substances, dividing animals such as rats into multiple groups of single administration and simultaneous administration, we estimated the response when there is no synergism based on the response in the single administration group to investigate whether the response in the simultaneous administration group exceeds the estimated response.

*Correspondence to: N. Matsunaga, Kyowa Pharmaceutical, Inc., 212 Carnegie Center, Suite 101, Princeton, NJ 08540, USA.

†E-mail: matsunaga.nobuhito@kyowa-kpi.com

For the data analysis method used in this type of animal experiment to investigate synergism, Kelly and Rice (1990) proposed a method to evaluate the dose–response curve by smoothing method. Gennings and Carter (1995) and Gennings *et al.* (1997), on the other hand, proposed a method to evaluate synergism by using a model in which the response becomes flattened when there is no synergism. Using a similar plan to that of Gennings *et al.* (1997), Matsunaga *et al.* (2003) proposed a method to evaluate the difference in the responses with simultaneous administration of two substances from those estimated by applying an additive model to the data of single administration of each substance. They applied their proposed method to the actual data analysis. Other data analysis methods are also cited in Laska and Meisner (1989) and Machado and Robinson (1994).

For the experimental design evaluating the synergism, Hasegawa *et al.* (1996) proposed the experimental design of animal experiment for five or ten chemical mixture, and Straetemans *et al.* (2005) investigated a fixed-ratio design on *in vitro* study. Abdelbasit and Plackett (1982) and Tan *et al.* (2003) are also related to this issue. However, the situation of these researches differs from our case study. Our interest is limited to simply checking whether the effects of the chemicals are additive or not. And in the animal experiments, some assumptions and limitations are generated for the applicable information and experimental conditions.

In this kind of research, we thought that the problem of the experimental design is to determine appropriate dose levels for simultaneous administration and to select the appropriate number of animals for allocation to the dose levels. However, in the past research, the research directly related to this problem by the animal experiment was not found. Accordingly, based on the analysis methods proposed by Matsunaga *et al.* (2003), we investigated what type of design would be appropriate.

The paper is organized as follows. Section 2 introduces the conditions in the case study that motivated this paper, while Section 3 formulates the issues. Section 4 derived the appropriate design corresponding to the case studies dealt with this paper. Finally, Section 5 provides Conclusions and Discussions for future issues.

2. MOTIVATING CASE

According to the World Health Organization, an endocrine disruptor is defined as “an exogenous substance or mixture that alters function(s) of the endocrine system and causes adverse health effects in an intact organism, or its progeny or (sub)populations.”

The effect of endocrine disruptors is not stimulated directly at the site of the adverse effect, but is mediated by the signal and occurs through nuclear receptor. Furthermore, there is more than one signal transduction system in humans and animals. Because nuclear receptors and transcription factors are redundant, there may be an interaction between different pathways, which leads to possible synergistic endocrine disruption action. Here the definition of “synergistic” is that if two chemicals produce the same endpoint, they bring about a larger response as compared to the anticipated response when these chemicals are purely added. Consequently, it is very important to realize the synergism between two endocrine disruptors such as “Genistein” and “bisphenol A (BPA)” through animal experiments for explaining the mechanism of action.

Both Genistein and BPA bind to estrogen receptors and elicit estrogenic responses to an organism including uterotrophic responses. Genistein is a phytoestrogen found relatively abundant in soybeans and its derivative foods. BPA is the basic monomer of polycarbonate plastic and epoxy resins widely used as a lining for food and beverage cans, in hard plastic baby and water bottles, toys, dental sealants, etc. It has been reported that BPA monomer can leached out to food and drinks especially when the

polymerization process is incomplete and/or the plastic is aged. These two estrogenic compounds can be found very commonly in our food environment. Therefore, it is of great importance from the point of the safety regulations to examine whether the combined effect is additive or synergistic.

For the experimental design of the research based on this background, some assumptions and limitations were generated for the applicable advanced information and experimental conditions. In the interests of simplicity, when explaining synergism research for two substances, the limitations are as follows.

First, before conducting the experiment that investigates synergism, advance knowledge can be obtained to some extent on dose–response curves for the single administration from preliminary experiments using each substance. Therefore, if the dose–response curve is nonlinear, by appropriate transformation of variables for the dose and response, it is possible to assume the dose–response curve to be approximately linear.

Second, the maximum dose of each substance used in the experiment is limited. We, for example, recognize in our experiments that the signal transduction system amplifies the signal at significantly smaller doses as compared to the dose used in normal toxicity studies. Other toxicities appear with higher doses, so that the endocrine disruption effect to be investigated is concealed. Because these maximum doses, $D_{A\max}$ and $D_{B\max}$, generally can be obtained through preliminary experiments, the range of dose levels used in the experiments can be limited.

Third, because various kinds of test substances are to be investigated, the number of animals, n , used in each experiment is relatively small. In our actual experience, the number of animals is approximately 40–50.

Fourth, the fundamental form of the experiment for investigating synergism is roughly decided. In an experiment using two substances A and B , we set single administration groups of G_{00} (at a dose level of 0), G_{A2} , and G_{B2} (at the maximum dose levels of $d_{A2} = D_{A\max}$ and $d_{B2} = D_{B\max}$, respectively), and groups G_{A1} and G_{B1} (at the middle dose levels of d_{A1} and d_{B1} , respectively), and administer the test substances after assigning the same number of animals n_s in every group. Independent of this, we set one or more simultaneous administration groups of G_{AB} with dose levels of the two substances at d_A and d_B . We measure responses by performing the experiment with this type of design, and estimate dose–response curves by forecasting synergism from the single administration group to confirm whether the response obtained in the simultaneous administration group is larger compared to that estimated.

Fifth, the observed response is usually quantitative variable such as the uterine weight of rats, which generally shows normal distribution, because it is difficult to define the additivity/synergism for the response in qualitative values.

Under the above conditions, what should be questioned for the design is what the most appropriate dose for simultaneous administration is, and whether to have more or less animals for the simultaneous administration group as compared to the single administration group. In order to obtain guidelines for these, this paper generalizes and formulates the above mentioned problems to make numerical evaluations under some conditions.

3. FORMULATION OF PROBLEM

3.1. Definition of synergism

In order to simplify the discussion, hereafter, we assume that there are two test substances denoted by A and B .

There have been many discussions in the past for how to define the terms additivity, synergism, and antagonism. Synergism is not defined unconditionally (Hewlett and Plackett, 1959; Berenberm, 1989).

Table 1. Difference between factorial design and triangular design

Dose of B	Dose of A		
	d_{A0}	d_{A1}	d_{A2}
(a) Factorial design			
d_{B0}	(1)	(2)	(3)
d_{B1}	(4)	(6)	(7)
d_{B2}	(5)	(8)	(9)
(b) Triangular design			
d_{B0}	(1)	(2)	(3)
d_{B1}	(4)	(6)	
d_{B2}	(5)		

Single chemical is administered at (1)–(5), whereas combination of two chemicals is administered at (6)–(9).

In fact, in the general remarks of these studies, many ideas are introduced for discussing synergism such as “independent joint action,” “potentiation,” “simple similar action,” “complex action,” and “dissimilar action.” We will first explain the definition of synergism that is adopted in this paper.

From the standard statistical viewpoint, the dosages set certain dose level for the respective two substances as shown in Table 1(a). If the response at the simultaneous dose level is the sum of the effects generated by single substances, then the effect is considered to be additive. On the other hand, if it is large, there is a positive interaction and the effect is synergistic.

However, with toxic responses like endocrine disruption action, this point of view is not appropriate. Because, for this toxic response, as pointed out by, for example, Hasegawa *et al.* (1996), it is impossible to establish response linearity at doses that exceed the maximum dose for the respective substances, and thus it is impossible to determine whether a positive interaction is attributable to synergism or nonlinearity. Therefore, the following definition that expresses the tenets of Hewlett and Plackett (1959) by isobologram is adopted in this paper.

Label the expected response at dose d_A and d_B (doses for A and B) as $f(d_A, d_B)$. Also, label the single administration dose of A that results in an arbitrary response E as D_A , so that $f(D_A, 0) = E$. Similarly, label the dose of substance B that has expectation E as D_B . In the cases that motivated this study, A and B generate their responses in a similar stimulation process, the expected effects are proportional to the doses of A and B, and the effects of the two substances are additive. If these conditions hold, then $f(d_A, d_B) = E$ whenever (d_A, d_B) satisfies Equation (1).

$$\frac{d_A}{D_A} + \frac{d_B}{D_B} = 1 \quad (1)$$

The reason is that because a combination of dose levels like this represents simultaneous administration of A and B at an arbitrary ratio, using the amount that brings about a response of the same magnitude. If a response of magnitude E is consequently generated as expected, there is no special combined effect between the two chemicals. In this paper, when this relationship holds, the effect of the two substances is additive, or the two substances satisfy additivity.

On the other hand, if the two substances generate a synergistic response in different stimulation processes, it is considered that $f(d_A, d_B) > E$ is established with respect to an arbitrary (d_A, d_B) that satisfies Equation (1). In this paper, when this relationship is established, the two substances are synergistic, or satisfy synergism.

3.2. Terminology, notation, and assumption

The two-dimensional plane by plotting d_A (dose of A) on the x -axis and d_B (dose of B) on the y -axis is referred to as the dose plane, and the three-dimensional space by plotting the response on the z -axis above the dose plane is referred to as the response space.

In the experiment, the response is measured for each individual animal. The response that is measured is called the response variable, and is generally expressed by the symbol Z . The response variable Z measured for each individual is set as a random variable that follows a normal distribution independent of other individuals. Since it is assumed that the endpoint is organ weight as a target for application in the case study, such as endocrine disruptor study, it is considered that the assumption of the normal distribution is empirically valid. When the dose of the two substances administered is (d_A, d_B) , the expected value is $E\{Z\} = f(d_A, d_B)$.

For single administration, that is, when the dose of one test substance is 0, Equation (2) can be assumed concerning the dose–response curve f .

$$f(d_A, 0) = \beta_0 + d_A\beta_A, f(0, d_B) = \beta_0 + d_B\beta_B \quad (2)$$

This assumes the dose–response curve for single administration to be linear. With this assumption, $f(d_A, d_B)$ is expressed by Equation (3). The two substances are additive if the hypothesis H_0 of Equation (4) holds, while synergistic if the hypothesis H_1 holds.

$$f(d_A, d_B) = \beta_0 + d_A\beta_A + d_B\beta_B + \Delta(d_A, d_B) \quad (3)$$

$$H_0 : \Delta(d_A, d_B) = 0, H_1 : \Delta(d_A, d_B) > 0 \text{ for all } (d_A, d_B) \quad (4)$$

The value of the dose used in the experiment is called the dose level, the collection of animals allocated for the dose level is called a group, and the number of animals for each group is called the group size, and the dose level of the group on the dose plane is called the group point. With this terminology, it is defined that “design is the set of group point and group size.”

For numerical evaluation described in the next section, five groups of group size n_s for single administration and one group of group size n_m for simultaneous administration as described in Table 1(b) are assumed as the design. For the single administration group, the group points are set to be $(0, 0)$, $(d_{A1}, 0)$, $(d_{A2}, 0)$, $(0, d_{B1})$, $(0, d_{B2})$, and the response variables are distributed as normal with variance σ_s^2 . For the simultaneous group, the group point is set to be (d_A, d_B) , and the response variable is distributed as normal with variance σ_m^2 .

Let the sample mean of response variable in each group be Z_{00} , Z_{A1} , Z_{A2} , Z_{B1} , Z_{B2} , and Z_{AB} , respectively. It is assumed that Z_{00}, \dots, Z_{B2} are distributed as normal with the mean of Equation (2) and the variance σ_s^2/n_s , while Z_{AB} is distributed as normal with the mean of Equation (3) and the variance σ_m^2/n_m .

3.3. Criterion for the appropriate design

As criterion for the most appropriate design, it is natural to use the power in the hypothesis test of “ H_0 versus H_1 .” Because the model is a linear model and the hypothesis is a linear hypothesis, a one-sided

t -test (or Welch test) with Equation (5) is naturally set as the test statistic.

$$T = \frac{\hat{\Delta}}{\sqrt{\hat{V}(\hat{\Delta})}} \quad (5)$$

Here, $\hat{\Delta}$ is the least square estimator of Δ , and the denominator of the statistics is the square root of the variance estimator.

The critical value of this test statistics is $t(\nu, \alpha)$, the upper 100α percentile of the t -distribution with degree of freedom ν , and the test is a one-sided test. In other words, A and B are judged to be synergistic when $T > t(\nu, \alpha)$.

This test, in short, detects the synergism when there is a statistically significant difference between Z_{AB} and the estimate obtained from the single administration groups assuming the dose-response surface under H_0 .

4. EXAMPLES OF RECOMMENDED DESIGN

4.1. Real examples motivated the problem

We conducted many experiments and selected endocrine disruptor study as case study. This study was performed using triangular design such as Table 1(b). This design consisted of seven dose groups which included five group points for single administration and two group points for simultaneous administration. The endpoint was uterine weight gain and the main purpose was to evaluate whether the combined effect was synergistic or not. In order to explain the characteristic of the data, we took up two real examples. The details of these data are as follows.

Example 1. Chemical A: genistein (mg/kg), chemical B: BPA (mg/kg)

1. Group points for the single administration: $(d_A, d_B) = (0, 0), (12.5, 0), (25, 0), (0, 35), (0, 70)$.
2. Group points for simultaneous administration: $(d_A, d_B) = (6.25, 17.5), (12.5, 35)$.
3. Group size: $n_s = 6, n_m = 6, n = 42$.
4. Mean \pm standard deviation of observed values

$$(d_A, d_B) = (0, 0) : 84.0 \pm 7.1$$

$$(d_A, d_B) = (12.5, 0) : 111.2 \pm 9.3, \quad (d_A, d_B) = (25, 0) : 149.5 \pm 33.7$$

$$(d_A, d_B) = (0, 35) : 138.4 \pm 16.2, \quad (d_A, d_B) = (0, 70) : 181.0 \pm 24.4$$

$$(d_A, d_B) = (6.25, 17.5) : 141.1 \pm 11.4, \quad (d_A, d_B) = (12.5, 35) : 180.2 \pm 26.8$$

5. Estimated value of Δ : 15.6.
6. Result of t -test: Significant in one-sided Welch test with significance level 2.5% $T = 2.19, \nu = 19, p = 0.02$.

Example 2. Chemical A: diethylstilbestrol ($\mu\text{g}/\text{kg}$), chemical B: genistein (mg/kg)

1. Group points for the single administration: $(d_A, d_B) = (0, 0), (0.1, 0), (0.2, 0), (0, 12.5), (0, 25)$.
2. Group points for simultaneous administration: $(d_A, d_B) = (0.05, 6.25), (0.1, 12.5)$.

3. Group size: $n_s = 6, n_m = 6, n = 42$.
4. Mean \pm standard deviation of observed values

$$(d_A, d_B) = (0, 0) : 91.2 \pm 16.2$$

$$(d_A, d_B) = (0.1, 0) : 92.3 \pm 9.7, \quad (d_A, d_B) = (0.2, 0) : 96.5 \pm 5.4$$

$$(d_A, d_B) = (0, 12.5) : 165.4 \pm 27.0, \quad (d_A, d_B) = (0, 25) : 220.8 \pm 27.5$$

$$(d_A, d_B) = (0.05, 6.25) : 141.8 \pm 12.8, \quad (d_A, d_B) = (0.1, 12.5) : 183.5 \pm 10.3$$

5. Estimated value of Δ : 19.5.
6. Result of t -test: significant in one-sided Welch test with significance level 2.5% $T = 3.98, \nu = 31, p < 0.01$.

From the above results, we suggested that the combinations for these two agents were synergistic. The synergism was observed in the real situations such as endocrine disruptor study.

4.2. Recommended group point selection

The example introduced in the preceding section has two simultaneous administration groups. In this section, we investigate the most appropriate dose level among several dose levels in the case of one simultaneous administration group. The conditions in the investigation are as follows. The conditions are set pursuant to the example in the preceding section except for assuming $\sigma = 1$ without losing any generality.

4.2.1. Fixed condition.

1. Group points for single administration: $(d_A, d_B) = (0, 0), (1, 0), (2, 0), (0, 1), (0, 2)$.
2. Group size: $n_s = 6, n_m = 12$.
3. Parameters in the dose-response curve: $\beta_0 = 1.0, \beta_A = 1.0, \beta_B = 1.0$.
4. Variance σ^2 : $\sigma_s^2 = \sigma_m^2 = 1.0$.
5. Nominal significance level of t -test: one-sided 2.5%.

4.2.2. Varied condition.

6. Group points for simultaneous administration: $d_A = 0.1(0.1)2.0, d_B = 0.1(0.1)2.0$.
7. Strength of synergism Δ :
 - Case 1 (constant case): $\Delta = 1.0$
 - Case 2 (square root case): $\Delta = 0.8\sqrt{(d_A + d_B)}$
 - Case 3 (linear case): $\Delta = 0.6(d_A + d_B)$

Numerical calculations were performed to calculate the power under the above conditions. The left-hand side of Figures 1–3 shows a three-dimensional display on the vertical axis above the dose plane of the power in Cases 1–3, respectively. On the other hand, the right-hand side of Figures 1–3 represents power functions when $d_A = d_B$ at the dose level for simultaneous administration group.

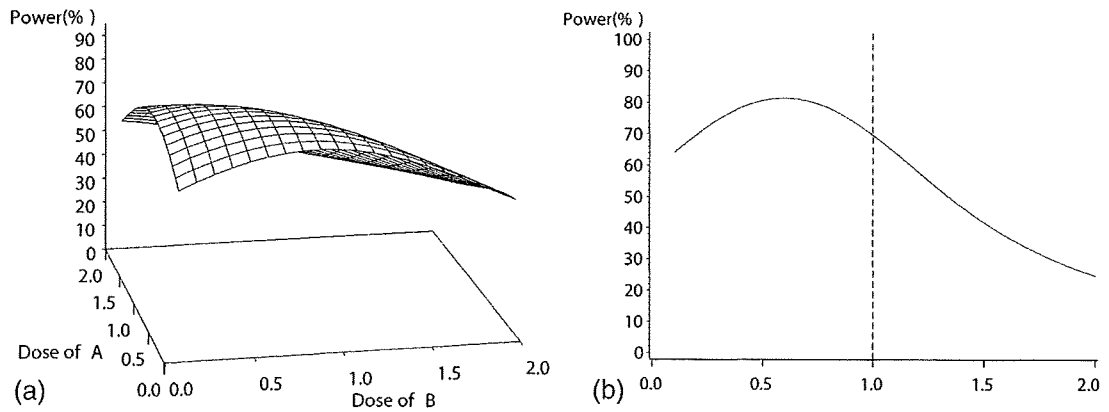


Figure 1. Power surface on the (a) dose plane and power function on the (b) dose for a simultaneous administration group with a constant surplus case (Case 1)

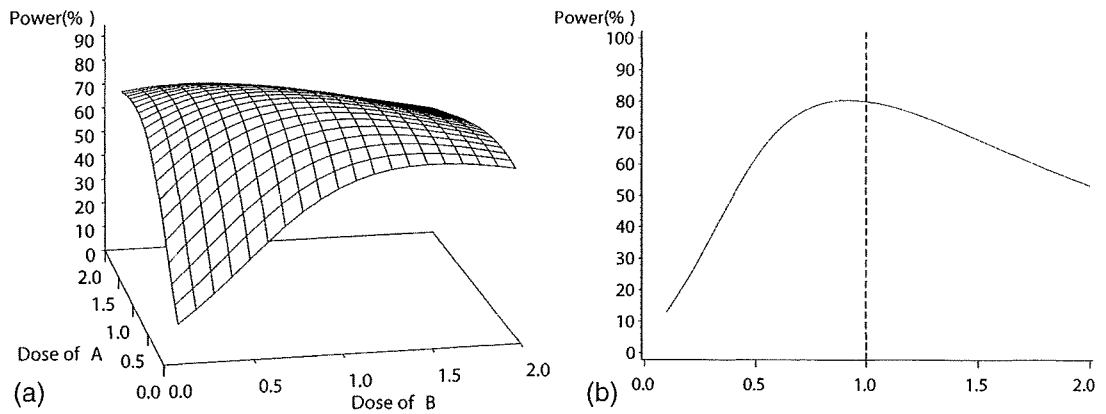


Figure 2. Power surface on the (a) dose plane and power function on the (b) dose for a simultaneous administration group with a square root surplus case (Case 2)

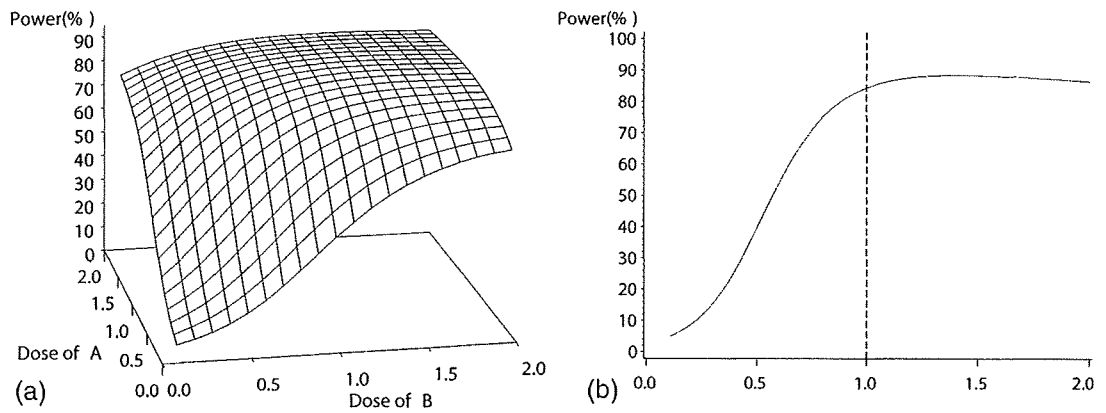


Figure 3. Power surface on the (a) dose plane and power function on the (b) dose for a simultaneous administration group with a linear surplus case (Case 3)

According to Figure 1, the power is high when the group point of the simultaneous administration group is within the triangular region and the power is low when it is located outside that region. Consequently, the design to evaluate synergism should have the group point within the triangle region. Figure 2 shows the power reaches a peak near the boundary of the triangular region and decreases apart from the triangular region as the dose increases. In consequence, the group point should be located on the boundary of triangle region. Finally, according to Figure 3, the power achieves the steady state on the boundary of the triangular region and reaches a peak slightly outside the triangular region when Δ becomes larger linearly as the dose increases. In numerical example, recommended group point in Cases 1–3 is, respectively, $(d_A, d_B) = (0.6, 0.6)$, $(0.9, 0.9)$, $(1.4, 1.4)$ and the power is 81.4%, 80.4%, 88.4%, respectively.

The sensitivity analysis was conducted to investigate the usefulness of location of the group point in the real study, such as two group points $(d_A, d_B) = (0.5, 0.5)$ and $(1.0, 1.0)$. In this section, only group points for simultaneous administration were changed in the following way and all other conditions were the same.

4.2.3. Group points for simultaneous administration.

- (1) $(d_A, d_B) = (0.6, 0.6)$ (recommended in Case 1)
- (2) $(d_A, d_B) = (0.9, 0.9)$ (recommended in Case 2)
- (3) $(d_A, d_B) = (1.4, 1.4)$ (recommended in Case 3)
- (4) $(d_A, d_B) = (1.0, 1.0)$ (on the boundary)
- (5) $(d_A, d_B) = (0.5, 0.5), (1.0, 1.0)$ (real study)

Here, the total sample size for simultaneous administration groups is fixed in all group points (1)–(5). Then n_m for two groups' simultaneous design such as (5) is half of that for one group's design.

The numerical calculation was conducted under the above conditions. Table 2 summarizes the reduction of power from a recommended group point. When the group point is located close to the boundary of the triangular region, the reductions of power from a recommended group point are small. Furthermore, in the situation of real studies, the loss of power is negligible compared to that in a recommended group point. It is considered that the configuration of the group points in real study is reasonable from the statistical viewpoint.

4.3. Recommended group size

In this section, under conditions in which the total number of animals is fixed and the effect sizes are varied, we investigate the group size of the simultaneous administration group that maximizes the power.

Table 2. Reduction of power from a recommended group point in three cases

Group point (d_A, d_B)	Case 1 (constant case)	Case 2 (square root case)	Case 3 (linear case)
(0.6, 0.6)	0.0	−10.2	−34.7
(0.9, 0.9)	−7.3	0.0	−9.3
(1.4, 1.4)	−34.6	−9.7	0.0
(1.0, 1.0)	−12.2	−0.4	−5.1
(0.5, 0.5), (1.0, 1.0)	−2.0	−2.6	−19.7

n_m for two group points is half of that for one group.