OECD/OCDE

2008/07/24

哺乳類の生殖毒性試験およびその評価に関するガイダンス文書

ご利用に関する注意事項

本文書および本文書に掲載している翻訳は OECD による公式文書**ではありません。**本文書における翻訳と原文の間に用語ないし表現等の齟齬がある場合、**原文のみが有効とみなされます**。内容について疑義が生じた場合、もしくは研究に公的な認証を必要とする場合は、原文に準拠してください。

本翻訳を使用したことに起因する如何なる損害について、当所は一切責任を負いません。

This document and translation are NOT official work by OECD. In the event of any discrepancy between the original work and the translation, **only the text of the original work should be considered valid**. Refer to the original document when there is any doubt about the translation, or legal or official verification is needed for your study. NIHS takes no responsibility even if you get any damage by using this translation.

Original work:

OECD (2008), Guidance Document on Mammalian Reproductive Toxicity Testing and Assessment, OECD Series on Testing and Assessment, No. 43, OECD Publishing, Paris, https://doi.org/10.1787/d2631d22-en.

©OECD(2008)

You are free to use this material subject to the terms and conditions available at

http://www.oecd.org/termsandconditions/

経済協力開発機構 （OECD）の化学物質の試験に関するガイドライン Section 4 Health effects
邦訳: 国立医薬品食品衛生研究所(NIHS) 安全性予測評価部

巻頭言

哺乳類の生殖毒性試験およびその評価に関する本ガイダンス文書（GD）は、化学物質のヒトおよび他の哺乳類に対する潜在的生殖毒性に関して、その試験を実施する際の方法論的側面およびデータ解釈についてのガイダンスを提供することを目的としている。また、神経毒性試験との関係についても取り上げる。本文書は、化学物質の潜在的生殖毒性に関する情報を入手するために用いられる現行の OECD 試験ガイドラインに対して、重要な捕捉をなす。

生殖毒性試験およびその評価に関するガイダンス文書を作成する計画は、1996 年の作業計画時に盛り込まれていた。数回の専門家会議を経て、2004 年 11 月、生殖毒性試験およびその評価に関するガイダンス文書の草案が、それについて意見を求めるため、試験ガイドラインプログラムの各国調整担当者らで構成する作業部会（Working Group of the National Coordinators of the Test Guidelines Programme：WNT）にて回覧された。

そこで出された意見をもとに GD 草案を修正すべく、GD を預かる起草班が 2005 年 10 月にパリで最終会議を開いた。改正された GD の新原案については、2007 年 12 月に WNT にて回覧され、それについても意見が出された。GD の最終案については、2008 年 2 月に事務局が意見に基づいて作成し、2008 年 4 月、第 20 回会議にて WNT によって承認された。その際、WNT は、化学物質の潜在的生殖毒性についての試験ガイドラインが改定される場合には、当該 GD を改正することでも合意した。

本文書は化学品委員会と化学品、農薬、バイオテクノロジーに関する作業部会との合同会合の責任において発行されたものである。

詳細についての問い合わせ先：
環境・保健・安全性部門
環境理事会
経済協力開発機構
2, rue André-Pascal
75775 Paris Cedex 16, France
E-mail：env.edcontact@oecd.org

目次

I. 緒言 15
本文書の履歴 15
本文書の目的 15
発達・生殖毒性の定義および生殖毒性に関する有害性評価の原則 16
生殖および発生における化学物質の作用の同定と評価 17
生殖毒性試験 17
成熟動物における全身作用評価 19
試験データの意義およびヒトとの関連性（外挿） 19
一世代および多世代生殖毒性試験 20
出生前発生毒性試験、催奇形性試験 22
発達神経毒性試験 23
生殖・発生毒性スクリーニング試験 24
その他の *in vivo* 試験 24
II. 出生前評価項目 25
評価項目の概要 25
実験動物を用いる試験 25
曝露の臨界期 25
潜在毒性 25
方法論的問題点 26
黄体数 26
胚の着床前および着床後損失 26
胎児および出生児の外表検査 27
胎児の内臓検査 27
骨格検査 27
データ解釈 28
奇形と変異 28
専門用語の国際統一 28
評価項目の相互関係 28
III. 出生後評価項目：出生児の成長、発育指標、ならびに機能および行動における神経毒性作用 33
評価項目の概要 33
方法論的問題点 33
同腹児数の標準化 33
曝露経路 34
クロスフォスタリング 35
児動物の日齢 35
検査を行う時間帯 35
児動物の体験事象 35
児動物の性別 36
行動学的検査結果に影響を及ぼす試験担当者側の因子 36
試験の自動化 36
行動評価の対象となる試験群について 36
身体的発達の指標 36
神経行動学的評価法概論 38
初期行動発達検査 39
運動量の測定法 41
運動機能および感覚機能の検査 43
学習および記憶の検査 46
水迷路 49
発達神経毒性（DNT）試験における神経系組織の死後検査 52
データ解釈 54
神経行動学的データの解釈における問題点 54
行動評価項目における試験の妥当性 55
神経病理学的評価項目 56

母動物の評価項目および児動物の評価項目の関連性 58
同腹児数、出生児の性別および平均体重 59
身体的発達および機能的発達の指標 59
IV. 成熟動物の生殖に関する評価項目 60
評価項目の概要 60
方法論的問題点 60
雄性生殖器の検査 60
精子パラメータ 61
雌性生殖器の検査 62
卵母細胞の定量化 62
膣細胞診 62
生殖能 63
データ解釈 64
雄性生殖器 64
精子パラメータ 64
雌性生殖器 65
生殖能 66
V. データの欠落 68
評価項目 68
曝露および毒性の臨界期 69
VI. 参考文献 70
Ison, J.R. (1984) Reflex modification as an objective test for sensory processing following toxicant exposure. Neurobehav. Toxicol. Teratol., 6:437-445 71
Koch, M. (1999) The neurobiology of startle. Prog. Neurobiol. 59:107-128. 72
Appendix 1 80
Appendix 2 82
Appendix 3 84

翻訳版注記：
本目次のページ表記は、英文ガイダンス文書原文ならびに本翻訳版の該当箇所と整合しているものではない。

I. 緒言

本文書の履歴

1. 1995 年 6 月､デンマークのコペンハーゲンにて OECD の生殖および発生毒性に関する作業部会が開かれた。その会合において、試験戦略、段階的試験方法、神経毒性試験との関係、データの解釈方法を盛り込んだ発達および生殖毒性試験に関するガイダンス文書が必要であるとの合意を得た。1996 年 10 月の第 7 回 WNT 会議にてガイダンス文書の作成が承認され、米国環境保護庁がこの動きに対して進んで先導的な役割を担った。

2. 本文書の草案をまとめるべく、専門家集団による会議が何度か開かれた。2003 年 6 月、最終的な問題の解決と文書化業務の割り当てのため、米国環境保護庁がワシントン D.C.にて専門家会議を主催した。2004 年 11 月、ガイダンス文書第 43 号の草案が加盟国に回覧され、意見が求められた。草案は概して好評であったが、出された意見はかなりの数に上った。意見の大半は単なる編集上の問題についてであったが、複雑な問題も持ち上がったため、2005 年 10 月、それらの問題に取り組み、ガイダンス文書案をしかるべく修正するために事務局は専門家集団をパリに招いた。

本文書の目的

3. 本文書は、化学物質のヒトおよび他の哺乳類に対する潜在的生殖毒性について、その試験の方法論的側面、データ解釈および全般的アプローチに関するガイダンスを提供することを目的としている。

4. 本文書は、化学物質の潜在的生殖毒性に関する情報を入手するために用いられる現行の OECD 試験ガイドラインに対して、重要な捕捉をなす。具体的な OECD 試験ガイドラインとして、一世代および二世代生殖毒性試験（TG 415 および 416）、出生前発生毒性試験（TG 414）、発達神経毒性試験（TG 426）および生殖・発生毒性スクリーニング試験（TG 421 および 422）が含まれる。ただし、全身毒性についての反復投与毒性試験（TG 407、408、409）など他の毒性試験によるデータから生殖毒性の可能性が示唆されることもあるため、評価時にはこれらのデータについても既知のヒトにおけるデータとともに考慮に入れるべきである。

5. 内分泌撹乱（ED）とは、内分泌系機能を変化させ、その結果、正常な個体、その子孫、または（下位）集団（an intact organism, or its progeny, or (sub) populations）の健康に悪影響を生じさせる外因性化学物質の能力を意味する。近年、EDに関する情報が広範に入手できるようになった（CSTEE 1999）。OECDでは、具体的な計画として内分泌攪乱物質の試験および評価に関する特別委員会（EDTA）を設置して、新たな試験ガイドラインの作成、およびEDに関する試験を包含すべく現行ガイドラインの改正に着手し、現在、生殖器系に焦点を絞っている。EDの問題に特別関心を置いて実施すれば、現行の試験ガイドラインでもEDに関連する生殖毒性の特定が可能かもしれない。しかし、ED関連作用の検出感度および検出特異性の向上を目的とするのであれば、現行の毒性試験ガイドラインには改善の余地がある。本ガイダンス文書ではED関連作用に関する考察も取り上げるが、ED自体が本文書の主目的ではない。

6. 関係諸国の多組織により生殖毒性データの評価に関する文書が作成された（ECETOC 1983, 1992, 2002、EHC 1984、LST 1989、NIOH（AMI）1994、DEPA 1995、IPCS-WHO 2001, OECD

2002a）。生殖毒性については、補遺 3 に有害性の評価方法に関するガイダンスを示す。

<u>発達・生殖毒性の定義および生殖毒性に関する有害性評価の原則</u>

7. 生殖とは、種の存続を確実にする生物学的過程である。現存する遺伝物質が次世代へ受け継がれるのである。

8. 生殖毒性には、成熟した雌雄の性機能および妊孕能に対する有害作用だけでなく、その胎児における発生毒性も含まれる。発生毒性と生殖毒性の区別については、発達期曝露が生殖に影響し、その逆の可能性もあるという点で、やや任意性がある。

9. 生殖能力に対する有害作用には、雄性および雌性生殖器官、または関連内分泌器官の変化が含まれ、以下のことに影響する。

- 性成熟の開始
- 配偶子形成および輸送
- 発情周期の正常性
- 性行動
- 妊孕能
- 出産
- 早期の生殖老化（premature reproductive senescence）、または
- 生殖器系の統合性に依存する他の機能の変化（泌乳など）

10. 発生毒性を最も広義に取った場合、受胎産物の出生前または生後の正常な発達を阻害する作用であり、受胎前の雌雄いずれかの親の曝露、または出生前発育期にある胎児もしくは生後性成熟期までの児の曝露により生じるものを指す。これらの作用は個体の一生のうちのいずれの時点でも表出しうる。発生毒性の主な発現には以下のものがある。

- 発育中の個体死亡
- 構造異常
- 発育の変化（成熟を含む）
- 機能的欠損

（次の文献を参照のこと：Clark (1999), Cooper and Goldman (1999), ECETOC (1983), EHC (1984), Heindel (1999)）。

11. 生殖毒性は、化学物質の単回投与または反復投与により生じる。単回曝露および反復曝露のいずれもがヒトへの曝露で予想される事態であるため、生殖毒性の評価にあたっては、理想的には両者について考えるべきである。

12. 生殖毒性評価については、各評価項目に関する項で論ずることとする。また、利用可能な方法論、試験動物種における正常な変異の範囲、評価項目（endopoints）間の相互関係、評価項目と機能面で生じた転帰（outcome）との関係、試験動物種間の相違、母体毒性との関係、結果の可逆性およびヒトとの関連性については、出生前評価項目、出生後評価項目および成熟動物における生殖毒性の章で論ずることとする。

生殖および発生における化学物質の作用の同定と評価

生殖毒性試験

13. 化学物質の健康評価において、生殖毒性試験は毒性試験計画の重要な部分を成す。毒性試験データからは、以下に関する必要な情報が得られるはずである。

- 危険有害性の同定
- 用量作用評価（可能であれば、NOAEL またはベンチマークドーズの推定*）
- 外挿（他の動物種、特にヒトにおける有害作用の予測）
- ヒトにおける曝露の安全レベルの予測

* 現行の OECD 試験ガイドラインはベンチマークドーズ法に対しては最適設計されていない。

14. 理想的には、生殖毒性試験によって、生殖に対する以下の作用のうち、1 つ以上の同定が可能でなければならない。

- 雄性のまたは雌性の生殖機能障害または生殖能力障害。すなわち、性欲、発情周期、性行動、精子形成や卵形成の諸段階、ホルモン活性と生理的反応に対する有害作用で、妊孕能、受胎、着床時までの受精卵の発育を妨げるもの。

- 子孫に対する非遺伝性有害作用の誘発。すなわち、最も広義には、出生前および出生後の正常な発育を妨げるあらゆる作用を含めるべきである。形態的な先天異常および機能不全（生殖内分泌や神経系への影響など）についても評価しなければならない。

15. 生殖および発生について、化学物質の毒性作用を調べるために用いられる実験的方法は多種多様にある（OECD 2002b、Toppari *et al.*, 1996、ECETOC 2002、Meyer and Svendsen 2003）。いくつかの試験は標準化されており、さまざまな政府機関および国際機関によって試験ガイドラインが公布されてきた。ここからは、表 1 に示すような、標準化され、承認を受けた規制上の OECD 試験ガイドラインに主眼を置く。優性致死試験、継代飼育による生殖能力評価、生殖腺の病理学的検査を行う反復投与毒性試験など表 1 に載っていない試験でも、正常な生殖を妨げる可能性のある化学物質の作用を明らかにすることができる。

16. 延長一世代生殖毒性試験に関する新たな OECD 試験ガイドラインについては現在作成中である。この試験ガイドラインは、国際生命科学研究所（ILSI）下の環境保健科学研究所（HESI）農薬安全性評価（ACSA）研究委員会が提案した Cooper ら（2006）によるライフステージ型延長一世代生殖試験（life stage F1 extended study）に広く基づいたものになる。試験デザインについては、いくつかの点で改良が加えられ、明確化がなされる予定である。さまざまな法的要求を満たし、動物実験の削減および苦痛の軽減という共通の目標にも対処するため、この試験ガイドラインでは、既存の情報に依拠して試験を開始し生存時観察を行うことで試験の方向を調整することの重要性が強調されており、柔軟性のあるより効率的な試験行程を示すものとなろう。この試験ガイドラインは 3 つのコホート群からなる。生殖発生の主要評価項目（コホート 3）に加え、発達神経毒性評価および免疫毒性評価の機会が提供されることになる（各々コホート 1 および 2）。コホート 1 および 2 の試験の実施にあたっては、既存の化学物質データベースの特性および各規制当局の要請に従うことになる。第二世代の産出は既存データと試験の生存時観察にお

ける所見次第である。これに関しては、Janer ら（2007）が第一世代と第二世代間の潜在的な相違を評価するために 176 件の多世代試験について回顧的分析を行っている。この分析結果が、現行の二世代生殖毒性試験を延長一世代生殖毒性試験に置き換えようという提案の根拠となっている。これらのことから、新試験ガイドラインが TG 415 および 416 の代替案となりうるわけである。新ガイドラインは現在作成中であるため、本ガイダンス文書には盛り込まなかったが、次回改定時には入れる予定である。

表 1. 生殖毒性試験の *in vivo* 試験法の要約

試験法	曝露期間	親および児動物における評価項目	関係ガイドライン
世代試験	一世代、二世代、または数世代にわたって連続曝露	成長、発生、生存能力。 妊娠期間および出産結果。 生殖器および標的器官の病理組織学的検査。 妊孕能。 TG 416 では発情周期および精子の質。	TG 415：一世代生殖毒性試験 TG 416：二世代生殖毒性試験
出生前発生毒性試験 （催奇形性試験）	着床から出産前日まで	1 腹の構成（吸収胚、生存胎児、死亡胎児など）。 胚発生。 胎児成長。 形態変異および奇形。	TG 414：出生前発生毒性試験
発達神経毒性試験	妊娠および授乳期間中	妊娠期間および出産結果。 身体的および機能的成熟。 CNS および PNS 作用による行動変化。 脳重量および神経病理学的検査。	TG 426：発達神経毒性試験
生殖・発生毒性スクリーニング試験	交配前 2 週間から分娩後 4 日まで	受胎能。 妊娠期間および出産結果。 生殖器および標的器官の病理組織学的検査（TG 422 では脳も）。 胎児および児動物の成長および生後 3 日目までの生存率。	TG 421 および 422

17. 近年、動物を用いる発生毒性試験の代替法としてさまざまな *in vitro* 試験~~系~~が提案されている（ATLA 2002）。通常、これらの試験は生殖周期の単一事象を対象にしているため、*in vivo* 試験における有害作用の評価としては不十分であり、動物を用いた化学物質のリスク評価試験に置き換えられるものではない。しかし、類似する化学物質のスクリーニングおよび基礎的な作用機序の解明には有用である。また、段階的な試験および評価戦略にとっても不可欠な要素となろう（付録 3 を参照）。非哺乳類でも研究が行われており、たとえば、米国の国家毒性プログラムは非哺乳動物を用いたミディアムスループットの発達神経毒性試験として土壌センチュウ（*Caenorhabditis elegans*）をモデルとした研究を行っている（US NTP, 2006）。その他の *in vivo* 試験については、段落 44～45 を参照されたい。

成熟動物における全身作用評価

18. 成熟動物を用いた反復投与毒性試験では、成長、一般症状および行動の観察、血液検査、生化学検査、臓器重量測定、ならびに生殖器官を含む臓器の病理学的および病理組織学的検査によって潜在的全身毒性についての情報が得られる。特定の生殖器官（精巣、卵巣など）における病理組織学的変化は生殖毒性の可能性を示す感受性のある指標であり、一般毒性試験でこれらの指標に陽性の影響が認められた場合は、生殖毒性の評価において価値があるものとなりうる（Dent 2007、Chapin *et al.*, 1998）。

19. 生殖毒性試験では、奇形、成長遅延、胚または胎児の死亡および出生後死亡、妊孕能、中枢神経系機能に対する作用など、いろいろな発生影響について情報が得られる。試験ガイドラインの中には特異的な情報を与えるものもあり、たとえば、妊孕能および生殖器官に対する発生影響は二世代試験によってしか分からず、また脳の発生および機能に対する影響は発達神経毒性試験によって検証される。

20. 生殖毒性試験における全身作用（systemic effects）の検証は、血液検査および生化学検査などが通常行われないなど、成熟動物を用いた反復投与毒性試験とは異なる。また、臓器重量の測定、病理学的および病理組織学的検査は生殖器官と特定の標的器官に限られる。したがって、出生前または出生後発育期間に生じた全身作用（肝臓および腎臓など）については、特定された標的器官を除き、標準の生殖毒性試験では特定されない。

試験データの意義およびヒトとの関連性（外挿）

21. ある物質がヒトの生殖毒性物質であることの最終的な証明は、ヒトの曝露結果からしか成立しない。一般的に、毒性作用に関しては、Calabrese（1983）が表した以下の表現が妥当である。

> *「細胞の構造および生化学は細胞のリポタンパク質膜をはじめとして動物界全体で非常に似通っており、細胞内への生体異物の吸収から、解糖、クレブス回路、その他非常に多くの中間代謝などの代謝過程にまでその影響が及ぶ。細胞レベルでの種間類似性が明白であることが、研究者らが機能について一動物種から別の動物種へ外挿または推論する根拠となる。」*

発生の基本過程は哺乳類間では高度に保存されることが知られている。このため、発生毒性を実証するために実験動物を用いて行った実験的試験の所見は、ヒトで生じうる反応を示唆するものである（Jakobsen and Meyer 1989、Schardein *et al.*, 1985、IPCS 2001）。ただし、種間変異というものもあり、一哺乳動物種で発生毒性が認められないことが、必ずしもヒトに毒性がないことを示唆するものではない。このような動物種または系統による差異は、化学物質の吸収、分布、代謝および排泄の違い、胎盤の構造、透過性および血流（Schröder 1995）および種差間の遺伝的背景（Kawakami *et al.*, 2006）で認められる。

22. 一動物種だけで投与物質に対するヒトの反応をいつも真に予測できるような動物種はない。複数の動物種を用いた試験では動物試験データの予測的妥当性が増すが、特定の作用のヒトとの関連性の評価に際しては試験動物種とヒトとの種差を考慮しなければならない（Jakobsen and Meyer 1989）。

23. げっ歯類の妊孕能に対する作用はヒトでの作用をよく表すと考えられており、ヒトの避妊薬の研究ではそのほとんどで最初にげっ歯類の試験を行う（Barlow and Sullivan 1982）。ただし、げっ歯類で妊孕能への影響が認められるには、精子数が大幅に減少しなければならないことを強調しておかなくてはならない（段落 181～183 を参照）。また、ほとんどの生殖試験は多胎動物で行われるため、産児数の有意な減少は軽度であっても精子パラメータに関連する機能障害を示唆している可能性がある（段落 194 を参照）。

一世代および多世代生殖毒性試験

24. 一世代および多世代生殖毒性試験の実施に関する試験ガイドラインは OECD、その他の機関から出されている。

25. これらの試験の目的は、雄および雌動物の妊孕能における物質の作用、ならびに催奇形性や周産期および出生後の母動物に対する影響を含め、卵子、胚・胎児および児動物に対する出生前、周産期、出生後の作用を同定するために継続的世代について調べることである。

26. 各種試験ガイドラインに共通する要件は多い。望ましい動物種はラットおよびマウスである。他の動物種を用いるのが妥当な場合（推奨される動物種とヒトで毒物動態に差異があると予想される場合、曖昧な結果を明らかにするため、またはすでに観察された作用についてさらに調べるためなど）には、他の動物種を用いても良い。被験物質は各群の動物に投与する（各群において、妊娠末期にそれぞれ大体 20 匹の妊娠動物が確保されるために必要な十分な数の動物を用いる）。通常、親世代動物への被験物質投与については、交配させる前に、少なくとも 1 精子形成サイクル全体および卵成熟末期を含む期間にわたって行う。雌動物については、交配期間および妊娠期間を通してその児動物の離乳まで連続曝露を行う。限度試験を採用しない場合は、少なくとも 3 段階の用量の投与群および 1 群の対照群（未投与群または溶媒投与群）を用いる。同腹児数の標準化は任意である（III 章を参照）。被験物質の物理化学的性質または生物学的効果によって制限を受けない限り、理想的には、最高用量は親動物に対して毒性を現わすが死亡させない量とする。低用量は、理想的には、親動物および出生児に対していかなる有害作用をも現さない量とする。

27. 記録は出生児の体重、出生後の成長および生存率について行うが、データを評価する際には、各腹の産児数または性比の違いおよび出生時の交尾後日齢（すなわち妊娠日数）の違いによる変動も考慮することが重要である。出生児の体重または体重増加量の変化については、連続変数であるとの理由などから、発生毒性の感受性指標となる。出生児の体重の違いが世代試験における発生毒性の唯一の指標となる場合もある。体重における差異が永続的な作用であるのか、一過性の作用であるのかは、いつも疑問の残る部分であるが、胎児または新生児の短期間の体重変化の長期的影響についてはほとんど分かっていない。したがって、NOAEL を決定する際には関連作用として体重の差異を考慮すべきである。

28. 二世代生殖毒性試験（TG 416）には、出生児についての精子の質や性周期に対する影響評価が含まれてきた。さらに、出生児の影響評価では、いくつかのオプショナルな行動指標と生殖器官、脳、および特定標的部位についての病理組織学的検査など、発達の指標に関するデータの記録が加えられている。

29. 多世代生殖毒性試験では生殖周期のすべての相で雄および雌動物の両者に対して曝露を行う。F1 動物については、他の試験では検討されない妊孕能および繁殖成績を含め、完全な評価が

行われる。F1 動物は、子宮内からの反復投与毒性試験に相当するもの（an equivalent of an *in utero* derived repeated dose toxicity study）であると考えられるという点で独特のものである。

30. 多世代生殖毒性試験は、発達神経毒性試験とならんで、通常の反復毒性試験では曝露開始期にあたる離乳期から約 5、6 週齢までの各個体の初期発達期間が対象となっている、げっ歯類を用いた唯一の毒性試験である。このため、これらの試験については、用いる指標に十分な感受性があれば、全評価項目について評価ができる。ただし、これらは労働集約的な試験であり、研究者間には、当該試験ガイドラインの要求事項にはデータの妥当性を妨げるような測定項目が多すぎるという意見がある。他方では、測定項目を追加すべきとの意見もある。これらの生殖試験に関するガイドラインが現在一斉に改定を受けており、新たな測定項目、および一部にはすでに含まれている苦痛の軽減または代替法などが取り入れられる予定である。新生児の死亡および奇形などの評価項目については、一般的に用いられる実験動物では、死亡した児動物、または重度奇形を有する児動物を分娩直後に食べてしまうことがあるという具体的な問題がある。そのため、同腹児数の減少から間接的にしか影響が分からないこともある。児動物の奇形または死亡がごくわずかであれば、同腹児数の減少が正常変動と比較して小さいために、検出されない、または統計学的有意差が認められない可能性もある。胚の着床前損失および吸収は同腹児数の減少によって間接的にわかるため、これらの指標としての感度はどちらかというと低い。

出生前発生毒性試験、催奇形性試験

31. 出生前発生毒性試験（TG 414）は妊娠中曝露の結果として生じる胚・胎児毒性を調べる試験法である。過去の催奇形性試験では、奇形および死亡のみを関連評価項目と捉える傾向があった。今日の試験では、成長遅延、構造異常および致死性に着眼されている。その結果、TG 414 のタイトルが「出生前発生毒性試験」に変わった。

32.推奨される動物種はげっ歯類（ラットおよびマウスなど）および非げっ歯類（ウサギなど）である。必要性が認められる場合には（推奨動物種とヒトとで毒物動態に差異があると予想される場合に、曖昧な結果を明らかにするため、または認められた作用についてさらに調べるためなど）、他の動物種を用いても良い。

33. 交配経験のない若齢成熟雌を人工授精させるか、雄と交配させる。交配の成立については、交配の観察（ウサギなど）、膣栓の確認（精子、凝固腺分泌物および膣部の細胞と粘液の混合物）、膣垢塗抹標本（ラット）により決定するか、または人工授精の実施時（ブタ、ウサギなど）とする。現行の OECD 試験ガイドライン 414 では、用量作用関係を確立するために 3 段階の用量および対照群（未投与群または溶媒対照群、1 群につき妊娠動物 20 匹、最低でも 16 匹とする）を用いる。改正前の旧 TG 414（1981）では、雌ラットに対して少なくとも妊娠期間の器官形成期、すなわち着床が起こる 6 日から 15 日まで（マウスでは 6～15 日、ウサギでは 6～18 日）曝露させることになっていた。この期間は催奇形性の最高感受期であることが分かっていた（ヒトの感受期に当てはめると妊娠 18～60 日）。しかし、生殖器官および脳の発達は 15 日目以降も続くため、15 日目に曝露を止めるとこれらの臓器の奇形については突き止められない。そこで、現行の TG 414（2001）では、投与期間が着床から帝王切開予定日までと延長されている。予備試験が可能である場合、それにより胚の着床前損失の可能性が高くないと考えられれば、投与を交配から屠殺予定日前日までの全妊娠期間としても良い。一般状態の変化について供試動物を毎日観察する。また、体重および摂餌量については妊娠期間を通して記録する。出産予定日の前日に帝王切開により子宮を摘出し、子宮および胎児の検査を行う。母動物については、構造異常や病理学的変化の有無について肉眼的に検査する。着床前に投与を開始した場合、着床前損失、すなわち着床する前に失われた胚の数についても評価を行う。

34. 総着床数、すなわち生存胎児、死亡胎児および吸収胚（早期に死亡し、吸収された胚）の数を記録する。吸収胚については、妊娠期間のどの時点で胚の死亡が生じたかを決定するため、その程度（すなわち胚が吸収された時点での成長度。総吸収数、早期吸収数、後期吸収数）を記録する。

35. 胎児の性別判定、体重測定を行った後、奇形について肉眼検査を行う。発育遅延、ならびに内臓および骨格の発達に対する影響について評価する。その際、生存胎児については、内臓および骨格評価を行う前に適切な方法で安楽死させなければならない。

36. 出生前発生毒性試験（TG 414）は着床後の子宮内死（胚吸収）の検証に最適である。着床前に投与を開始する試験においては、着床前胚損失の評価も可能である。

37. 胎児体重の評価を行うが、そのデータの解析にあたっては、同腹児数または性比の差異による影響を考慮に入れることが重要である。

発達神経毒性試験

38. OECD 試験ガイドライン「発達神経毒性試験」（TG 426）は近年採択されたものであり、米国環境保護庁のガイドライン（US EPA 1998a）を基にしている。発達神経毒性試験は、妊娠期間および授乳期間中の母動物への曝露により、出生児の神経系の機能および形態に生じる潜在的危険性についてデータが得られるよう設計されている。これらの試験では、中枢神経系（CNS）および末梢神経系（PNS）に対する作用に起因する行動変化の特定が可能である。行動は肝臓、腎臓、内分泌系など他の臓器機能の影響を受けるため、これらの臓器に対する毒性作用も出生児の行動の一般的変化に反映される。それだけで複雑に入り組んだ行動機能をすべて反映できるような試験はない。行動について検査を行うためには、広範な指標を用いて各機能の変化を同定する。発達神経毒性試験で用いる方法論について検証した文献がいくつかある（Adams 1986、Francis *et al.*, 1990、IPCS 2001）。一般的に各種検査は身体発達検査、単純反射の検査、運動機能検査、感覚の発達検査および機能検査、自発運動試験、学習および記憶試験、ならびに神経学的検査に分かれる。

39. 推奨される動物種はラットである。OECD 試験ガイドラインでは、各用量群とも 20 腹とし、妊娠期間および授乳期間中投与することが望ましいとされている。出産後、出生児数を記録してから、同腹児数がすべて等しくなるように調整する。被験物質の影響については、母動物の母乳を通じた直接的影響であるのか、母乳産生量の変化、または曝露を受けている母動物の行動変化の結果による間接的影響であるのか決定するため、クロスフォスタリングを行う場合もある。クロスフォスタリングとは、曝露を受けている母動物とその児動物、対照群の母動物とその児動物を相互交換して育成させる方法である。

40. 児動物の評価については、出生後の成長期と成熟後に行い、神経および行動全般の異常を検出するための状態観察、身体発育、初期反射の発達、自発運動量、運動および感覚機能、学習および記憶の評価、ならびに脳重量測定および神経病理学的検査（形態計測評価を含む）で構成されている。

41. 発達神経毒性試験は独立した試験としても実施できるが、一世代または二世代生殖毒性試験などの生殖試験に組み込んで実施することも可能である。ただし、発達神経毒性試験を別の試験に組み込むか追加して実施する場合には、両種の試験の完全性を保つことが必須である。一世

代生殖毒性試験および多世代生殖毒性試験の節で述べた、出生児の死亡、奇形、着床前胚損失および胚吸収などの評価項目に関する制約は発達神経毒性試験にも当てはまる。

生殖・発生毒性スクリーニング試験

42. 近年、より短期間で、より少ない資源を用いて生殖発生毒性を調べるスクリーニング試験が開発された。定義上、従来の試験と比較して、スクリーニング試験の規模は制限されている。

43. 1990 年代半ば、OECD が生殖毒性のスクリーニング試験に関する試験ガイドラインを導入した。「生殖・発生毒性スクリーニング試験」（TG 421）および「反復投与毒性試験と生殖・発生毒性スクリーニング試験の複合試験」（TG 422）には高生産量（HPV）化学物質に対して用いられるスクリーニング情報データセット（SIDS）の一部を含んでいる。TG 421 は規模が限られた一世代生殖毒性試験であるが、TG 422 は 28 日反復投与毒性試験と規模が限られた一世代生殖毒性試験との併合試験である。

44. 本試験の目的は、被験物質が雌雄の生殖能、すなわち生殖腺の機能、交尾行動、受胎、受胎産物の発生、分娩などに与える影響について限定的な情報を得ることである。ただし、既存の一世代および多世代生殖毒性試験、ならびに発生毒性試験に関する試験ガイドラインの代替法ではなく、それらを置き換えるものでもない。

45. 動物への投与は交配の 2 週間前に開始し、交配期間中継続する。雌動物については、分娩後 4 日の試験終了時まで投与を継続する。雄動物については、総投与期間が最低でも 28 日間となるまで投与を続ける。使用動物数は、各群雌雄各 10 匹以上とする。これにより通常、各群 8 匹以上の妊娠雌動物が得られると予測される。記録は妊孕能および分娩に対する影響について行う。生存児数、性別、体重について分娩後 24 時間以内と出生後 4 日に記録する。数ある指標の中でも、少なくとも最高用量群および対照群の動物の卵巣、精巣および精巣上体の詳細な組織学的検査を入れる。

その他の *in vivo* 試験

46. ここまで解説してきた毒性試験以外の *in vivo* 試験でも、正常な生殖を阻害する可能性のある化学物質の影響を明らかにすることができる（TG 407、408、409 など）。このため、がん原性試験など反復投与試験を含む毒性試験では、生殖腺および副生殖器を病理組織学的検査を含む病理学的検査に供する。

47. 米国国家毒性プログラム（NTP）が継続的な繁殖による生殖評価試験（RACB）を開発した。それによると、分娩前に処置した雌マウスを未処置の雄マウスと継続的に同居させ、一見して妊娠がわかるようになったら雌を離し、分娩後すぐに雄のケージへ戻すという継続的な繁殖パターンを確立して、累積出生児数を計数する（Chapin *et al.*, 1997、Lamb 1985）。この試験法によって、各世代の産児数を継続的に調べて、妊孕能および繁殖能に対する遅延性作用を監視することが可能となり、親動物の曝露期間の長期化とともに悪化するような作用、または対照群と比較してわずかだが生物学的に重要な連続的な産児数の減少が投与群で確認される可能性がある。

II. 出生前評価項目

評価項目の概要

実験動物を用いる試験

48. 化学物質曝露後の有害性評価については、出生前曝露後の周産期評価（成長遅延、構造異常および致死性）を取り入れた多くの試験ガイドラインが利用できる。これらには、出生前発生毒性試験（TG 414）、一世代生殖毒性試験（TG 415）、二世代生殖毒性試験（TG 416）、発達神経毒性試験（TG 426）、および二つの生殖・発生毒性スクリーニング試験（TG 421 および 422）が含まれる。

曝露の臨界期

49. 解説してきた試験ガイドラインでは、子宮内胎児に対する曝露は特定の時期に（たとえば、出生前発生毒性試験では、着床時から出産予定日の前日まで、または器官形成および性分化の主要期間）、または全妊娠期間を通して（たとえば、一世代および多世代生殖毒性試験、ならびに生殖・発生毒性スクリーニング試験では、受胎から出産まで）行うことになる。これらの広範な試験シナリオは、脆弱性のある臨界期の多点にわたって子宮内胎児を確実に曝露することにより、出生前曝露の有害転帰を広くスクリーニングできるよう設計されている。理論上は、供試動物の出生前曝露により発生に対する特異的有害作用が生じた日、または期間を特定することが可能であるが（Wilson 1965、Selevan *et al.*, 2000）、特定の有害作用を生じる曝露の臨界期を正確に特定できる十分な情報が、試験ガイドラインによって得られることはめったにない。単回曝露と反復曝露では生じる発育の結果が非常に異なるが、通常リスク評価における有害性評価時にこれらのデータを考慮する場合は、出生前発生段階に認められたあらゆる影響は、曝露期間のいずれの時点であれ、被験物質の単回投与により生じた可能性があるという安全側の推定がなされる。

潜在毒性

50. 発達期曝露による潜在（遅延）毒性については、発達神経毒性試験において、未投与期間（出生後約 21 日から 60 日まで）を経て神経行動学的評価と神経病理学的評価が行われる際に間接的に評価される。しかし、出生前発生毒性試験では、出生前に胎児の検査を行うため、発生に対する潜在的影響の評価はしない。多世代生殖毒性試験では、二世代目（および三世代目以降）の動物が子宮内および出生後発育期間を通して被験物質の曝露を受けるという点で特殊であり、潜在的な作用が成熟期（試験デザインにより限定された期間ではあるが）に発現しても検知できるよう若齢成熟期の大半にわたって毒性の評価を行う。しかし、出生前曝露による潜在性有害作用を、出生後や成長後の曝露、または成長の各段階にまたがる曝露の蓄積により生じた有害作用と識別することは著しく困難である。この問題に対処するために、同一試験内で、出生児における反応を連続した世代間で比較することによって限定的な情報を得ることがある。

方法論的問題点

黄体数

51. 母動物の卵巣については、剖検時に摘出して検査を行い、黄体数を計数する。黄体（CL）は、排卵後に卵胞の莢膜細胞と顆粒膜細胞によって形成される一時的な内分泌器官である。プロゲステロンは、CL が産生する主要ステロイドであり、卵子の着床と妊娠の維持に必要なホルモンである（Nelson and Gibori, 1993）。黄体数に関する情報は、受胎産物の生存能力についてのデータ解釈時に有用である（下記、着床前胚損失を参照）。黄体数の計数は、生の組織または固定組織を用いて、通常は解剖顕微鏡下で行う。卵巣組織の切開は必要最小限とする。計数の正確性は動物種間の大きさの差異の影響を受ける（マウスではやや困難である）が、技術的な問題は克服できないものではない。出生前発生毒性試験では児動物を出産する前に母動物の剖検を行うが、生殖試験では児動物が離乳する（出生後 21 日）まで母動物の剖検は行わないため、黄体数の正確な測定に関して技術的な課題が増えることになる。さらに、多経産の雌の黄体数は信用できない。すなわち、多胎妊娠後の雌動物では、前回の妊娠で遺残した黄体も計数される可能性があるため、妥当とは言えないのである。試験データの記録を再検討することで、1 腹の着床数が黄体数を上回っている例などでは、計数した黄体数が妥当でない可能性が示唆されることがある。そのような場合、該当する腹の黄体数、およびそれを元に算出した着床前胚損失については通常、群データの統計学的評価には含めない。

胚の着床前および着床後損失

52. 出生前死亡は、帝王切開時または分娩時の生存産児数の減少により肉眼的に明らかとなる。また、出生前発生毒性試験では、子宮内容物の詳細な検査によって早期および後期吸収率がわかり、前回の着床後に生じた子宮内死の証拠となる。出生前発生毒性試験の帝王切開時には、ごく初期の胚吸収が観察されることがあり、「着床痕」として記録される。生殖試験において授乳 21 日目に剖検を行う母動物の子宮でも同様である。肉眼的に着床痕が確認できなかった雌については、2 枚のスライドグラスの間に子宮組織を挟んで押しつぶすか、10%（体積比）硫化アンモニウムで染色する（Narotsky *et al.*, 1997）などして子宮の可視検査を進めていくことを推奨する。各腹の着床数を生存胎児数および死亡胎児数、または産児数と比較することで着床後胚損失の定量化が可能である。また、着床前胚損失については、各母動物の黄体数と着床数との比較により定量する。多世代生殖毒性試験で着床前損失数の増加を認める場合、配偶子の輸送または機能に対する有害作用によって受胎率が低下したものか、着床前胚に対する直接作用、または母動物の子宮もしくは内分泌系に対する作用の間接的影響によるものである。出生前発生毒性試験では交配・受胎後に投与を開始するため、着床前胚損失は後者によるもののみと推察される。着床前胚損失があったことを示すデータを解釈する際には、曝露変化状況と用量反応パターンについて考察することで有用な情報が得られることもあるが、具体的な毒性機序の同定は追加試験なしに成立しない。

胎児および出生児の外表検査

53. 帝王切開時または分娩時に目視による胎児または出生児の外表検査を行うことで、被験物質の子宮内曝露により生じた肉眼的な構造欠陥または運動性欠損の同定が可能となる。その際、出生後に死亡した児動物を見分けるための判断基準として肺浮遊試験などを行い、生存能力の評価を行う。各個体の体重測定とともに頭臀長を計測することもある。通常、げっ歯類ではこの時点で外生殖器の目視検査と、場合によっては肛門・生殖結節間距離の計測により性別判定を行い、後で（胎児の軟組織評価時など）再確認する。口蓋の完成および閉鎖については、各胎児の口を

優しく開けて内部を調べることで確認できるが、これについては後の軟組織検査まで保留しても良い。

胎児の内臓検査

54. 通常、げっ歯類の出生前発生毒性試験では、各腹の約半数の胎児を屠殺して、軟組織の固定および骨格構造の脱灰を促進する固定液に浸漬する。各胎児の検査は連続切片法によるか、または臓器の肉眼検査（固定前または固定後）と選抜した臓器の切片検査とを組み合わせて行う。ウサギの胎児については、比較的大きいために連続切片の作製が容易ではない。そのため、ほとんどの試験機関では各胎児の剖検を詳細に行って、特殊な方法による頭蓋の検査（冠状断切片を用いた脳、眼および鼻腔の内部構造の可視化など）を行うとともに、胸部および腹部器官の構造と完全性を評価する方が好まれる。この手順はげっ歯類でも用いることができ、それによって100%の胎児で軟組織の異常に関する情報が得られる。また、特殊な方法（Staples, 1974 など）を標準試験法として用いたり、または標的となることが分かっている臓器の発生に関わる変化についてさらに追究するための試験として用いることもできる。胎児組織の病理組織学的評価はガイドラインの標準的な試験手順に含まれていないが、観察された異常、または予想される異常の追加評価に有用であろう。

骨格検査

55. げっ歯類の出生前発生毒性試験では通常、各腹の約半数の胎児（すなわち、軟組織検査に割り当てられなかった胎児群）を骨格検査に供する。連続切片法を用いないウサギ（ウサギの場合、胎児の 50%は頭部が付いた標本で、50%は頭部のない標本で内臓検査を行う）については、軟組織検査終了後に全胎児について骨格異常を調べる。適切に処理を進行させるため、ウサギの胎児の内臓を摘出し、剥皮を行い、肩部に過剰な沈着脂肪があれば、それを取り除く。この手順はげっ歯類でも用いることができ、それによって 100%の胎児で骨格の評価が可能になる。次に、胎児（げっ歯類、非げっ歯類ともに）を脱水性の固定液（エタノール、アセトンなど）に浸漬し、水酸化カリウム溶液中で軟化させてから、カルシウムに特異的なアリザリンレッド S で染色し、軟部組織の余分な染色剤を除去するためにグリセリンを基剤とした透徹剤に浸漬する。胎児の軟骨の評価を行うことを推奨する。これにより、骨格構造の骨化に関わる軟骨前駆細胞について有益な情報を得ることができる。軟骨の評価に際しては、二重染色法（胎児の処理に、軟骨特異染色であるアルシアンブルー染色などの透徹前処置を加えたもの）、またはアリザリンレッド S の分染による色調勾配を用いた代替手法を用いる。二重染色法はげっ歯類を用いた試験で用いられるが、ウサギを用いた試験では通常は用いない。染色後、十分透徹し、胎児の頭蓋、胸部、骨盤部および体軸の骨格構造について異常の有無を調べる。各胎児について、骨化の範囲、正常な構造形成（形、大きさ、完全性）、体位および関節を評価する。

データ解釈

奇形と変異

56. 通常、胎児および出生児の検査により得られた所見については、奇形と変異に分類する。奇形の一般定義は、生存、発育、または機能に有害な影響を与える恒久的な構造変化であり、変

異の一般定義は、生存または健康に有害な影響を与えることのない、構造の通常範囲からの逸脱であるとされている（US EPA, 1991、Chahoud *et al.*, 1999、Solecki *et al.*, 2001, 2003）。奇形と変異の分類方法について一般的な合意はないが、その理由として一つには、正常な発達と異常な発達の間に反応の連続性が認められるためである。また、特定の所見が、ある動物種では奇形に、別の動物種では変異に分類される場合があること、または検査実施時の妊娠日数によって分類が変わることも知られている。さらに、発育遅延による変異と構造変化と考えられる変異とを区別するなどして、これらの定義をさらに細分化する試みを行っている試験機関もある。以上のことから、奇形と変異の定義については、試験機関ごとに、各試験報告書の文中で慎重かつ明確になされなければならない。

専門用語の国際統一

57. 国際先天異常学会連盟（IFTS）の発生毒性分野における学術用語の国際協調に関する委員会（Committee on International Harmonization of Nomenclature in Developmental Toxicology）が、国際合意が得られた胎児および出生児の形態所見を記載する際に用いる一般的な学術用語の用語集を作成し、公表した（Wise *et al.*, 1997）。この取り組みの目的は、専門用語の調和を推進し、特に国をまたがって規制当局へ提出する際に生じうる、発生影響に関する説明の混同と曖昧さを減らすことであった。その後、1998 年から 2007 年の間にベルリンで専門用語学会が数回にわたって開催され、その結果、文献（Chahoud *et al.*, 1999、Solecki *et al.*, 2001, 2003）およびインターネット上（http://www.DevTox.org）に公表された。外表、内臓および骨格検査において国際統一をみた用語を熟知しておくこと、データの収集、報告および再検討時に適切な用語を用いることが奨励されている。しかし、この取り組みによって作成された一般的な用語体系が広く利用可能となり、国際的合意を得てもなお、専門用語の使われ方が化学物質の有害性評価を実施している全試験機関で等しいとは限らないことも認識されている。

評価項目の相互関係

- 母体毒性

58. 通常、出生前毒性の評価を目的とした試験は、ある程度の母体毒性を誘発する群が少なくとも 1 用量群は含まれるよう計画される。母体毒性に関する評価項目はプロトコールの規定にもよるが、有症率または死亡率（これらを誘発しないように投与量を設定しても生じうる）、ならびに妊娠期間、生化学検査および血液検査項目、一般状態、体重、体重変化量、摂餌量または飲水量、臓器重量、剖検時所見および病理組織学的データにおける変化などがある。母動物の状態とその出生児の状態には高い相関関係があり、非常に毒性の高い投与レベルでは特に顕著に表れる。ただし、ほぼ例外なく、毒性物質への胎児の直接曝露に起因する子宮内発育に対する作用と、母体毒性に起因する、または母体毒性によって悪化を受ける作用とを完全に識別することは不可能である。その理由の一つは、このような試験では母体毒性の評価が限られているためである（ECETOC, 2004）。発生毒性の要因としての母体毒性の潜在的役割を解明するためには、さらなる作用機序研究が必要となる。たとえば、Daston（1994）や Hood および Miller（1997）が行った研究などである。発育中の個体に対する有害作用は、その原因が何であれ、投与による毒性徴候であることには変わりない。したがって、母体毒性が認められたからといって、同じ投与レベルでの胎児毒性所見の価値が自動的になくなってしまうわけではない。また、ラットおよびウサギを用いた最近の給餌制限も関する二つの研究（Fleeman, 2005 and Cappon, 2005）では、重度の体重減少または体重増加量の低下によって、ラットでは骨格の発達に軽微な変化が認めら

れたものの、生存や奇形発生には影響がないことが明らかにされた。ウサギでは、給餌制限が最も厳しかった群で流産する例があったが、奇形は認められなかった。しかし、母体毒性が試験データの解釈に影響している研究もある。たとえば、胎児の健在（well-being）が損なわれるほど母体毒性が非常に重度である（死亡発生など）場合は、発生影響に関するデータの解釈は困難となる。

• 死亡率および奇形発生率

59. 出生前発生毒性試験の試験デザインでは、生殖試験で見られるような母動物の喰殺による奇形胎児の損失を確実に防ぐため（Schardein *et al.*, 1978）、出産予定日の前日に母動物を屠殺する。しかしながら、出生前発生毒性試験によって出生前死亡の起因や原因（source or cause）が研究者に突き止められることはない。子宮内死亡は、生存を維持できない奇形が原因で生じることがある。しかしながら、その中で発育異常の証拠となりうるのは発育後期に死亡した胎児だけである（浸軟の進んでいない後期吸収胚または死亡胎児の評価による）。これを識別することは可能ではないかもしれないが、一方では、奇形、吸収（早期および後期）および死亡（帝王切開時に満期の状態で死亡していた胎児）の腹当たりの合計発生率を算出し、群データを適切に統計解析することによって、同腹児生存率への影響全体に対する胎児奇形の寄与を適切に分析することができる。

60. 出生前発生毒性試験では、使用動物数が比較的少ないため、奇形のように頻度の低い事象については感度に限界がある。通常用いる妊娠ラット数は 1 群 20 匹であるが、高用量投与群を設けるか、被験物質が高い胚・胎児毒性を有していない限り、主要な奇形の増加を検出できない（Palmer 1981）。このため、化学物質の発生毒性を評価する際には、軽度異常、変異、胎児の死亡および発育など、他の発生影響に関する情報を取り入れることが重要である。さらに、主要器官形成期以降に発達する生殖器や脳などの奇形については、過去の催奇形性試験のガイドラインに従って実施された試験では検出できないことがある。たとえば、妊娠後期の雄性分化期間中の曝露によって雄性生殖器系の形態および機能を変化させる発生毒性物質には、フタル酸ジブチル（Mylcreest *et al.*, 1999）、弱いアンドロゲン拮抗剤であるリニュロン（McIntyre *et al.*, 2002）、抗アンドロゲン作用を有する除草剤ビンクロゾリン（Gray *et al.*, 1999）がある。そのため、改正後の TG 414 では、少なくとも着床から屠殺予定日の前日、すなわち出産予定日の 2 日前まで投与が行われるよう投与期間が延長された。（ただし、この投与期間の延長により、被験物質が内分泌攪乱物質である可能性を示唆する変化すべてを検出できることを保証するものでも、陰性結果が内分泌撹乱作用のないことの確かな証拠となるものでもない。）投与期間については、被験物質のデータから奇形の評価に交絡を生じるような着床前胚損失の発生が示唆されなければ、着床前期間に生じる作用についても調べるために、屠殺予定日までの全妊娠期間を含めるべきである。

• 同腹児数および平均胎児体重

61. 各胎児の体重および同腹児の平均体重に対する同腹児数の影響については、これまでによく分かっている。一般的に多胎動物では、胎児または出生児の体重は同腹児数に逆相関し、用量反応曲線の上端は同腹児数が少なく、胎児または出生児の体重が大きいほど影響を受けやすい。未投与群の平均生存産児数は狭い範囲内に収まる場合が多いが、同腹児数については、広範な変動が普通にみられる（MARTA and MTA, 1995, 1996）。同腹児数の増加傾向は長い時間をかけて認められるようになってきたが、同腹児数が多いものを選択するような繁殖計画がその一因で

ある。帝王切開時または分娩時の同腹児数は、自然変動だけでなく、生殖細胞毒性、着床不全、胚死亡など、出生前の化学物質への曝露の影響も受ける。胎児または出生児の平均体重について統計分析する際には、共変量解析などを用いて、同腹児の体重を同腹児数で統計的に補正しなければならない。

• 骨格変異の発生率に対する影響因子

62. 骨格検査に割り付けられた各胎児について、 死亡時（すなわち、出産予定日の約 1 日前に予定されている母動物の帝王切開時）の骨格の骨化度を測定する。骨格は標準的な時系列に沿って規則正しく発達していく（Spark and Dawson, 1928）。それでもやはり、動物の系統の違いなどによって、骨化過程に正常変動が認められることもある。このため、帝王切開を行う時期については、科学的に正当な妊娠期間とすることが非常に重要である。さらに、骨化遅延（通常、骨格変異に分類）の発生率が胎児の在胎期間に直接関係するものであって、投与による有害作用ではない場合もあるため、帝王切開の実施日については、対照群および投与群ともにできるだけ管理することが重要である。日程の概要としては、試験に供している各母動物について、正確な交配時間を確認し、帝王切開の実施はほぼ同等の妊娠期間（時間単位）となるよう行うことになっているが、ガイドラインに従って出生前発生毒性試験を大規模に行っている試験機関では実際的でない場合がほとんどである。それでもやはり、潜在的な偏りを生じないような予定を立てることが重要である。

63. 胎児の骨化の程度はある程度胎児の大きさに関係し、同一投与群中でも大きな胎児（同腹児数の少ない腹）と比較して小さな胎児（同腹児数の多い腹）の方が骨化遅延の発生率が高い。骨化データの統計学的評価時には、胎児の体重で補正するか、少なくともデータ解釈に胎児体重についての考察を入れるとよい。

• 統計学的評価

64. 胎児および出生児のデータについて統計解析を行う際には、試験デザイン、解析中の評価項目、標本の大きさ、性別の影響、解析結果に対する腹ごとの影響因子について慎重に考慮しなければならない。統計解析時に個々の同腹児を独立観測データとして処理しないようにすることが重要である。これに関して、Holson および Pearce（1992）は体重データの解析で、1 腹から 2 匹の同腹児を独立観測データとして扱っただけで、危険率 α 値（訳者注：第 1 種の過誤に対応する確率）をほとんど 3 倍に設定したのと同様となり得ることを示している。適切な統計的手法の選択とデータ解析のために、各腹のデータを適切に解析する方法に精通している統計の専門家に相談することを推奨する。統計学的解析方法について細かく論じることは本ガイダンス文書の範囲を超えているが、本章、および次の出生後評価項目に関する章では、懸念される具体的な問題について取り上げる。

65. 出生前毒性データの評価時には、試験の検出力、すなわちその試験によって真に存在する影響が実際に示される確率が重要となる。統計的検出力に影響しうる要因として、その試験の標本の大きさ（1 腹を解析の基本単位として）、所見の自然発生率、その評価項目の発生率の変動、データの頑健性および解析方法がある。各投与群あたり 20 腹を用いた出生前発生毒性試験で全胎児を評価した場合、統計学的に検出可能な最小変化量は以下である。

- 奇形発現率については、対照群レベルの 5 倍から 12 倍の増加。

- 子宮内死亡率については、3 倍から 6 倍の増加。
- 胎児体重については、0.15 倍から 0.25 倍の減少（US EPA, 1991）。

多世代生殖毒性試験における F1 児および F2 児の形態構造異常の検出については、評価する腹の数だけでなく、各評価項目で検査対象となっている腹ごとの同腹児数、および出生児間における作用の類似度にも依存することが指摘されてきた。出生児間で作用が全く同じというわけではないため、発生毒性試験で実施されているように、各腹の全出生児を調べて奇形の有無を確認することに統計学的な価値がある（検出力向上）。検査対象となる F1 世代の 1 腹当たりの児動物数が多ければ、低用量での作用の検出力が非常に増強される。腹ごとの平均値を解析する場合であっても、1 腹につき雌雄それぞれ 2 匹以上の児動物を検査できれば、統計精度が向上する（F 統計量の計算に用いる平均二乗誤差が小さくなるため）。一般的に多世代生殖毒性試験では、「同腹効果」の大きさは全評価項目で等しいわけではなく、投与量によっても異なり（作用の表出がより明らかになる高用量群になるほど効果大）、また、各臓器に対する同腹効果も化学物質の作用機序によって異なる。

66. 投与関連作用の生物学的意義の妥当性を立証する上で、統計的有意性は不要である。このことは特に発生率の低い所見（すなわち稀な奇形）、もしくは変動の大きい所見、または同時対照データに発生率の異常が認められる状況下に関して当てはまる。それと同様に、統計的有意性が必ずしも生物学的意義を示すわけではなく、偶発的所見と真の所見とを見分けるためには、科学的判断および関連のある背景データを用いなければならない。

• 同時対照群／背景対照データ （Concurrent/historical controls）

67. 試験では毎回同時対照データをとることが求められる。一方、背景対照データは、通常良く解析されている複数の試験の陰性（溶媒）対照群のデータで構成されるもので、必須というわけではないが、利用可能である場合も多く、試験時の所見の解釈に有用で適切であると考えられている。試験の投与群動物から得られたデータについては、背景対照データとの比較よりも、同時対照群から得られたデータとの比較を優先すべきである。背景データを利用するのであれば、実験動物群内の遺伝的浮動を避けるために、同一試験施設内で、試験データの解釈を行うより前の妥当な期間内（± 2 年間など）に、同じ試験条件下（動物種、系統、供給元、週齢、溶媒、投与経路、投与期間、担当技術者などが同じ）で実施された試験のデータを用いるのが最適である。背景対照データには、現在実施している研究の文脈において意味のある情報が十分に含まれていることが重要である。たとえば、用語の定義付けがなされていること、付随データおよび連続データの特性化が完全になされ、適当なデータ範囲、最大値、最小値、中央値および平均値を用いて要約されていること、データの分散について説明がなされていることが必要となる。動物供給者により蓄積された背景データ、または複数の試験機関の調査を通じて集められた（遺伝的浮動を含む）背景データ（Clemens *et al.*, 1994, MARTA and MTA, 1995, 1996）も場合によっては有用であるが、適切な注意のもとで用いられるべきである。総じて、背景データの解釈および利用時には慎重に検討することが必要であり、科学的判断および専門知識を要する。万が一、同時対照データが背景データと明らかに違う場合には、その試験には致命的欠陥がある可能性がある。最も極端な例では、投与群のデータの解釈時に背景データを用いることが適当でないことも考えられる。

III. 出生後評価項目：出生児の成長、発育指標及び機能／行動に関する神経毒性

評価項目の概要

68. 出生後の発育に関する評価項目を取り入れた多くの試験ガイドラインが利用可能である。一世代および二世代生殖毒性試験（TG 415 および 416）、生殖・発生毒性スクリーニング試験（TG 421 および 422）、発達神経毒性試験（TG 426）などである。

69. 各評価項目とそれを含む試験ガイドラインについて表 2 に示す。

表 2. 各種試験ガイドラインと評価項目

評価項目	TG 415	TG 416	TG 421 および 422	TG 426
出生児体重	+	+	+	+
生存率（周産期）	+	+	+	+
生存率（授乳期）	+	+	-	+
生存率（成熟期）	-	+	-	+
成長（周産期）	+	+	+	+
成長（授乳期）	+	+	-	+
成長（成熟期）	-	+	-	+
身体的発達（性成熟）	-	+	-	+
機能的発達	-	(+)	-	+
行動	-	(+)	-	+
神経病理	-	(+)	-	+
生殖機能	-	+	-	-

＋：必須、（＋）：任意

方法論的問題点

同腹児数の標準化

70. 世代試験および出生前試験では、いわゆる「間引き」により、出生後数日目に児動物を無作為に最大 8 匹（雌 4 匹、雄 4 匹）まで選抜淘汰してできるだけ同腹児数を標準化する操作が用いられることが多く、その手順については試験ガイドラインに記載されている。同腹児数の調整は試験ガイドライン 416 では任意であるが、試験ガイドライン 426 では使用動物の平均的な産児数に近くなるよう同腹児数を調整することが推奨されている。

71. 同腹児数の標準化に賛成する意見としては、児動物の体重が同腹児数に関係するため、離乳時の児動物の体重をより均一化できるというものである。約 500 腹のデータを用いた試験では、同腹児数の標準化によって、体重の変動は変えず、同腹児の平均体重だけが増加したように思われた（Palmer 1986）。一方で、間引きにより 25～40%もの児動物が除外されることになり、たとえば無作為抽出により矮小児の除外が生じた場合など偏りが生じかねない（Palmer 1986）。

72. 児動物の成長率および成熟率は同腹児数によって異なるが、同腹児数の差異が行動にも影響を及ぼすか否かについては、はっきりと分かっていない（Barlow and Sullivan 1975）。同腹

児数の少ない児動物と比較して、同腹児数が多い児動物では行動に変化が認められたという文献もあるが、差異が認められなかったという報告もある（Lore and Avis 1970, Barlow and Sullivan によりすべて引用 1975）。いずれの試験でも、同腹児数が 2、3 匹であれば、なんら有意差が認められていない。

73. 同腹児数の標準化を行った試験であっても、出生時の同腹児数は出生後の成長に関係してくる。通常、統計学的評価に際しては産児数を考慮に入れなければならない。

曝露経路

74. 一般的に、生殖期を含む試験パラダイムでは、供試動物への曝露方法に様々なものがあるばかりでなく、児動物が実際に受ける曝露の不確実さも様々である。通常、詳細データ（被験物質の組織内濃度など）がなくても、出生前の動物または未成熟動物に有害な投与関連反応が認められることによって用量選択の妥当性が実証されたことになる。また、多くの場合、投与の影響が児動物に認められなくても、児動物が妊娠期間中（母動物の循環系を通じて）および授乳期間中（母乳を通じて）に被験物質の曝露を受けたと仮定する。どの曝露経路でも、多くの場合、被験物質への児動物の直接曝露はある程度生じる。たとえば、授乳期間 3 週目になれば、児動物は被験物質混餌飼料を食べ始めるため、被験物質の実質摂取量（mg）が体重（kg）比にして母動物よりも高くなる場合もある。混餌投与による場合、母動物と同居飼育された同腹児は真の独立集団とみなすことはできない。このことが用いる動物数（F0）および児動物数（F1）に関係してくる。同腹児を曝露チャンバーいれて曝露を行う吸入試験では、児動物は吸入投与と経皮投与を同時に受けることになる。混餌または飲料水による投与では、出生時期および新生児期を通した連続投与が容易である。児動物は生後 14 日頃から飼料および飲料水を摂取し始めるため、授乳期間後期の曝露については、母乳を介した間接曝露と直接曝露の両者ということになる。また、強制経口投与であれ非経口投与であれ、出産の前後には、1 日ないし数日投与を延期しなければならないことがある。

75. 発育の臨界期に十分な用量を確実に児動物に曝露する必要がある場合（発達中の神経系または免疫系における作用を評価する場合など）、または児動物の曝露量を定量する必要がある場合には、離乳前の特定の期間に児動物に直接投与することを考えなければならないことになる。適切な訓練を受けた者により正しく投与されれば、児動物にとってストレスとはならない。ただし、そのような操作が毒性反応およびデータの解釈へ与える影響については、慎重に考慮すべきである（ILSI, 2003、Moser *et al.*, 2005）。

76. 通常、吸入試験では、母動物および児動物の双方が曝露される。児動物から母動物を 1 日 6 時間だけ引き離しても、最初の日に可逆的な軽度の体重減少が生じるだけであるという報告（Pryce and Feldon, 2003）もあり、母動物または児動物を分離して、片方のみの吸入曝露を実施することも考えられるようになってきた。しかし、児動物から 1 日 6 時間母動物を引き離すことで、母性剥奪の動物モデルで認められるような生理学的変化が児動物に顕在化することがあるかもしれない。このため、分離吸入曝露試験を是とする前に、母動物単独で吸入曝露を行うことについて、その結果と実施上の問題点を明らかにするための研究が必要である。分離吸入曝露を採用するのであれば、母動物を児動物から引き離さない対照群の追加を推奨する。

77. 生殖毒性試験の曝露経路として経皮投与を用いることは推奨しない。経皮投与による生殖毒性試験に関しては、ヒトにおける通常の曝露を正確に映し出しているという利点よりも技術的問題の方が大きい。必要であれば、吸収・分布・代謝・排泄に関する試験（ADME 試験）など他の試験を実施して、経口経路から皮膚経路へ外挿する。

クロスフォスタリング（里親哺育法 cross-fostering）

78. 出生前および出生後投与による出生後の影響は、児動物における出生前もしくは出生後、またはその両者に障害としてあらわれる。発育上の変化は、出生前および出生後の発育中の動物に対する化学物質の影響によるばかりでなく、母動物に対する影響を介して生じることがある。たとえば、授乳または母動物の管理が影響を受ければ、母動物の生理または行動の変化が児動物の行動に影響を与える可能性がある。この点の調整を図るため、里親哺育法（cross-fostering）を用いて、出生前曝露を受けた児動物と非曝露群の母動物、またそれぞれの母動物および児動物とを交差するように組み合わせて哺育させる実験手法が開発された。この試験法には明らかにより多くの動物および資源が必要であり、それゆえスクリーニング試験としては用いられない。しかし、追跡試験など、初期スクリーニング試験で認められた作用についてさらに調べるような場合には有用である。

児動物の日齢

79. 児動物の日齢には、出生時から計算した出生後日齢、または交配時から計算した交尾後日齢を用いる。特に妊娠期間に群間差異が認められる場合、離乳前の発達指標が発現する日齢の予測因子としては、出生後日齢よりも交尾後日齢の方が良いという研究結果がある（Hughes 1986, Raimondo and Draghetti 1990）。交尾後日齢は発達神経毒性試験で用いられた経緯がある（Goodlett *et al.*, 1987, Kelly *et al.*, 1988, Hass *et al.*, 1994a, b, Hass *et al.*, 1995）。例えばWistar ラットでは妊娠 22 日が出産予定日となるが、全出生児について（実際に娩出された日に関係なく）この日を出生後 0 日とするような場合、これは交尾後日齢を用いていることになるが、「日齢」自体はやはり出産予定日に絡めて数えられる。

検査を行う時間帯

80. 行動学的検査における個体の反応は、検査の実施時間帯による。ラットは夜行性動物であるため、正常な睡眠時間帯に動物を覚醒させると正常な覚醒時間帯に行った検査結果とは異なる行動を誘発する可能性がある。このため、夜行性動物の検査は夜間に行うか、より利便性の高い方法として、飼育室の照明の時間帯を逆に設定して、日中に赤色照明下で検査を行うと良い（Barlow and Sullivan 1975）。いずれの条件下でも、頑健なデータを得るために最も重要なことは、同一条件下で試験を行うことである。

児動物の体験事象

81. 幼齢期に児動物が体験した事象はその後の行動に影響を与える。たとえば、幼齢期にラットを頻繁に手で触ると、ストレスに対する生理反応（Levine *et al.*, 1967, Meaney *et al.*, 1988, 1991）および情動性試験および学習試験時の行動に変化を生じる（Levine and Broadhurst 1963, Nunez *et al.*, 1995, Meaney *et al.*, 1988, 1991）。環境エンリッチメントを用いて飼育された動物では、行動変化だけでなく、大脳皮質重量の増加、コリンエステラーゼ活性の上昇および脳内モノアミン濃度の変化も認められる（Rosenzweig and Bennett 1969）。

82. 幼齢期の環境的体験を管理するために、児動物の飼育条件については、室温、湿度、騒音レベル、照明、飼育ケージ、個体の取り扱いおよびケージ洗浄などの可変因子に関して試験法の範囲内で標準化すべきである（Barlow and Sullivan 1975）。環境エンリッチメントも含めた試験デザインの変更による影響については、背景データとの比較、また関連性があれば陽性対照試験による評価を行わなければならない。

児動物の性別

83. 児動物の行動に性差があることは、よく知られていることである。したがって、発達神経毒性試験では雌雄ともに検査を行って、結果の解析時には雌雄別々に解析を行うことが重要である。さもなければ、作用が隠れて表面化しない可能性がある。

行動学的検査結果に影響を及ぼす試験担当者側の因子

84. 試験担当者の期待が行動学的検査結果に大きく影響する場合があることが、研究によって分かっている（Rosenthal and Fode 1963）。試験担当者の影響を完全に除外するためには、試験担当者に動物が属する群を知らせるべきではない。このような試験の施行には、試験担当者がデータの解析および評価に関与する時期などに実際的な問題が伴う。それでも、実験者による先入観を回避するためには、検査に際して投与群について何も知らない方が良い。

試験の自動化

85. 行動学的検査を自動化することで実験者による先入観と必要な人的資源が減少し、また 1 実験から得られる情報量が増加して、実験結果のより良い解釈のための詳細なデータ解析が可能となる。自動機器の利用は OECD 試験ガイドライン 426 の発達神経毒性試験における運動量測定で求められている。各種行動学的検査用の測定装置が市販されている。しかし、自動化したデータの収集では行動の種類を分類する能力は低下する。たとえば、ビーム遮断を利用して活動性を測る自動測定装置を用いた場合、「首振り行動」を示している動物の首の動きによって単一ビームが何度も遮断されれば、それは移動行動量として計数される可能性がある。一般的に、行動試験を自動化するにあたっては、収集されるデータの解釈に影響しかねないような行動について慎重な検討が行われなければならない。

行動評価の対象となる試験群について

86. 試験ガイドライン 426 では、出生後の発達および行動の評価を行うため、一世代および二世代生殖毒性試験と同じ群構成、すなわち、3 処置群と 1 対照群が含まれている。しかし、明らかな発生毒性（同腹児数の顕著な減少、または出生後死亡数の増加など）が観察される群では、検査対象となる出生児数が足りない、または生存児動物の健康状態が不良であるために検査を行えないなどの事情から、行動学的検査の実施が不可能になる場合もある。母動物または児動物に重度の毒性を認める場合は行動学的検査は必ずしも必要ではない。

身体的発達の指標

87. 出生児の身体的発達については通常、離乳期前の数回にわたる体重測定、および離乳後には 1 カ月に 1、2 回行われる体重測定によって観察される。

88. その他の身体的および機能的指標については、いわゆる発達指標、または発達の目安を記録していくことで追跡していく。これらの観察は、「起こるか起こらないか」ではなく、「いつ起こったか」を表す場合が多く、各種発達指標が最初に認められた時期を確認して、試験の特定の評価項目における発達の時間的経過の遅延または早期化の確認に用いられる（Lochry *et al.*,

1986）。通常、このような検査では、発達指標が発現するおよその予想日、またはそれ以前の日から始まって、連続的に各指標の有無について評価を行う。

89. 身体的発達の指標として提示されることの多い項目としては、耳介展開、被毛の発生、上下切歯の萌出、眼瞼開裂、全身の毛生および性成熟の開始がある。これらの指標が異なる観察者によって観察された場合の結果の信頼性については、Hughes および Palmer（1986）によって評価されている。その結果、3 つの身体的項目、すなわち耳介展開、上下切歯の萌出および眼瞼開裂については信頼性があるとみなされ、評価項目として用いられることになった。通常、発達指標の変化は体重の増加と関連しているため、体重についても身体的発達の指標として同時に記録していく。

90. 試験ガイドライン 416 では、F1 の性比や性成熟の時期の変化から必要と考えられる場合には、F2 児について生後 0 日に肛門・生殖結節間距離（AGD）を測定する。AGD は、雌よりも雄の方が長い。ホルモン、または内分泌攪乱作用が疑われる化学物質によって AGD が変化することを示した研究がいくつかある（Gray *et al.*, 2001, McIntyre *et al.*, 2001 など）。さらに、雄の出生児における AGD の減少は、その後の尿道下裂および停留睾丸など、他の作用の予測因子となりうることが示されている。このため、内分泌攪乱作用が予想または疑われる化学物質について試験を行う場合は、F1 世代の AGD 測定も考慮する。F2 児については、離乳時に安楽死させるが、F1 児については、出生時の AGD を測定した場合、成熟期に測定した値との相関性を検討する。AGD の永続的変化が引き起こされていれば（すなわち、出生時および成熟期ともに変化が認められれば）、永久的な構造変化が示唆される。AGD の測定はノギスを用いるか、または測定器具付きの実体顕微鏡下などで行う。肛門中心から性器突起中心までというように明確な測定基準を確立しておくことが重要である。さらに、動物の取り扱い時に、一部の動物では他の動物よりも測定部位が伸長するなどして、指標に変動が生じることがないよう慎重に測定作業を進める。AGD は児動物の体重と関連性があるため、統計解析時に AGD を共変量として用いる場合には、体重に対する AGD の補正方法を標準化しておかなければならない。

91. ラット出生児の雄の乳頭・乳輪遺残の評価については、現行の OECD 試験ガイドラインの測定項目として含まれていないが、本指標はホルモンおよび特定のホルモン攪乱物質（特に抗アンドロゲン作用物質）への曝露によって変化する（Gray *et al.*, 2001, McIntyre *et al.*, 2001）。さらに、乳頭・乳輪遺残は、尿道下裂および停留睾丸などの予測因子となりうることが分かっている。このため、内分泌攪乱作用が予想または疑われる場合には、F1 世代の評価項目に本指標を含めると有用である。F2 児については、離乳時に安楽死させるが、F1 児について出生時検査を行った場合には、成熟期に測定した値との相関性を検討する。雄動物に永続的に乳頭が存在すれば、永久的な構造変化、すなわち、奇形ということになる。正常な雌の児動物では、生後約 13 日から 12 個の乳輪が認められるようになるが、雄の児動物では乳輪がごくわずかに認められるか、全く認められない。児動物の被毛の発生により乳輪の確認が困難、または不可能になるので、試験に用いる動物の観察時期を適切に設定することが重要である。雄の乳輪については、1 匹あたりの数ではなく、有無のみ記録される場合が多い。ほとんどの試験では対照群の雄ラットにおける乳輪の確認率は 5%未満と低く、そのような場合には有無による評価で検出感度は十分となる。ただし、対照値が 30%近くになる例も報告されており（Hellwig *et al.*, 2000）、そのような場合には有無による評価だけでは検出感度が十分でない。特に対照値が高値である場合には、各雄の乳輪数を評価に用いることを推奨する。

92. 試験ガイドライン 416 および 426 には*性成熟*の評価が含まれている。雌の性成熟の開始については、腟開口の確認によって検査を行う。ラットの場合、腟開口は生後約 30～35 日の間に認められる。生殖毒性試験ではラットが最も一般的な動物種であるが、マウスを用いる場合、腟開口の日齢は性成熟の指標とならないため、検査者は最初の発情期の日齢を用いなければならないことを念頭に置く必要がある。雄では通常、陰茎亀頭と包皮の分離を性成熟指標とする。精巣下降の検査に関しては、被検査動物の取り扱い方次第であり、評価が難しい。陰茎亀頭および包

皮の分離日齢は雄ラットの性成熟日齢と合致しているため（Korenbrot *et al.*, 1977）、試験ガイドライン 426 および 416 の評価項目として採用されている。ラットでは、生後約 40～45 日の間に認められる。性成熟時の日齢について、体重とともに記録する。雄および雌の性成熟に関する記録については、慎重に評価する必要がある。このため、被験動物の検査が毎回同じ判断基準で行えるようになるまでには、訓練を要する。また、検査は毎回同一検査者が（または数人で）同一判断基準によって実施することが望ましい。

神経行動学的評価法概論

93. 発達神経毒性試験（OECD TG 426）には、初期行動発達、慣れを含む自発運動量、運動および感覚機能、ならびに学習および記憶の評価が含まれている。これらの神経行動学的評価項目の評価法および解釈に関しては、多くのガイダンス文書が作成されている（OECD 2004、IPCS 2001）。

94. 米国で開催された研究会において、鉛、メチル水銀、いくつかの乱用薬物、フェニトイン、PCB、エタノールおよび電離放射線を対象に、ヒトと実験動物のデータ間に、定性的および定量的な同等性（comparability）がどの程度あるかが検討された（Francis, 1990）。定性的比較については、運動発達および運動機能、認知機能、動機付け行動／覚醒行動、感覚機能、及び社会的行動に関して評価された。異種間の同等性については、多くの制約条件が同定されたたものの受容可能な程度の同等性があるとされた（Francis *et al.*, 1990）。定量的に厳密な比較を行うことについては、用量と評価項目との関連性を検討した結果、当時は可能ではないとされたが、検討された物質のいくつかについて、認知機能が最も感度が高いことが示唆された。用量反応データは、特にヒトでは不十分である場合が多い。有効投与量の比較では種間の差が大きかったが（最大 10,000 倍）、用量として体内濃度測定値（血中濃度、脳内濃度など）を用いた比較では顕著な相関関係が認められた（平均差 1～2 倍）。発生毒性の他の評価項目については、本研究会で得られた知見は、動物試験で機能的な変化が認められた場合、その作用因子はヒトの発生過程にも影響を与える可能性を示唆するという仮説を支持するものであった。

95. OECD の発達神経毒性に関するガイドライン（TG 426）では、実施する具体的な行動評価の選択に柔軟性が持たされており、多くの場合、明記されているのは検査の対象となる機能および検査を完了する日齢のみである。このため、各試験の処置に関連した変化の検出感度および対象とする機能の検出に関わる特異性について、試験施設ごとに文書記録しておくことが必要不可欠である。検査実施手順については、行動評価分野に詳しい者が計画と実施を担当し、各検査に携わる技術者は使用される個々の測定手順を実施できるよう訓練を受けておかなければならない。これらの評価を行うにあたり、動物の取り扱い、機器洗浄の時期および方法の変更、試験環境（照明または騒音レベルなど）の変更など手順細部の変更によって被験動物の行動が大きな影響を受ける可能性があるため、SOP を忠実に遵守する必要がある。検査手順の変更により、検査結果の変動性が上昇し、被験物質の行動影響の検出感度が低下する可能性がある。

96. 幼齢期の体験は特にその後の行動に対する影響が大きい。たとえば、幼齢期にラットを頻繁に手で触ると、ストレスに対する生理反応、および情動性試験および学習試験時の行動に変化を生じることがある。幼齢期の環境的体験を管理するために、児動物の飼育条件については、騒音レベル、個体の取り扱いおよびケージ洗浄などの可変因子に関して試験法の範囲内で標準化すべきである。行動学的検査の個体の成績は検査時間帯、個体のストレスレベルなどの影響を受ける。このため、対照群および投与群について検査が同時に行われ、環境条件が標準化されている試験のデータが最も信頼性の高いデータということになる。

97. 試験機関が行動学的検査を新たに取り入れようとする場合、新たな検査法を取り入れる前に対照動物の検査結果について評価を行い、検査回ごとに、検査成績（performance）のベースラインが、妥当な範囲で安定しており、変動が許容範囲内であることを確かめなければならない。さらに、各試験機関は、使用するほとんどの検査手順について陽性対照物質を用いた試験を行い、行動における処置関連変化を検出する能力を有することを実証しなければならなす、これには行動学的検査成績が低下する場合と上昇する場合の両者について、検出力を示すための試験が含まれる。行動学的検査手順に変更を加える場合、背景データおよび陽性対照データの再評価を行って、その検査手順の変更によって処置関連変化の検出力が低下するようなことがないことを確認しなければならない。

98. OECD の発達神経毒性試験ガイドライン（TG 426）では検査項目ごとに推奨する標本の大きさが記載されている。ほとんどの検査項目では 10 匹／性／群（1 腹につき雌雄各 1 匹）とされている（2007 年 OECD 試験ガイドライン 426 および 1998 年米国環境保護庁 OPPTS 870.6300 の「発達神経毒性試験」を参照のこと）。試験機関で用いる特定の手順のもとで、各検査項目における投与による影響の統計的検出力が十分となるだけの標本サイズを確保しなければならない。行動の制御が不十分である場合、または行動に変動が大きい場合には、標本サイズを大きくする必要がある。ほとんどの場合、対照群のレベルと比較して 5～25%、またはそれ以下の行動変化を検出できるはずである（Buelke-Sam *et al* 1985, CBTS pp 591-634）。検査対象が幼齢または若齢動物である場合、成熟動物用の検査手順または測定装置を調整する必要がある。たとえば、通常は成熟動物の測定に用いられるセンサー（自発運動量測定装置の光電セルなど）を若齢またはより小型の動物に用いる場合、動作の検出に適当な高さに調節しなければならない。離乳直後または若齢ラットと成熟ラット間で試験感度が同程度になるよう装置の大きさ（モリス水迷路試験における水槽の直径など）または検査手順（記憶課題における遅延時間の長さなど）の調節が必要となる場合もある。

初期行動発達検査

99. 初期行動発達については、発育期の動物における反射の発達を調べることで評価を行う（表 3 の例を参照）出生後初期における運動のパターン（the pattern of motor activity）の変化も初期行動発達の評価に用いることができる。出生後の発達時期が異なる 2 種類の行動を指標とすべきである。初期行動発達検査に多くの指標を用いる場合は、初期行動発達の特徴を捉えるために、指標ごとに適当な発育段階で複数の時間点を設けて評価を行う。

100. 機能的評価項目の検査に際しては、全個体について同一基準によって測定できるように訓練が必要となる。このため、各機能的評価項目の測定は全個体について同一人物が、または数人で同一測定基準を用いて行うことが望ましい。機能的評価項目の測定を複数の観察者が行う場合は、訓練を行って、測定基準が同等となり、観察者間の測定基準の信頼性が確保されるようにしなければならない。

101. 身体的発達指標または機能的評価項目の検査時の取り扱いがその後の動物の行動に影響を与える可能性があるため、全投与群の児動物を同様に扱うことが重要である。陽性結果が得られるまで（眼瞼開裂など）動物を取り扱えば、投与群で取り扱いに差が出る（最も作用が出やすい群を最も長時間扱うなど）ことになり、そのような取り扱いがその後の行動評価項目に影響しかねない。

表 3. 反射の発達および初期行動発達の検査項目

検査	概要	評価項目	検査日齢（1）（出生後日齢）	参考文献
表面立ち直り反射	平坦な場所に児動物を仰臥位にして置く。立ち直るまでの時間、または制限時間内における反射の有無を記録する。	神経運動機能	2～4	Altman & Sudarshan, 1975
負の走地性反射	傾斜面に児動物の頭部を下方にして置く。180°向きを変えるまでの時間、または制限時間内における反射の有無を記録する。	反射の発達 神経運動機能	7～10	Adams, 1986
帰巣反応	飼育に使用されている床敷および洗浄済みの床敷の間に児動物を置く。選択した床敷および選択するまでの時間について記録する。	反射の発達 神経運動機能、感覚機能（嗅覚）	6～10	Adams, 1986
空中立ち直り反射	空中で仰臥位の姿勢にした児動物を落下させる。四肢による着地能の有無について記録する。	反射の発達 神経運動機能	12～17	Altman & Sudarshan, 1975、Adams, 1986
聴覚性驚愕反射	児動物に対して突然音を提示する。驚愕反応の有無について記録する。	反射の発達 神経運動機能	10～14	Buelke-Sam & Kimmel, 1979
泳力発達検査	遊泳力の各発達段階について評価を行う。体位、四肢の使い方、移動方向について記録する。	反射の発達 神経運動機能	6～25	Schapiro *et al.*, 1970、Vorhees, 1983
運動量	運動量測定装置内に児動物を入れる。自発運動量およびセッション内馴化について記録する。	神経運動機能	13～21（通常は 13、17 および 21 日齢）	Ruppert *et al.*, 1984, 1985

（1）検査日齢は一般的な推奨日齢であり、試験機関ごとに各検査の日齢を設定しなければならないことに注意。

運動量の測定法（Measures of motor activity）

102. 運動量の検査では、全般的な運動量に関するものと非連合学習（馴化）に関するものとがある。運動量はさまざまな検査によって評価することができる。これまでに用いられてきた検査の種類について、表 4 にその例を示す。OECD 試験ガイドライン 426 では、活動性検査については自動測定装置によって行うこと、測定装置へ入れる動物は 1 匹ずつとし、一定時間測定することが推奨されている。また、OECD 426 では、運動量の測定は少なくとも離乳期前および成熟期に行うことが推奨されている。運動量の馴化とは、1 試験セッション内における活動性の経時的減少であり、試験セッション内で連続して測定された一定間隔での活動性レベルを経時比較する

ことにより評価される。試験セッションの終了時の運動量スコアは漸近値になっていなければならない。馴化は離乳直後から発達し、幼齢動物（生後（PND）12 日齢など）では運動量の正常なセッション内馴化は通常示されないため、離乳期前の動物の初期行動発達の評価に運動量検査を用いてもよい（表 3 参照）。試験セッションについては、運動量の正常なセッション内馴化の測定を行うために十分な時間とする。これに関して、短時間検査（2～10 分間）では正常なセッション内馴化が認められないため、そのような短時間検査は採用すべきではない。

103. オープンフィールド試験およびホールボード試験は、通常短時間の活動性および新奇環境に対する探索行動の測定に用いられる（通常、試験セッションは 2～5 分）。ただし、この 2 試験は長時間の運動量や馴化の検査に応用され、そのような場合は試験ガイドライン 426 に準拠した検査法として有用である。

104. 自動測定装置には 8 字型迷路、放射状迷路、ホームケージに似た新型ケージなどを用いたものがあり、馴化検査と同様に、長時間の運動量の測定に用いることができる。数種のなるタイプの化学物質について、各種装置について試験機関内および試験機関間における信頼性が示されている（Crofton *et al.*, 1991, Moser *et al.* 1997, Buelke-Sam, 1985）。

105. オープンフィールド試験：オープンフィールド試験は通常、長方形または円形の測定フィールドを用いて行う。被験動物を一定時間測定フィールド内へ入れ、運動量に関わるいくつかの測定指標が記録される。測定指標には、移動量計測値や移動距離、立ち上がり回数（前足が同時に両側とも床から離れた回数）や垂直運動計測値、立ち上がり時間などがある。また、運動パターンの指標として、踏み込み行動や運いていた位置（測定フィールドの周辺部か中央部か）がある。立ち上がり回数（number of rears）を計数してもよい。オープンフィールド試験で得られる活動性レベルは、一般状態の詳細症状観察として適切と認められる短時間（約 2 分）の測定で評価される場合と、運動量および馴化の評価として長時間試行（約 30 分）によって評価される場合がある。新奇なオープンフィールドにおける運動量の検査は、特殊なオープンフィールド試験とみなされる。この場合、測定チャンバーが当該供試動物にとって初めての新奇なものであること以外、測定方法は同じである。オープンフィールド試験によって運動量を測定された経験がない個体にとっては、オープンフィールド試験は事実上新奇チャンバー試験（訳者注：新奇場面への適応性を見るもの）となる。運動量の測定は、他の運動量検査の場合と同様に、被験動物を測定チャンバー内に入れて光電セルを用いるかビデオ撮影による記録を用いて行う。運動量の定量化方法には何種類もあるが、最もよく用いられる方法は、一定時間内における動作回数または動作時間を用いる方法である。

106. 8 字型迷路：8 字型迷路はオープンフィールド試験の改良型とみなされる。そのデザインの特性から、8 字型迷路は自発運動量の検査項目の中でも環境因子の影響が少ない指標であると考えられている。本試験の装置は 8 字型の走路および中央交差部から外方に伸びる 2 本の袋路からなる。迷路内の動物の動きは迷路内の光電ビームによって記録され、運動量は一定時間内におけるビーム遮断回数として測定される。8 字迷路内の動物の動きについては、ビデオによる記録装置を用いて測定することも可能である。間隔のあいたビーム遮断回数を数えたり（移動運動と見なす場合が多い）、または迷路内の各地点における活動の割合を調べる（袋路内での活動の割合など）ことで活動パターンの評価も可能である。8 字型迷路については、活動性の増加や低下を生させる薬物処理（d-アンフェタミン、クロルプロマジン、メチル水銀など）に対して感受性があることも分かっている（Reiter and MacPhail 1982、Elsner *et al*, 1988、MacPhail, 1999）。

107. ホールボード試験：ホールボード試験は運動量および探索行動の両者の測定に用いられる。ホールボード試験は、動物の鼻部は入るが体全体は入れない程度の大きさの穴が床にある以外、オープンフィールド試験と同じである。用いるホールボードの構造はさまざまである。探索行動については、穴を探索する時間を測定することで定量する。ホールボード内の穴に飼料を入れておくことで、記憶の検査課題である放射状迷路と同様の課題としてホールボード試験を利用することができる（File, 2001、Van der Staay, 1999、Brosnan-Watters, 1997）。

108. 放射状迷路：放射状迷路試験は空間記憶を評価するために開発された試験であるが、自発運動量の評価に用いることも可能である（Olton, 1987）。放射状迷路試験の手順については後述の学習および記憶の検査の項に記す。

表 4. 運動量の検査方法

検査	概要	評価項目	参考文献
オープンフィールド試験	四角形または円形の箱。 移動行動、立ち上がり行動。	短時間または長時間の活動性 探索行動	Adams, 1986、Denenberg, 1969、Schiorring, 1979、Walsh, 1976
ホールボード試験	床に穴がある箱。穴の底に目的物を置く場合もある。	短時間または長時間の活動性 探索行動	Adams, 1986
放射状迷路	中心部から放射状に伸びた 8 本の走路。	短時間または長時間の活動性 移動行動、学習行動	Adams, 1986、Walsh & Chrobak, 1987
8 字型迷路	相互につながった数本の走路。8 字型を形成。	長時間の活動性	Adams, 1986、Reiter & MacPhail, 1982、Crofton *et al.*, 1993
新奇ケージ検査	赤外線ビームを張り巡らした標準ケージ（飼育ケージではないこと）。	長時間の活動性	Gårdlund *et al.*, 1991

運動機能および感覚機能の検査

109. 運動機能および感覚機能の評価は少なくとも青年期（adolescent period）に 1 度と成熟期初期に 1 度行う（OECD 試験ガイドライン 46 参照、訳者注：原文誤記・試験ガイドライン 426 参照）。各評価は定量的評価でなければならない。

110. 近年まで、発生毒性試験における感覚機能の検査はカテゴリー評価（反応の有無）に限られていて、反応の大きさによるより鋭敏な測定ではなかった。しかし、洗練され自動化された行動測定技術が利用できるようになり、感覚機能のより高感度で定量的評価が可能になった(表 5)。感覚機能をより高感度で測定することが可能なため、そのような定量的検査法を定性的検査よりも優先して使用することを推奨する。

111. 反射または協調運動の発達評価に神経運動機能検査がしばしば用いられる。反射および運動の発達に関する検査は、催奇形性試験における行動学的検査で使用される機能的評価項目の中で最も普及したものである（Adams, 1986、Buelke-Sam and Kimmel, 1979）。一般的に用いられる検査手順のほとんどは文献で確認できる（Barlow and Sullivan, 1975、Adams, 1986）。したがって、ここでは表 3 に概略を載せるのみとする。神経運動機能の評価に用いられる検査項目としては、ロータロッド（および変法として加速ロッド）、握力、平均台試験、着地開脚幅がある（表 5）。

表 5. 神経運動機能および感覚機能の検査項目（1）

検査	概要	評価項目	検査を行う日齢	参考文献
ローターロッドおよび加速ロッド	一定速度で回転しているロッドに乗せる。ロッドに乗っていた時間、または制限時間内における落下の有無を記録する。加速ロッドも同様だが、ロッドの回転速度を増加させていく。加速ロッドはより要求条件の多い、より感受性の高い検査である。	神経運動機能、前庭系機能、運動協調性	通常、離乳後だが、離乳前の実施も可	Jones and Roberts, 1968、Kaplan and Murphy, 1972、Bogo *et al.*, 1981、Adams 1986、Pryor *et al*, 1983a、Bushnell *et al.*, 1994
握力	細いワイヤーにつかまりぶら下がっている時間または専用装置により前肢および後肢の握力を測定する。	神経運動機能	生後 13 日以降	Meyer, *et al.*, 1979、Pryor *et al.*, 1983a、Crofton *et al.*, 1990, 1994a, 1994b
着地開脚幅	30 cm の高さから紙上に落下させる。両足間の距離を測定。	神経運動機能	通常、離乳後	Gabriel *et al.*, 1991
平均台試験	動物を細い梁の上に乗せ、避難台またはホームケージにたどり着くためには渡らなければならない状況に置く。渡りきるまでの時間、滑る・落下について記録する。	神経運動機能	通常、離乳後	
負の走地性反射	45 度の傾斜面に顔を下方にして動物を置き、方向転換により顔を上方にできるか否か、できた場合の時間について記録する。	神経運動機能、前庭系機能	生後 7～10 日	Pryor *et al.*, 1983a
プレパルスによる驚愕反応の変容	聴力。 視覚的または触覚的プレパルスも用いられる。	定量的測度、反射の変容、感覚運動統合	離乳期近く、または離乳後に測定	Adams *et al.*, 1986、Crofton *et al.*, 1990
聴覚性驚愕試験	聴覚の反射反応。	定量的または定性的	生後 10 日以降	
聴覚性驚愕反応の馴化	定期的に大きな音を提示。驚愕反応の大きさを測定。	馴化（非連合学習）		
瞬目条件付け	瞬目反射を惹起するエアパフと組み合わせて手がかり刺激を提示。	聴力 視覚、触覚、学習	通常、離乳後	Stanton & Freeman 1994
オペラント条件付けを利用した試験	学習訓練で獲得される反応を用いて感覚の閾値や感度を測定（条件付け回避など）。	感覚機能の評価に広く用いられる。	通常、離乳後だが、飼料による動機付けを行う場合は、成熟期	Elsner, 1991、Gabriel *et al.*, 1991、Bushnell *et al.*, 1994、Pryor *et al.*, 1983b

（1）反射の発達指標として用いたもののいくつかが、感覚機能または神経運動機能の検査項目にもなっていることに注意。

112. 負の走地性の試験は前庭機能の評価に用いられる。本試験では、傾斜面（通常約 60°）に鼻部を下方にしてラットを置き、ラットが 180°体位を変えて鼻部を上方に向けるまでの時間を測定する。本試験は若齢動物の感覚運動機能の評価に有効であるが、成熟動物では変動が大きく、その有用性が限られると考えられている（Pryor *et al.*, 1983a）。

113. ローターロッド試験は平行感覚と運動協調性の評価に用いられる。ローターロッド試験（Bogo *et al.*, 1981）では、モーター駆動によって動くシリンダー上にラットを置き、回転するシリンダー上でラットがバランスを保っている時間を測定する。本試験は加速シリンダー（加速ロッド）または一定速度のシリンダーのいずれかを用いて行う。加速シリンダーによる試験では、床敷にラットが落下した時点の速度を記録するのに対し、一定速度下での試験では、シリンダーから落下した時間、または制限時間（約 60 秒）後の運動性（performance after a fixed time period）について記録する。加速ロッド試験の方が加速しない回転ロッド試験よりも感度が高いと考えられている。

114. 行動を用いた感覚機能の定量的な測定は、主として反射の変容を利用した試験や条件づけによる行動を用いた弁別試験として行われる。感覚的な手がかりに対して特定の反応をとるよう動物を訓練することで、感覚的な手がかりに対する検知能力が評価できる。嗅覚的、聴覚的、または視覚的な手がかりを用いれば、被験動物が手がかりを知覚できるか、または手がかりの変化（可聴周波数の変化など）を知覚できるか調べることができる。刺激条件（音声または光の周波数または強度）を変化させても反応に変化が認められない場合、その個体は刺激の性質の変化を検知できなかったことを意味する。刺激の性質または強度を変化させることによって、感覚機能の検知閾値の測定が可能である。

115. 聴覚性驚愕反射馴化試験（*Auditory startle habituation*）では驚愕反射を繰り返し誘発する。一定の間隔で反復していくうちに、驚愕反射誘発刺激に対する馴化が生じ、驚愕反応の強度が低下する。本試験は反射能の評価だけでなく、単純な非連合学習（すなわち慣れ）の評価にも用いられる。通常、げっ歯類で本試験を実施する場合、驚愕試行を約 50 回与え、反応強度を 10 試行を 1 ブロックとして平均化し、試験セッション中の反応強度の変化を評価する。驚愕刺激の大きさ、時間および間隔、ならびに背景雑音レベルなどの設定については、試験機関、被験動物の系統および日齢に合わせて試験感度が最大となるよう最適化する。驚愕反応はほとんどの哺乳類で明らかになっている。最も普及している驚愕反応は聴覚性驚愕反応であるが、試験手順および検査方法の感度について十分な裏付けがあり、妥当性が立証されていれば、他の反射的驚愕反応も利用可能である。

116. 驚愕反応のプレパルス抑制（*Pre-pulse inhibition of startle response*）は、感覚情報の統合能を評価する場合に用いられる。（Ison, 1984、Crofton and Sheets 1989、Koch, 1989）プレパルス抑制とは、通常では驚愕反応を誘発しない別の刺激が驚愕反射誘発刺激に先行した場合に生じる驚愕反応強度の抑制である。プレパルス抑制はほとんどの哺乳類で明らかになっている。一般的に、プレパルス刺激自体が驚愕反応を誘発するものであってはならない。通常、驚愕刺激の 30～100 ミリ秒前にプレパルス刺激を提示すると、正常動物では驚愕反応の大きさが強度依存性の減少を示す。また、驚愕反応の潜時は、プレパルスが提示されると正常動物では変化しないか増加する。プレパルス抑制は複数の感覚入力の統合を反映していることから、プレパルス抑制の障害は情報統合能力障害を反映していると考えられている（Koch, 1989, Ison 1984）。プレパルス刺激の性質を変えることで、プレパルス刺激を検知する感覚能力の変化を測定することができる（Crofton *et al.*, 2000、Wecker *et al.*, 1985）（訳者注：プレパルス音の強度を弱くして感知されなくなると抑制が生じなくなることから聴覚閾値の測定が可能）。

117. 瞬目反射条件付けは古典的学習またはパブロフ型学習（訳者注：classiscal or Pavlovian

learning とあるが古典的あるいはパブロフ型条件付け－conditioning－とするのが一般的）の一種であり、無条件刺激（通常、眼にエアパフを吹き付ける）の直前に条件刺激（通常、音）を提示する。条件刺激のみでは正常時は何の反応も誘発しない。無条件刺激は無条件反応（瞬目反応など）を起こす。ただし、無条件刺激（エアパフ）を提示する直前に条件刺激（音など）を提示することを繰り返すと、無条件刺激がなくても条件刺激（音）で瞬目反応が誘発されるようになる。この場合、学習した瞬目反応は条件反応であり、基礎的な学習能力の指標となる。（総説は Stanton and Freeman, 1994 を参照）。ラット、ウサギおよびヒトにおける瞬目条件付けによる試験法がある。ヒトおよび動物モデルを用いた試験を比較すると、生物学的変数の多くで作用に種間類似性が認められ、また実験的事実から生物学的メカニズムの種間類似性があることが示唆される。瞬目反射条件付けモデルは主として学習効果の研究に用いられてきたが、条件刺激の性質（周波数、強度など）の変化が条件反応に与える影響を調べることで、感覚系機能の評価にも本試験を適用することが可能である。

118. ここまで解説してきた試験法の他に、オペラント条件付けを用いた試験も感覚機能の評価に用いることができる。古典的条件付けの手続きでは反応の前に条件刺激を提示するのに対して、オペラント条件付け手続きでは反応後（レバー押しなど）に無条件刺激（報酬としての飼料など）の提示を行うが、両者は完全に分離されるものではない。オペラント条件付けを用いた試験を感覚機能の評価に用いる場合、瞬目反応試験と同様に、反応を制御する感覚的手がかりの性質（音の周波数または強度、光の周波数またはスペクトル特性など）の変化に対する動物の検知能力を評価することになる。オペラント法は、行動を制御する感覚的手がかりとしてさまざまな種類の手がかりを用いることができることから非常に柔軟性に富む。このため、オペラント条件付けを用いれば、現在ある感覚機能評価試験の中で最も感受性の高い試験となりうる。

119. 条件性回避反応（*Conditioned avoidance response*）（CAR）では、警告信号（音または光など）から予知される嫌悪刺激を回避するよう動物を訓練する。警告の手がかりの提示は嫌悪刺激を与える前に行い、被験動物はそれに反応することで嫌悪刺激を回避することを学習しなければならない。試行によって、嫌悪刺激を回避するためにさまざまな行動（壁を超える、レバーを押す、棒を登る）を取るよう動物を訓練することは容易である。たとえば、警告信号（条件刺激）を先に提示すると、足への電気刺激を回避するために棒に登ったり、棒を引くことをラットは学習するようになる（Pryor *et al.*, 1983a, b）。多種多様な刺激（視覚的、聴覚的、嗅覚的）を条件刺激として用いることができるため、警告の手がかりの強度または周波数を変えることで、その感覚的警告刺激の変化に対する動物の検知能力または識別能力の評価が可能である。たとえば、測定する反応を棒登りとし、手がかりを聴覚刺激として用いると、トルエン曝露を受けたラットでは 8 kHz を超える周波数で聴覚障害が認められることが分かっている。陽性対照物質を用いた実験によって、CAR を用いた試験の試験感度や特定の感覚障害に対する検出力が示されている。

120. オペラント条件付けを用いた試験は体性感覚機能の評価にも用いることができる。たとえば、Elsner（1998, 1991）は、強度感受性のあるレバーを押すようにラットを訓練して、ラットにおける触覚・運動感覚能力試験法を開発した。試験試行を繰り返す内に、強化子が得られるレバー押し反応圧の範囲を徐々に減少させていくことで（訳者注：一定範囲の力でレバーを押した時のみ報酬を与える手続きで訓練し、その範囲を徐々に狭くして行く）、レバーを押す力の制御能力の評価が可能となり、体性感覚機能の尺度として本試験を用いることができる。ラットを用いた本試験によって、メチル水銀が、運動の微調整に関連した障害など、触覚・運動感覚系に影響を及ぼすことの実証に成功している（Elsner *et al.*, 1991）。

学習および記憶の検査

121. 児動物における連合学習および記憶の検査については、離乳後に 1 回および成熟期に 1 回

の合計 2 回行う。具体的な検査法の選択肢としてはかなりの種類があるが、検査法は a）試験セッション間または 1 試験セッション内における行動変化を調べることで学習を明らかにするもの、かつ b）学習（習得）および記憶（保持）の両者に関する指標を取り入れたものでなければならない。2 回の検査時期に用いる動物は同一個体であってもよいが、1 回目の検査で学習したことが 2 回目の検査の学習指標に交絡を生じさせる可能性を最小限にするために、通常、各検査時期の被験動物には異なる個体を用いるとよい。若齢期および成熟期の検査に同一検査法を用いる場合、特定の動物群を両時期の検査に供すると、学習の評価が十分なものとならない可能性がある。

122. 学習および記憶の評価に用いられる連合学習課題は非常に多い。通常、動機付けによって飲食欲よって動機付けられた課題（appetitively-motivated task）と嫌悪によって動機づけられた課題（aversively-motivated tasj）のいずれかに分類することができる（Lochry *et al.*, 1986）。飲食欲を動機付けとした学習では反応の維持に強化子を用いる（すなわち、動物は強化子を獲得するために何らかの反応をする）。このような課題では、被験動物に動機付けをするために蔗糖やサッカリンなどの嗜好性のある物質を利用しても良いが、通常は動因の水準を管理するために給餌制限または給水制限が必要となる。嫌悪を動機付けとした学習では反応の動機付けに嫌悪的刺激、不快刺激、疼痛性刺激（足への電気刺激、まぶしい照明、大きな音、水など）を用いる。このような嫌悪刺激による課題は、さらに嫌悪刺激の提示前に反応すれば完全にそれを回避できるという選択が動物にあたえられるか、嫌悪刺激から単に逃れることができるだけであるかによって、逃避課題と回避課題のいずれかに分類される。通常、飲食欲を動機付けとした課題では給餌制限、給水制限のいずれかが必要となるが、嫌悪を動機付けとした課題（回避課題、水迷路）では、そのような制限は必要ない。発達期の若齢動物における給餌・給水制限の潜在的影響を考えて、動機付けにそのような制限が必要な飲食欲を利用した課題は原則的に未成熟動物に対しては行わない。

123. 能動的または受動的回避試験、味覚または嗅覚嫌悪試験、嗅覚条件付け、ならびにモリス水迷路、ビール迷路またはシンシナティ迷路、放射状迷路、T 迷路などの各種水迷路試験、さらにはオペラント条件付けを基礎とした各種学習および記憶試験（遅延位置見本合わせ、遅延場所交替反応、その他の交替反応課題、各種遅延反応課題など）など、連合学習としての基準を満たすと考えられる検査法の一部を表 6 に示す。離乳ラットおよび成熟ラットを用いた他の検査法については文献を参照されたい。

表 6. 学習および記憶試験の例（詳細については本文参照）

試験法	概要	利用可能な評価指標（1）	参考文献
受動的回避	被験動物は、嫌悪刺激を回避するためには反応しないということを学習しなければならない。	基準に達するまでの試行回数、誤反応数、反応潜時、暗箱に入るまでの時間	Adams 1986、Wier *et al.*, 1989、Lochry and Riley 1980
能動的回避	被験動物は、嫌悪刺激を回避するためには反応しなければならない（扉の向こう側に行く、レバーを押すなど）。	基準に達するまでの試行回数、回避成功回数、回避に失敗した場合の逃避回数	Adams 1986
オペラント条件付け（スキナー箱）	被験動物は、強化因子の獲得または嫌悪刺激の回避のためには反応しなければならない。試験日程、用いる刺激および反応の選択肢が広い。	正解回数、強化因子の獲得数、正解率、誤答率、（試験による）	Adams 1986、Hass *et al.*, 1994a
放射状迷路	中心部から放射状に伸びた 8 本の走路。各走路に強化因子を置く。	誤反応数、すなわち 1 度行った走路へ再び入った回数、運動性、基準に達するまでの試行回数、時間	Adams 1986
ビール型水迷路、シンシナティ迷路	連続 6 地点で左折・右折を選択させる迷路。	誤反応数、完走するまでの時間	Peele *et al.*, 1990
モリス水迷路	空間手がかりを頼りに円形水槽の水面下に隠れた避難場所を見つける。開始地点を変えることが重要。	泳動距離および泳動経路、避難場所に避難するまでの時間	Kelly *et al.*, 1988、Hass *et al.*, 1995
T 迷路	T 字型の迷路の 2 本のアームのうち、強化因子が置かれたアームを選択する。床面でも水面上でもよい。	誤反応数、時間、基準に達するまでの試行回数	Wier *et al.*, 1989、Lochry and Riley 1980

（1） これらの試験の多くは、試験の目的や特殊な用途にあわせて、さまざまな評価指標を測定出来ることに注意。ここでは当該試験で測定可能な評価指標のうちよく使用されるものについて概説した。

124. 受動的回避課題では、被験動物は嫌悪刺激を回避するために反応を控えなければならない（何らかの能動的反応を示さなければならない能動的回避課題との相違点）。たとえば、最もよく用いられる受動的回避課題には、2 区画に仕切られた箱の中で、足に対する電気ショックなどの嫌悪刺激を受けた経験のある区画を回避する課題がある。本回避課題では、電気ショックが与えられる区画に入るまでの時間を測ることで連合学習の評価を行う。受動的回避課題は技術的に容易で迅速に行うことができ、学習訓練試行がほとんど必要ない場合が多い。ただし、結果に変動が認められる場合があり、運動機能障害によっても変化すると考えられる。本試験の感度は受動的回避の試行を繰り返すことで上昇する。反復試行を実施する場合には、被験動物が暗箱に入らなくなるまでの試行回数などを評価項目として用いる（Robbins, 1977）。能動的回避課題は、

嫌悪刺激を回避するために被験動物は能動的な反応を取らなければならない点を除いては、受動的回避課題と同じである。通常、嫌悪刺激の開始を予告する手がかり刺激が嫌悪刺激と連合され、被験動物は嫌悪刺激を回避するために反応を示す（別の場所へ移動する、レバーを押す）。受動的回避課題と同様に、能動的回避課題も学習が迅速に行えるが、運動機能障害によっても変化すると考えられる。学習の尺度としては、回避反応回数または反応するまでの時間が用いられる。

水迷路

125. 学習および記憶の評価を行うための水迷路が数多く開発されてきた。水迷路はすべて主たる動機付けが水から逃避することであることから、嫌悪を動機付けとした課題である。課題の多くは、それ以前に開発された食欲を動機付けとした迷路（T 迷路など）の遊泳版であるが、特定の学習および記憶の能力を測定するために開発された課題もある（モリス水迷路など）（総説としては Vorhees, 1987、Vorhees *et al.*, 1991、Morris, 1984、Brandais *et al.*, 1989、D'Hooge *et al.*, 1991 を参照）。水迷路には、給餌制限を要しないことを初めとして、若齢動物にも利用できる、迅速に学習できる、食欲または味覚による交絡を受けない、嗅覚的な手がかりの影響を比較的受けない、限度はあるが体格の差異による影響を比較的受けないなど、利点が多い（Vorhees *et al.*, 1991）。水迷路は運動要求性であるものが多いため、特に選択の正確さではなく時間や速度を尺度とする水迷路では潜在的な運動障害を考慮することが重要である。

126. モリス水迷路（*Morris water maze*）はげっ歯類における空間の学習および記憶について評価するために開発された（Morris, 1984、Brandais, 1989, D'Hooge *et al.*, 2001）。この迷路は大型円形水槽と水面下に隠れる避難台からなる。試験方法はさまざまだが、最も典型的な方法は、4 方向のうち無作為に選んだ 1 方向から水迷路内へ被験動物を入れ、水迷路から逃れるために隠された避難台を見つけさせるというものである。測定回ごとに開始地点を変えるため、動物は迷路外の空間的手がかりを頼りに、隠れている避難台を探すことを学習しなければならない。（Morris 1981, Morris 1984）。避難台を見つけるために動物に確実に空間的手がかりを利用させるよう、試行時ごとに開始地点を変えることが必要不可欠となる。学習指標は避難台を発見するまでの遊泳時間および距離、ならびに遊泳速度である。遊泳時間は個体の遊泳速度の影響を受けるため、通常は遊泳距離の試行反復による減少の方が学習指標として適している。本試験の利点は、学習試行中に各個体の遊泳速度（運動能力）に関する情報が得られることである。各個体の遊泳経路を記録しておくことで、さらに詳細な能力分析も可能である。通常、記憶の評価は試験の最終日に実施し、避難台を取り除いて遊泳させ、避難台があった位置で留まっている時間を記憶の指標とする。本課題は、関連する脳部位を様々な方法で実験的に損傷する操作や加齢による障害に対して感受性を有することがわかっている。げっ歯類を用いて、エタノール、工業用キシレン、N-メチルピロリドン、トルエンの出生前または出生後初期における曝露や低栄養などの影響を調べる試験に用いられている（Blanchard *et al.*, 1987, Kelly *et al.*, 1988、Hass *et al.*, 1994a, b, Hass *et al.*, 1995）。モリス水迷路には、作業記憶（または短期記憶）の評価に用いられる変法もある。その場合、避難台を用いて 1 日 2 回測定を行い、避難台の位置については 1 日 2 回の測定時では変えず測定日ごとに変える。こうして、1 日のうち 1 回目では避難台の位置に関する情報を提供し、その情報の記憶を 2 回目に測定する。指標としては、避難台を発見するまでの遊泳時間、遊泳距離および遊泳速度がある。本課題に位置に関する手がかりを付加した方法もあり、手がかりによる学習試験、または運動機能試験の対照課題として用いられる。

127. ビール型水迷路（*Biel water maze*）は迷路の開始地点から目標地点まで連続的に6回の左右位置選択を求める課題である。ビール型水迷路を用いた試験では、泳力を評価するために直線水路による初期試行を行い、その後走路を6回連続選択させるビール型水迷路試行を実施し、さらにスタートボックスとゴールボックスとを学習習得試行時とは逆にして試行を実施するよう試験を構成する場合もある。通常、学習および記憶の評価に用いる尺度は各試行時の誤反応数、すなわち、右折すべき地点で左折する、またはその逆の回数である。迷路走行に要した時間についても原則的に試行ごとに記録するが、学習評価の尺度としては選択の正確さよりも劣る。各種化学物質の生殖毒性試験において、ビール型水迷路の信頼性の高さが示されており、作用の検出に関して検出係数（coefficient of detection）10～20%を有する感度があり、神経毒性物質の影響を反映する試験として妥当である。

128. T 迷路（*T-maze*）試験に用いる装置は、スタートボックス、左右走路およびスタートボックスと左右両走路とを結ぶ選択地点からなる。T 迷路は、被験動物が正しい走路を選択した時に報酬が与えられる水を用いない迷路としての使用（飲食欲を動機付けに用いた試験）と、水迷路としての使用（嫌悪を動機付けに用いた試験）がある。多くの試験で用いられる課題は、位置に関する習慣形成課題（position habit task）であり、この課題では試験中常に同一選択肢が正解とされる。その他には正反応を左右毎回変える遅延交替課題や、2 回の走行を 1 試行として分散試行型（訳者注：試行間間隔を空けて実施する）遅延交替課題がある。1 回目は強制走行、すなわち被験動物が選択できるのは 1 走路のみとし、2 回目には、強制走行時に入った走路とは別の走路を報酬が与えられる選択として、自由に選択させる。各試行間の遅延時間を変化させることにより、短期記憶または長期記憶の評価が可能であり、また、記憶障害による遂行成績の変化をはっきりさせるために、遅延時間を延長することもできる。常に同一走路を正解とした学習課題はラットにとって比較的容易であり、化学物質の曝露による影響を受けないが、遅延交代課題を用いた試験結果では明らかに認知過程における障害が示される（Robbins, 1977）。

129. ラットでは空間学習過程に海馬およびその関連構造が大きく関与している（Olton *et al.*, 1979, Morris *et al.*, 1982）。空間学習の評価に一般的に用いられる方法としては、モリス水迷路および放射状迷路の 2 つがある。

130. 放射状迷路（*radial arm maze*）は飲食欲を動機付けとした課題であり、その装置は中心部から放射状に伸びた 8 本の走路からなる。通常、迷路の 8 本の各走路の端に強化子を置き、試行時は被験動物を中心部に置いて各走路の端にある強化子を探索させる。放射状迷路では短期記憶（作業記憶）および長期記憶（参照記憶）の両者が評価可能である。記憶の尺度には誤答数（走行済みの走路に再び入った回数）がある。放射状迷路はエタノールやトリメチルスズなどの化学物質の影響（Miller *et al.*, 1982）、母性フェニルケトン尿症や出生前フェニトイン曝露の影響（Weisenburger *et al.*, 1990）、メチルニトロソウレアにより誘発された小頭症の影響（Akaike *et al.*, 1988）の評価試験で感受性があることが示されている。

131. スケジュール制御オペラント条件付けは、既述の感覚機能だけでなく、学習および記憶を評価するための様々な試験条件を提供する。オペラント課題は、注意、反応抑制（衝動性）、弁別、空間記憶および非空間記憶など、他の認知機能の評価試験として開発された。利用可能なオペラント試験課題には学習および反応に関する実験条件変数を様々に制御することが可能なものがあるため、そのような課題を用いれば、特定の認知機能に対して感受性の高い試験が行える。ただし、オペラント課題を用いた試験では、行動に厳密なスケジュール制御を与える必要があるため、発達毒性試験における学習および記憶の初期評価としては課題の訓練に非常に長い時間がかかる。通常、オペラント課題を用いた試験では飲食性の強化子（appetitive reinforcer）が用いられるが、嫌悪性の刺激も用いることが可能である。飲食性の強化子を利用したオペラント課題の方が多用されているが、給餌制限や給水制限を要する上、長期間の訓練が通常必要であることから、一般的に若齢動物の学習および記憶の試験には適さない。完全な一覧というにはほど遠いが、オペラント課題を用いた試験の例を以下に示す。（概論：Buccufusco 2002、Laties, 1978）

132. 空間弁別学習（*Spatial discrimination learning*）：強化子を獲得するために、オペラント課題用装置内での空間的位置（左右の選択など）を頼りに反応を選択するよう（通常、2 つのうち 1 つのレバーを選択させる）被験動物に学習させる。

133. 手がかり刺激弁別学習（*Cued discrimination learning*）：強化子が獲得できるレバーを示唆している手がかり刺激（音、光）に基づいて反応を選択するよう（通常、1 つのレバーを選択させる）被験動物に学習させる。弁別学習を終えたら、手がかりを変化させて（音刺激周波数、光刺激周波数など）、感覚刺激特性の変化を動物が検知できるか否かを調べてもよい。通常、誤答数、一定基準に達するまでの試行回数、誤答の種類を尺度とする。

134. 交替反応課題（*Alternation tasks*）：試行ごとに交互に反応するよう被験動物に学習させる。空間（左右など）や手がかり刺激（照明のあるレバー対照明のないレバー）に基づいて交互に反応することとする。前の試行時の正解に基づいて正しい選択を行わなければならないため、動物は前の試行時に選択したことを記憶しておかなければならない。本課題では各試行間の時間間隔を延長させていくことで記憶の測定（訳者注：短期記憶の保持忘却曲線を得ることが可能）が可能となる。通常、誤答数、一定基準に達するまでの試行回数、誤答の種類、反応が基準以上のレベルで維持されなくなる遅延時間の長さが、尺度として測定される。

135. 遅延空間交替反応課題（*Delayed spatial alternation*:）：交替反応課題の特殊型であり、選択は空間的位置（左右）に基づくが、試行と試行の間に遅延時間を導入することで記憶の評価を行う（訳者注：134 後段で説明されている課題であり、短期記憶の保持忘却曲線を得ることが可能）。通常、誤答数、一定基準に達するまでの試行回数、誤答の種類、反応が基準以上のレベルで維持されなくなる遅延時間の長さが、尺度として測定される。

136. 見本合わせおよび非見本合わせ課題（*Matching to sample and non-matching to sample tasks*）：選択直前に提示される手がかりを頼りにレバーに反応するよう被験動物を学習させる。見本合わせの場合は、試行直前に手がかりや見本となる刺激（光など）を動物に提示して、手がかりと対応するレバーを選択させる。手がかり刺激の提示から選択試行の開始までの時間を増加させていくことで、手がかり刺激に対する記憶が評価される。遅延非見本合わせはほぼ同様にして行うが、被験動物には直前に提示された手がかりや見本刺激に合致しない方のレバーを選択させる。通常、誤答数、一定基準に達するまでの試行回数、誤答の種類、反応が基準以上のレベルで維持されなくなる遅延時間の長さが、尺度として測定される。

137. 低頻度分化強化課題（*Differential reinforcement of low rate tasks*）：強化子を獲得するためには、定められた時間間隔のあいだ待ってから反応しなくてはならないことを被験動物に学習させる。一定時間を待たずに反応した場合は試行をやり直す。本試験法では、動物の待時間を増加させることによって、反応抑制能や衝動性が測定できる。通常、反応が基準以上のレベルで維持されなくなる遅延時間が尺度として測定される。

138. 弁別逆転課題（*Discrimination reversal tasks*）：以前は正反応であったレバー（右側のレバー、照明の当たっているレバーなど）が正反応ではなくなり、以前は正反応ではなかったレバーが正反応となったことを被験動物に学習させる。本試験法では、新たな正反応を学習するまでの試行回数を測定し、学習と反応の両者の柔軟性を測定する。通常、逆転学習レベルが設定した基

準に達するまでの試行回数を尺度として測定する。

発達神経毒性（DNT）試験における神経系組織の死後検査

139. 発達中の神経系に対する毒性学的な障害が、成熟期と比較して若齢期において大きくなる可能性があるのは、組織形成過程（細胞の増殖、分化、移動、再構築、髄鞘形成など）関係していることによる。このような考え方は、神経発達の臨界期に起こる事象の時間軸上で種特異性はあるものの、ヒトを含めた全動物種間で不偏的である（Moore, 1989、Rice and Barone, 2000）。DNT 試験では機能的および神経行動学的評価項目とともに形態学的指標についても検査を実施して、発達期曝露後の神経系における投与関連有害作用を評価する。

140. 発達神経毒性試験のガイドライン（TG 426）では、離乳期（生後約 21 日）またはそれ以前（生後 11 日など）、および試験完了（生後約 60 日）の時点で、選抜した児動物（10 匹/性/群）から採取した神経系組織の病理組織学的評価が求められている。これによると、離乳直後の児動物については活発に神経が発達している期間かつ投与期間後期に採材することとなり、成熟した（生後 60 日）児動物については脳の成熟後かつ投与終了後約 40 日に採材することになる。いずれの時点でも、神経系構造の肉眼病理検査と脳重力測定の後に（10 匹/性/群）、脳組織の採取を行う。末梢神経系組織の採取は成熟期の児動物についてのみ行う。組織の定性的評価については、通常、最初に対照群と高用量群との比較を行い、そこで投与関連作用が認められた場合、さらに低用量群および中間用量群についても評価を行う。投与関連所見については、主観的検査（どれが投与群が知らない担当者が実施すること）も行う。さらに、両時点について定量的評価を行う。ガイドラインでは少なくとも大脳皮質の主要部位、海馬、小脳の計測が規定されている。

141. 脳重量については、各群雌雄各 10 匹の児動物について測定する。各試験項目に別個体を割り当てる変法もあるため、脳重量測定に割り当てた児動物についてはさらに病理組織学的検査を行う場合もあれば、行わない場合もある。このため、日齢の異なる動物の検査を予定通りに進めていく試験実行上の都合によって、脳標本固定の状況が決まることとなる。具体例については試験ガイドライン 426 付録 1 を参照されたい。脳重量の測定は固定前でも固定後でも良いが、用いる手順については全試験群で統一し、試験報告書には用いた手順を明確に記載しなければならない。

142. 神経系組織の肉眼的な形態検査は顕微鏡検査のための組織処理期間中のいずれの時点に行っても良い。

143. 神経系組織の病理組織学的評価を適切に行うためには、その前段階の固定を適切に行うことが必要不可欠である。成熟動物については、神経系組織に生じるアーチファクトを最小限にするため、剖検時に灌流固定（*in situ* perfusion fixation）を行う（OECD, 2004）。また、離乳期の児動物についても、脳組織の灌流固定の必要性（浸漬固定ではなく）について一般的な合意されるものの、この点については歴史的な論争が行われている（OECD, 2000）。通常、幼齢動物（生後 11 日など）から採材した場合には浸漬固定も適切な方法とみなされる。ただし、浸漬固定を行う場合は、固定液との接触面が十分となるように、必ず頭蓋冠を除去する必要がある。その

際、不注意によって組織を損傷させないように注意する。浸漬固定法と灌流固定法に関するさらに詳細な解説については、OECD のガイダンス 20 の段落 51～53（OECD, 2004）および多くの文献に著されている標準的な組織学的検査法を参照されたい（Prophet *et al.*, 1994、Fix and Garman, 2000、Bancroft and Gamble, 2002）。用いる固定液および固定液中に組織を浸漬する時間は、特殊染色または免疫細胞化学的染色の利用など、予定されている組織学的検査手順による（発達神経毒性試験に用いると考えられる特殊染色および補助的な手法の一覧表は Garman *et al.*, 2001 を参照）。ホルマリン固定法では、固定時間が長くなると脳が収縮する。このため、脳切片の作製、処理および包埋に関しては、最終的に形態計測評価を行う全個体について、ほぼ同じ時間点となるようにする（Garman *et al.*, 2001）。中枢神経系組織にはパラフィン包埋法を用いても良い。末梢神経系組織については、高解像度である必要がある場合（末梢神経障害が疑われる場合や形態計測を行う場合の末梢神経など）、オスミウム後固定およびエポキシ樹脂包埋が有効かつ適当である。

144. 採取および病理組織学的検査の実施対象となる中枢神経系および末梢神経系構造については試験ガイドライン 426 に定められている。規定上、脳の主要領域すべて（嗅球、大脳皮質、海馬、大脳基底核、視床、視床下部、中脳（中脳蓋、中脳被蓋、大脳脚）、橋、延髄、小脳）が含まれている。さらに、成熟期の児動物については、視神経および網膜を含む眼球、脊髄の頸膨大および腰膨大ならびに後根線維および前根線維、坐骨神経近位部（膝部）および腓腹筋枝など、脊髄および末梢神経系の代表的な部位の切片を作製する。脊髄および末梢神経の切片は横断面および縦断面の両者について作製する。

145. 神経病理学的検査における定性評価時には、通常成熟動物の神経病理学的検査で認められる細胞変化（神経細胞の空胞形成、変性および壊死など）および組織変化（神経膠細胞の増殖、白血球浸潤および嚢胞形成など）に加え、神経系の正常な発達における変化の有無に焦点を絞ることが重要である。発達障害を示唆する重大変化の例については、試験ガイドライン 426 に詳説されているが、以下が挙げられる。

- 嗅球、大脳、小脳の全体的な大きさ、または形の変化

- 脳の各部位の相対的な大きさの変化。正常では一時的にのみ存在する細胞集団や軸索投射の消失または残存による部位の拡大や縮小（小脳外胚芽層、脳梁など）を含む。

- 増殖、移動、分化の変化。これらは、過剰なアポトーシスや壊死を示す部位、異所性、配列不整、奇形などを示す神経細胞の集塊または散在性の分布、皮質構造各層の相対的な厚さの変化などにより示唆される。

- 髄鞘形成パターンの変化。全体の縮小や髄鞘構造の染色性の変化を含む。

- 水頭症の徴候。特に、脳室拡張、中脳水道狭窄および大脳半球菲薄化。

定性的評価の手順は、最初に対照群と高用量群の検査結果の比較を行い、必要であれば、より低い用量群の検査を行って、認められたあらゆる投与関連所見について用量反応関係を明らかにする。また、投与関連所見については、どれが投与群か知らされていない担当者による主観的検査を実施して確認する。

146. 病理組織学的検査における定量的評価（すなわち形態計測データ）は、定性的病理組織検査、脳重量、体重の変化によっては予測できない神経毒性作用に対して感受性のある評価項目であることが示されている（Raffaele *et al.*, 2005）。試験ガイドライン 426 に、大脳皮質の主要部位、海馬、小脳の計測が規定されているが、計測部位の追加、または脳の特定領域を標的とした計測が必要となることもある。解剖および組織の処理を慎重に行って、形態計測に適した同じ平面の切片を作製することが重要である。有意義な指標とするためには、切片が相同である必要がある（包埋法の標準化、目印とする解剖学的部位および細切面の一致によって達成される）。計測にはさまざまな方法が用いられるが、通常は直線計測法が用いられる。デジタル技術の利用が進んだことで、定量的検査法の精度が向上して標準化が推進され、データの保管と検証が容易になったことは特筆すべき点である。代替定量法（面積測定、細胞計数など）は、形態計測によって認められた投与関連所見について、さらにその特徴を明らかにすることができる（Garman *et al.*, 2001）。

データ解釈

神経行動学的データの解釈における問題点

147. 行動評価項目を解釈する際は、行動評価項目間の相関性を認識しておくことが重要である。たとえば、学習や記憶の試験課題遂行に必要な能力に影響を及ぼすような運動機能や感覚機能の障害があれば、学習および記憶の検査結果に影響する。

148. 学習および記憶に対する影響を評価するために、さまざまな試験が用いられてきた（表 6）。その多くは動物の動きに基づいた試験や感覚を用いた試験などであるため、学習試験で認められた機能障害は、学習障害とは別の行動学的影響によるものである可能性がある。学習または記憶の評価時に認められた明らかな投与関連機能障害が、交絡因子によるものであると考えられる場合には、適切な統制手続きやフォローアップ試験により、その所見についてさらに詳しく調べる。たとえば、直線路を用いた遊泳速度の測定は、水迷路課題における運動機能障害と学習障害または記憶障害との識別に非常に有用である。また、記憶の検査結果の解釈は、同一試験で得られた学習の検査結果なしには解釈できないということに留意しておくことが重要である（US EPA 1990）。

149. 行動評価を行う時期については、特に児動物に被験物質を直接かつ連日投与する場合には、被験物質が行動評価項目に対して急性作用を生じる可能性を考慮しなければならない。投与期間中に神経行動学的指標に変化が認められた場合、被験物質の動態によっては、投与経路に関係なく全身作用と発達に対する作用とを識別できる可能性が限られる。発達神経毒性試験の標準的な試験デザインでは、神経行動学的検査は離乳時期、性成熟期および若齢成熟期（機能観察バッテリー試験（functional observational battery）のみ）、ならびに試験終了直前の生後約 60 日に行うが、投与については生後 10 日または生後 21 日を過ぎたら中止する。

150. 生後まもない出生児に観察された作用について、成熟後に明らかな可逆性が認められた場合、それは代償性発育または代償性行動によるものであって、真の回復を示すものではない。同様に、若齢期には観察されなかった所見が成熟期に認められた場合、神経発達初期の変化の潜伏性発現を示すものである可能性があるため、不一致として無視してはならない。また、明らかに

可逆的な作用についても解釈しなければならない。可逆的な作用は曝露継続時、または急性作用が懸念される場合に問題となるのに対し、不可逆的な作用は明らかに重篤である。作用が真に可逆的なものであるか否かの決定は難しい場合が多いということも認識しておかなければならない。神経系は可塑性を有するため、個体は被験物質に関連した作用を相殺することができるが、結果として生じた神経系機能の変化については有害作用とみなさなければならない。発達神経毒性が一生のうち一時期にのみ観察される場合は代償作用を疑う。また、たとえば学習課題の初期に認められ、後期には認められない学習効果については、可逆的な作用とは解釈しない。むしろ、そのような結果は学習速度の低下を示すものである。

151. 投与計画（treatment paradigm）が行動学的検査に影響を与える可能性について考慮しなければならない。被験物質の薬物動態学的および薬力学的特性によっては、1 日のうちの投与時間帯に対して検査の時間帯をいつにするかで結果が大きく左右される。また、投与期間中に認められた神経行動学的指標の変化が、本質的に全身性（systemic）のものか発達に関するものであるのかの決定は不可能であることも多い。通常、そのような識別については、薬力学的試験、作用機序試験（mode of action）、段階的投与（phased dosing）および里親哺育などの実験を追加実施して初めて確証できるものである。このことは、発育期の高感受性集団に対する潜在的リスクを評価する場合には、通常は必要不可欠であるとみなされない。それでも、有害性に基づいて化学物質を分類する上では重要である。

行動評価項目における試験の妥当性

152. 試験の妥当性（validity）はその試験が何を測定し（関連性(relevance)）、どの程度適切に測定したか（信頼性(reliability)）にかかってくる。陽性対照と考えられるもの、すなわち行動催奇形物質（behavioural teratogen）を用いた行動学的検査の妥当性検証試験が数回実施されている。複数の欧州の研究機関がグルールとなり実施したメチル水銀とエタノールについての行動奇形学的研究（European Inter-laboratory Study group on Behavioural Teratology）では、メチル水銀で最も明確な結果が示されている（Elsner *et al.*, 1986, Elsner *et al.*, 1988）。また、プロポキシフェン、クロルプロマジン、ビタミン A が異なる行動奇形学的特性を生じることから、それらについて調べた研究では、使用されたテストバッテリーニ関して、ビタミン A が陽性対照物質として最適であると判定されている（Saillenfait and Vannier 1988）。

153. 試験の妥当性に関して重要な点は試験の関連性（relevance）であり、その試験が計画どおり測定すべきものを測定しているかどうかである。たとえば、学習試験における変化は常に学習能力に対する影響のみを表すだろうか。答えは否である。学習能力以外にも動機付けおよび活動レベルが学習試験の成績に影響しうるためである。したがって、そのような場合には学習試験の結果に対するこれらの因子の影響を評価してから学習能力に関する結論を出さなければならない。

154. 一連の試験に含まれる各行動学的検査は、それぞれ妥当性が立証された方法でなければならず、陽性対照物質については文献上にその例を見ることができる。たとえば、2,5-ヘキサンジオンはローターロッドにおける影響を見るための陽性対照物質として用いられる（Ladefoged *et al.*, 1989）。また、握力測定装置では、トリエチルスズ、アクリルアミド、2,5-ヘキサンジオンについては影響が見られるが、インスリン、四塩化炭素、ハロペリドールの影響は認められないことが示されている（Gerber and O'Shaughnessy 1986）。このことは、握力測定法が既知の神経毒性物質に感受性を有し、これらの神経毒性物質の影響と、血漿グルコース濃度、肝臓機能、中枢神経系のドーパミン遮断作用による影響を識別しうることを示している（Crofton *et al.*, 2007）。

試験感度

155. 試験感度は生物学的な種間差異の影響を受けるが、このことは行動学的検査を行う上で重要である。ただし、行動評価項目の多くは同一動物で複数回測定を行う、すなわち反復測定デザインであるため（Tamura and Buelke-Sam 1992）、そのようなデータを反復測定分散分析にて解析する場合は、個体間変動と反復による変動が考慮に入れられることになり試験の感度は増加する。

156. 試験の感度は、評価の対象となっている機能の変化の検出力にかかってくる。行動学的検査結果を評価する時に、その検査法の複雑さと被験動物の能力との関係を検討することが重要である。被験動物の能力を過小評価すると、検査法が簡単すぎて感度が低いものとなり、その試験で適切な遂行成績を示すことができないのは大きな障害を負った動物だけとなる。過大評価をした場合は、対照群もその課題を正確にこなすことができないため、曝露群と優劣がつけられないこととなる。例を挙げると、キシレンによる児動物の出生前曝露試験で、モリス水迷路の生物学的感受性は標準飼育条件下の児動物では十分であったが、環境エンリッチメントを取り入れた飼育条件下の児動物では非常に低感受性（試験が簡単すぎた）であった（Hass *et al.*, 1995）。また、トリブチルスズの出生前曝露を受けたラットを放射状迷路およびモリス水迷路の 2 つの空間学習課題に供したところ、放射状迷路では明らかな学習遅滞が認められたが、モリス水迷路では差が認められなかった（Gårdlund *et al.*, 1991）。ただし、この場合、モリス水迷路については各試行時の開始地点と避難場所を変えず、また非常に短期間の日程で行われたために、感度が低下していた可能性もある。

157. モリス水迷路、8 字型迷路、オープンフィールド試験、ローターロッド試験など、行動学的検査の多くはおよそ対照値から 30～50%の差を検出することができることが、種々の研究によって示されている（Hass 1993、Reiter and MacPhail 1982）。1 群当たり 16 匹を用いた研究で、検出係数、すなわち与えられた有意水準（0.05）を用いて有意な変化を検出するのに必要な変化率は、聴覚性驚愕反応馴化試験では 7～23%、オペラント条件付け用いた試験では 2～26%、複合水迷路試験では 13～20%であった（Vorhees 1985）。これらの結果は、生きている動物の行動評価指標には、標準化された条件下で実施されれば他の科学的評価指標と同等の正確があるということを示している（Annau and Cuomo, 1988）。

158. しかしながら、たった 1 種類の陽性対照物質では、すべてのタイプの行動に対応することは期待できない（Saillenfait and Vannier 1988）。したがって、テストバッテリーとして使用される各検査法について、陽性対照物質のデータを検査感度を示すものとして使用することができるが、陽性対照データは必ずしも出生前曝露試験によるデータである必要はない（US EPA 1990, Crofton *et al.*, 2007）。

159. 行動的影響の程度および性質について考慮しなければならない。一般的に、単独の、または複数ではあるが相互関連のない変異よりも、一定の作用パターン（一連の試行反復時に認められる学習障害など）が得られた場合の方が、発達神経毒性の証拠として説得力がある。しかし、ガイドラインで指定されている各行動評価項目は、それぞれ動物の異なる行動機能を評価するよう設定されていることを認識することが重要であり、異なる検査項目にまたがる障害が認められないからといって、個々の課題で検出された障害が投与に関係ないものと解釈してはならない。

神経病理学的評価項目

160. 通常、神経病理学的検査では、発生率データ（病理組織学的所見など）か、連続データ（脳重量、形態計測評価など）として結果が得られる。データ解析に用いられる統計学的手法としては、分散が均一なデータであれば分散分析（ANOVA）と *t* 検定（Dunnett の方法など）、不均

一分散データであればノンパラメトリック分散分析がある。多変量分散分析などの他の方法も投与、性別、用量の相互作用の解析に有用である。カイ二乗検定、Fisher の直接法などのノンパラメトリック法は発生率データの解析に利用できる。統計解析の基本単位として各腹を用いることが、極めて重要であるため（Holson and Pearce, 1992）、試験デザイン（すなわち試験に供する動物の選択および割り付け）にもデータの統計学的評価時にも、このことを考慮すべきである。得られたデータの統計解析時に、脳重量と病理組織学的検査結果については、標本の大きさ（すなわち 10 匹/性/群、1 腹につき雄または雌 1 匹、異なる腹を代表する場合もある sometimes representing different cohorts of littermates）が原因で統計的検出力が限られたものになることを認識することが重要である（Tilson, 1994）。性差による違いが認められなければ、雌雄合わせると標本の大きさが 1 群 10 匹から約 20 匹に増加して検出力が上がるため、雌雄まとめてデータ解析を行っても良い。データの統計処理に正確を期する場合、または解析時に生じた困難な問題もしくは特殊な問題を解決したい場合には、統計の専門家に相談すると良い。

161. DNT（訳者注：developmental neurotoxicity test）試験では機能的評価項目および行動評価項目が評価において使用されるため、有害な神経病理学的変化として観察されることとなった毒性学的有害作用の起始および時期を正確に特定することは通常不可能である。成熟動物を用いた試験結果との比較から、この点に関して何らかの情報が得られることもあるが、常に得られるわけではない。たとえば、DNT 試験では母動物の剖検データは収集しないし、他の神経毒性の研究から急性または反復曝露を受けた成熟ラットのデータが得られることがあるかもしれないが、発達段階における臨界期に曝露を行うという点で DNT 試験の被験体は特異なものとなっている。DNT 試験の剖検時データの解釈を困難にしている要因は他にもある。たとえば、病理組織学的検査において有害作用を同定する際には、発生過程で生じる事象にはプログラム細胞死などのように、形態的に類似した特徴を示す投与関連事象との鑑別が難しいものもあることを考慮に入れることが重要である（Shield and Mirkes, 1998）。また、神経発達初期の事象が阻害されると、成長と分化に変化が生じ、細胞死や関連する他の組織の反応を伴わない神経系の構造変化を生じる（Garman *et al.*, 2001）。神経病理学的データの標本の大きさが限られているために、群間で反応に統計学的有意差が認められないことがある。このため、定性的検査所見については、発生率が低値で統計学的有意性が認められなくても、生物学的または毒性学的意義を有する可能性があることを認識しておくことが重要である。生物学的意義の解釈には試験実施施設の背景データおよび陽性対照データが有効である。試験結果の解釈をさらに複雑にしていることは、早期に屠殺した出生児の神経病理学的検査で認められた所見には、それ以降の所見と関連性があるものもあれば、ないものもあるということである。後期の所見は早期（すなわち発生期）曝露による潜伏性の作用が表れたものであり、潜在的に投与と関連している可能性があるため、無視してはならない。ただし、後期に有害作用所見を認めないからといって、それ以前の所見を軽視してよいわけではなく、また、必ずしも有害な発生影響から回復したことを示すわけでもないのであって、その理由は、（i）早期の検査で認められるような発生過程または発生事象における変化は成熟期には生じるもはや生じなくなっているかもしれない、（ii）動物の成長に加えて大規模な脳の再構築が生じることによって成熟期の構造異常を示す所見が隠されてしまうことがある、（iii）（同じ同腹児母集団からのものではあっても）各検査時点で用いられるのは別の組の動物であることが、潜在的な反応のバラツキの原因となっている可能性があるためである。成熟後の児動物に認められた投与関連所見については、DNT 試験では若齢成熟期（2 カ月齢）に屠殺して検査を行うことから、正常な加齢に伴う退行変性過程（げっ歯類の慢性試験で投与 2 年後に認められるような所見など）とは無関係であることに留意することも重要である。最後に、神経病理学的変化は、投与に関連して認められた行動学的検査結果とは関係がない可能性もある（Jensen and Catalano, 1998）。観察された神経病理学的変化と行動学的変化との間に何らかの関連性があるようなら、関連性の評価も行わなければならない（OECD, 2007a）。

母動物の評価項目および児動物の評価項目の関連性

162. 母動物および児動物の評価項目については、特に児動物が母動物に頼っている授乳期間で関連性が認められることがある。投与を受けた出生児を未投与の母動物に、未投与の出生児を投与を受けた母動物に交差育成させることによって、その関連性を立証することが可能である。母体毒性を介した作用であっても、母体毒性を介した作用でなくても、発生毒性における NOAEL を設定する際には考慮に入れなければならない。

同腹児数、出生児の性別および平均体重

163. 出生児の体重に対する同腹児数の影響については、これまでによく分かっている。通常、少なくとも離乳前の出生児の体重は同腹児数に反比例する。この関係は離乳後には認められなくなるものではあるが、児動物の体重データなどの統計解析に際しては、児動物数の標準化（つまり同腹児数がそろうように間引くこと）は常時実施されるわけではないので、同腹子数により調整して解析すべきである。性別も体重および体の大きさに影響する。性による影響の差を考える際には、体重または体の大きさが間接的に関係する指標について、体重または体のサイズの潜在的な役割について考慮する。また、関連性があれば、体のサイズ性差に合わせて試験を改良する（平均台試験における平均台の幅など）。通常、離乳期前は雌よりも雄の体重の方がわずかに重いだけであるが、性成熟期以降、その差が著しく開く。児動物の統計解析に雌雄ともに含まれる場合、性別による解析も入れる。通常、児動物の体重に統計学的に有意な変化を認めれば、有害作用とみなす。

身体的発達および機能的発達の指標

164. 初期の身体的発達および反射の発達の遅延は授乳期の児動物体重の減少と合わせて認められることが多く、それ自体は発育遅延の徴候である。曝露を受けた児動物の身体的発達または機能的発達における変化に対応した成長面の変化が伴わない場合、その変化は特異的な発生影響を反映するものであり、通常の発育遅延ではない可能性がある。

165. AGD は個体の体のサイズの影響を受けるため、データの評価時にその点を考慮しなければならない。児動物の体のサイズや体長は通常計測されるものではなく（頭臀長を計測するときもあるが）、体重の測定が行われている。肛門・生殖結節間距離インデックス、すなわち AGD を、体重で除したものを使用する場合もある。しかし、児動物の体重はバラツキが大きいため、（訳者注：体重で除す補正によって）肛門・生殖結節間距離インデックスのバラツキが大きくなることがある。これにより、AGD に対する実際の影響が隠れてしまう可能性があるため、この補正法は推奨できない。代わりに、動物の体のサイズに関する共変量を用いて、その影響を考慮した解析を行う。体重を（訳者注：共変量として）用いても良いが、この指標が 3 次元的指標であるのに対し、AGD は一次元的である。それゆえ、最適な共変量は体重の立方根であると考えられる（Clark, 1999）。動物の体のサイズによって説明がつかない統計学的に有意な変化が AGD に認められたら、曝露による影響が示唆されるため、それを用いて NOAEL を設定する。

IV. 成熟動物の生殖に関する評価項目

評価項目の概要

166. 「生殖・発生毒性スクリーニング試験」（TG 421）、「反復投与毒性試験と生殖・発生毒性スクリーニング試験の複合試験」（TG 422）、「一世代生殖毒性試験」（TG 415）、「二世代生殖毒性試験」（TG 416）など、生殖関連項目の評価を含む試験ガイドラインは多い。これらの試験はすべて、成熟動物への曝露後の成熟期の生殖器系についてなんらかの情報を提供する。二世代生殖毒性試験では、さらにあらゆる発達段階で曝露した後の生殖器系についての情報が得られる。すなわち、F1 世代は子宮内、授乳期間、性成熟期から成熟期を通して連続曝露を受けることになる。最近の OECD ガイダンス文書案には雄および雌の生殖器官、ならびに膣垢塗抹標本の病理組織学的評価による情報も含まれており、これらの試験の評価および解釈時に非常に有用なものになると思われる（OECD, 2008）。

方法論的問題点

雄性生殖器の検査

167. 肉眼検査による雄性生殖器の評価はほとんどの生殖毒性試験ガイドラインに含まれている。精巣、精巣上体、精嚢、前立腺および下垂体の絶対重量および相対重量（すなわち、体重で補正した重量）について記録する。精巣の適切な組織検査法については多くの文献で総説を読むことができる（Russell *et al.*, 1990、U.S. EPA, 1996、Chapin and Conner, 1999、Lanning *et al*, 2002、Creasy, 2003、Latendresse *et al.*, 2002）。精巣および精巣上体の組織検査に用いられる方法は、それらの組織検査が含まれる各生殖毒性試験法に則る。精子評価とともに精子形成期の全期間を通じた雄への投与を規定している試験法（TG 416 など）では、投与計画がより短期間で精子評価も実施しない試験方法（TG 421、422 など）と比べて組織検査の範囲は狭いものとなっている。一世代生殖毒性試験（TG 415）および二世代生殖毒性試験（TG 416）では、精巣、および頭部、体部、尾部を含む精巣上体について適切に固定・薄切した後に組織学的検査を行う。枯渇（上皮内生殖細胞数の減少）、セルトリ細胞の空胞形成、多核細胞（精母細胞および精子細胞からなる巨細胞）の形成、精細管内における生殖細胞死および生殖細胞層の異常配列、上皮における円形精子細胞および細胞残屑の存在など、精巣変性について多面にわたり認めることができる。

168. より短期間の試験（TG 421、422 など）では、精上皮周期の特定の段階で精巣が損傷を受けた可能性について調べることができる。確実に性成熟期に達していることを確認するために、投与開始時の被験動物の日齢を考慮することも重要である。ラットの 14 段階の成長ステージを分類する最も一般的な方法については、Leblond および Clermont（1952）が先体の発育および伸長精細胞の核の形に基づいて解説している。標準的な毒性試験では、精子形成の全段階における正常性について定性的評価を行うだけで十分であり、各管腔断面を用いて精子形成段階を定量的に分類する必要はない（OECD, 2008）。

精子パラメータ

169. 二世代生殖毒性試験（TG 416）で採用されている検査項目に、精子の数、形態および運動性がある。精子の適切な検査法については多くの文献で総説を読むことができる（US EPA, 1996、

Seed *et al.*, 1996、Chapin and Conner, 1999）。精子の運動性については、精巣上体尾部または輸精管から精子を採取して評価に用いる。採取時、室温から 37°C の気温であれば、採取場所を問わず、また、緩衝生理食塩水中にて最長 1 時間まで保存可能である。精巣上体尾部から採取する場合、緩衝生理食塩水を入れたシャーレ内に組織をとって、カミソリの刃などで精巣上体尾部に数カ所切り目を入れる。精子が生理食塩水中に拡散したら組織を取り出す。輸精管から採取する場合、緩衝生理食塩水を入れたシャーレ内に組織の一部を入れると精子が生理食塩水中に拡散してくる。いずれの場合も、組織への余分な操作を避けるよう注意する。

170. 精子の運動性の評価は手動またはコンピュータによる精子運動能自動解析装置（CASA）により行う。精子の運動性は室温によって決定的に決まるため、形態および精子数に優先して評価を行う。運動性の解析は試料を 34～37°C に保つ。正確な運動性の評価にはチャンバーの深さが重要である。げっ歯類の場合、深さが 20 µm を超えるものが望ましい。通常、200 個の精子について評価を行う。対照群における運動性の要件は最低 70%とする。手動で検査を行う場合は血球計を用いて運動精子数を計数する。簡単なのは、不動精子数を計数後、試料を固定してから総精子数を計数する方法である。CASA にて運動解析を行う場合、精子の進行方向性速度の平均値がユーザーの設定した閾値よりも高ければ運動精子として計上される。閾値は所定の試験機関において決定する。

171. さらに、前進運動精子率を求める。この指標は、前進運動をしている精子と小刻みな不規則運動を繰り返しているだけの精子とを識別するものである。手動検査によって計測する。CASA にて解析を行う場合、前進運動については、1 種類または複数の CASA データについてユーザーが設定した閾値を超えた精子を計上するが、通常は、進行方向性速度の平均値と直進性、または直線速度のみを尺度とする（Seed *et al.*, 1996、Perreault and Cancel, 2001）。

172. 精子の形態については、精巣上体尾部または輸精管から採取した試料を用いて評価を行う。緩衝生理食塩水に含まれるウシ血清アルブミンが精子の形態評価に用いられる染色法の妨げになるため、通常、精子の運動性の評価とは別に採取された試料を用いて形態評価を行う。形態評価用の試料を少量スライドグラスの上にとって湿潤標本として観察するか、風乾してから観察する。風乾によって精子の尾部にねじれが生じるため、湿標本による観察を好む機関もある。標本の染色には通常、エオジン Y を用いるが、精子の適切な観察が可能であれば、その他の各種染色法も許容される。標本の観察は光学顕微鏡下にて 400 倍で行う。

173. 精子の形態について、統一された分類法はない。通常、精子の形態異常は頭部および尾部の奇形について記録され、例として、頭部の矮小化、または極端な鎌状形態、尾部の切断およびらせん化、多尾、ミトコンドリアの変位、細胞質小滴の遺残などが挙げられる。また、頸部と頭部および尾部との接合部は脆弱な部分である。尾部の異常が比較的高いと他の異常の増加が隠れてわからなくなるため、尾部の異常を含む発生率と含まない発生率の両者を求めると有用である場合が多い。げっ歯類では正常精子率が高いため、通常はラット 1 匹につき 200～400 個を評価対象とする。対照群、投与群間の正常精子率の差異に関する統計的検出力は高い。

174. 二世代生殖毒性試験（TG 416）では、精巣上体尾部および精巣の精子数を測定する。精巣上体尾部の精子数は精子貯蔵能の指標となる。精巣の精子数については通常、界面活性剤および均質化抵抗性精子細胞の頭部を計数する。精子細胞核は精子形成期に高度に濃縮される。成熟精子細胞の核は未熟な精子細胞と違い、均質化抵抗性が非常に高い。このため、上記の方法によれ

ば、精子形成後期の精子細胞数を適性に推計することができる。精子数の計数は剖検時に行うか、凍結保存組織を用いて行う。精巣上体尾部、精巣のいずれの場合も、Triton X-100®などの界面活性剤の存在下で組織を均質化する。精子数の計測には血球計、細胞計数分析装置、または CASA を用いる。測定値の変動および統計的検出力の水準については各試験機関で決定する。

雌性生殖器の検査

175. 肉眼検査による雌性生殖器の評価は生殖毒性に関する全試験ガイドラインに含まれている（TG 414、421、422、415、416）。子宮の着床痕数について記録する。卵巣、子宮、腟および下垂体の絶対重量および相対重量（すなわち、体重で補正した重量）について記録し、これらの臓器の病理組織学的検査を行う。卵巣の詳細な組織検査時には卵胞、黄体、間質の区分だけでなく、胚上皮および卵巣支質について観察する。

卵母細胞の定量化

176. 二世代生殖毒性試験では、F1 雌動物について原始卵胞の定量的評価を行う。適切な固定法、染色法および定量分析法については、近年になって Heindel（1999）によって総説が出されている。卵胞は以下の 3 つに分類される。

- 未発育の原始卵胞：単独卵母細胞、または卵母細胞が部分的もしくは未破裂の単層顆粒膜細胞に囲まれている卵胞。
- 発育卵胞：卵母細胞が多層化した充実性の顆粒膜細胞層に囲まれている卵胞。液体に満たされた腔（卵胞腔）は認められない。
- 胞状卵胞：内部に卵母細胞を有し、液体に満たされた卵胞腔が発達した顆粒膜細胞層で囲まれている。

177. 試験ガイドライン 416 の標準試験法では、各投与群につき雌 10 匹の評価を行い、卵胞が最初に観察された連続切片から 10 枚目ごとに卵胞数を計数する。しかし、計測する切片数、または評価動物数がそれより少なくても同等の統計的検出力が得られたとする報告もある。動物および切片の適正数については、各試験機関で決定する。

腟細胞診

178. 二世代生殖毒性試験では、母動物および F1 世代の雌の腟細胞診を行い、発情周期の長さおよび正常性について検査する。近年、検査法に関する検証を Cooper および Goldman（1999）が行っている。また、OECD（2008）も腟垢塗抹標本の作製と解釈に関するガイダンスを公布している。正確な発情周期を求めるためには、腟スメアを 1 日 1 回、少なくとも 2 週間は採取しなければならない。腟スメアの採取方法はいくつかある。その 1 つは、点眼器に水または生理食塩水を約 0.25 mL 入れて腟部を洗浄し、洗浄液を点眼器に回収するという方法である。その際、腟内へ挿入するのは点眼器の先端部だけとする。操作による偽妊娠を確実に防ぐため、子宮頸部を刺激しないようにする。腟洗浄液を顕微鏡用スライド上に均一に載せる。すぐに約 100 倍でスメアを観察し、その後の検査のために染色するか、湿潤標本の観察後に保存してから染色を行う。腟垢塗抹標本は常法（トルイジンブルーO 溶液、メチレンブルー溶液など）によって染色する。

179. 細胞分布が均等でない可能性があるため、腟スメアすべてを調べる。観察した細胞の種類によって腟スメアの評価を行う。ラットの発情周期は 4、5 日間である。1 周期は発情後期および発情間期（2～3 日）、発情前期（1 日）、ならびに発情期（1 日）の 3 期に分かれる。発情間期中の腟スメアには各種細胞が混在する。白血球が優勢であるが、量の多少はあれ、角化上皮細胞も認められる。発情前期の腟スメアは円形の有核上皮細胞塊が優勢であり、発情期は角化細胞主体である。試験機関によって腟スメアの分類基準は異なるが、雌の発情周期がわかる腟スメアを比較するためには雌動物の検査を長期間にわたって行うことが重要である。また、変動を減少させ、周期パターンを特定するために、腟スメアの採取は毎日同じ時間帯に行うことも重要である。

生殖能

180. 生殖能とは雄および雌動物が正常に交配し、生存能力のある児動物を生みだすことのできる能力のことであり、試験ガイドライン 421、422、415 および 416 によって評価される。通常、生殖能に関する情報については、試験から得られたデータに基づいて各種の比率を求め、それを指標とする。用いられる指標の種類は多く、試験機関によって異なるが、主要な指標について下表に示す。

表 7. 各種生殖指標の算出方法

指標	計算式	定義
雄の交尾率 Male Mating Index	$\frac{\text{交配が確認された雄の数}}{\text{同居させた雄の総数}} \times 100$	雄の交配能の指標
雌の交尾率 Female Mating Index	$\frac{\text{精子陽性の雌の数}}{\text{同居させた雌の総数}} \times 100$	雌の交配能の指標
雄の受胎率 Male Fertility Index	$\frac{\text{雌を受胎させた雄の数}}{\text{同居させた雄の総数}} \times 100$	雄における卵子との受精が可能な精子産生能の指標
雌の受胎率 Female Fertility Index	$\frac{\text{妊娠した雌の数}}{\text{精子陽性の雌の数}} \times 100$	雌の妊娠能の指標
出産率 Gestation Index	$\frac{\text{生存産児を有する雌の数}}{\text{妊娠した雌の数}} \times 100$	1 匹以上の生存児動物が得られた妊娠についての指標
生存率 Survival Index	$\frac{\text{生存児動物数（所定時点における）}}{\text{産児数}} \times 100$	授乳期間中の複数の時点で算出する児動物の生存に関する指標

<u>データ解釈</u>

雄性生殖器

181. 雄性生殖器の絶対重量の減少は体重増加量の減少に関係なく生じる可能性があるため、絶対重量および相対重量の両者について検討しなければならない。ただし、体重に対する影響が明らかな場合はデータの解釈を慎重に行う。精巣重量の個体間変動は低いことから、精巣の絶対重量に有意な変化（増加または減少）が認められた場合、有害作用が示唆される。精巣重量の変化は、精細管の損傷を初め、浮腫、炎症、細胞浸潤、ライディッヒ細胞の過形成、輸出管の閉塞による液体貯留などさまざまな原因によって生じる。前立腺および精嚢の重量については、アンド

ロゲン依存性であるため、精巣機能または内分泌状態の変化が反映されやすい。下垂体重量も生殖状態の指標となりうる。ただし、下垂体には、生殖とは関係のないものも含めて、多くの生理機能の調節に関わっている細胞が各種含まれている。このため、影響が認められる細胞の種類を同定するためには、性腺刺激ホルモンに特化した病理組織学的評価を行うとよい。

182. 生殖系組織の病理組織学的評価は比較的感受性の高い損傷指標となり、雄に対する生殖毒性の検査に有効である。通常、病理組織学的所見は定性基準に基づいて分類され、データは投与群ごとに影響が認められた動物の数として表される。さらに損傷の程度を定量化するための方法は標準化されていない。通常、病理組織学的所見は、具体的な試験手順に照らしつつ、精子パラメータおよび生殖能のデータと併せて検討しなければならない。用いた試験手順と損傷の程度によっては、病理組織学的所見と妊孕能の間に明らかな関連性が認められないこともある。投与期間が精子形成サイクル全体を含まない短期試験（すなわち TG 421 および 422）については、精子形成に対する影響が妊孕能に影響する精子数の減少として顕在化するのに、十分な時間でないかもしれない。通常、投与に関連して有意な病理組織学的変化が認められた場合、いかなる変化であっても、ヒトで同様の影響が認められる可能性を示唆するものである。

精子パラメータ

183. 精子の運動性に関しては、通常、2 つの指標を算出する。運動精子率は運動精子数を総精子数で割って 100 を掛けたものであり、前進運動精子率は前進運動精子数を総精子数で割って 100 を掛けたものである。精子の運動性と妊孕能とに関連性があることは研究結果から認められてはいるが、運動性の変化についてどの程度で有害とすべきであるのかという点に関しては、一般的な合意を得た基準がない。また、他の精子パラメータおよび病理組織検査のデータだけでなく、具体的な試験手順についても検討しなければならない。精子の運動性は精巣および精巣上体の機能によるため、組織学的に明らかになった精巣の障害が精子の運動性の変化に関係している可能性がある。ただし、精巣上体または精子そのものに対する直接作用によっても運動性は変化する。このため、組織病変が認められないからといって、運動性の変化を無視すべきではない。通常、精子の運動性に認められた用量反応性および統計学的に有意な変化については、ヒトの妊孕能に対する潜在的影響を示すものと解釈する。

184. 精子の形態と妊孕能とに関連性があることは研究結果から認められているが、形態の変化についてはどの程度で有害とすべきであるのかという点に関しては、一般的な合意を得た基準がない。精子の形態の評価に際しては、精子の運動性および精子数のデータ、ならびに病理組織学的検査結果を共慮しなければならない。程度が十分であれば、組織病変も精子の形態に影響する。ただし、正常な精子形態の構築にはプロタミン類の構造の正確性など非常に多くの要因が関わっているため、組織病変が認められないからといって精子の形態変化を無視してはならない。

185. 通常、精子数に関するデータの表現形式には 2 通りある。射精精子数に関するデータは精巣上体尾部ごとの精子数と精巣ごとの均質化抵抗性精子細胞数から得られる。組織の効率に関するデータは精巣上体尾部 1 mg あたりの精子数と精巣 1 mg あたりの均質化抵抗性精子細胞数から得られる。これまでの研究で検査対象となった全動物種において、精子数と妊孕能に強い関連性があることが実証されている。また、重複するが、全体を考察する際に他の精子パラメータおよび病理組織検査のデータを考慮しなければならない。程度が十分であれば、精巣病変が精子数に反映されるが、組織病変が認められないからといって精子数の変化を無視してはならない。同様

に、とりわけげっ歯類の試験では、精子数の減少が妊孕能の低下を生じるわけではない。これは、ラットおよびマウスでは射精時の精子数が非常に多いためであり、それゆえ、精子数が妊孕能を変化させるためには 90%の減少を認める必要がある。ヒトの精子濃度については変動が大きく、また、一般的にげっ歯類よりも低いということを念頭に置いておくことが重要である。精子数の分布について言えば、精子濃度が WHO の示す妊孕能の参照値に近いか、それを下回っている男性は多い。このため、ある集団において、精子濃度に少しでも減少が認められれば、その集団の妊孕能もシフト変移して、不妊または低受精能の範囲に移ってしまう男性も出てくると予想される。したがって、げっ歯類を用いた試験で統計学的に有意な変化が精子数に認められた場合、ヒトの妊孕能に対する潜在的影響を示すものと考えられる。

雌性生殖器

186. 雌性生殖器の重量データおよび病理組織学的検査結果は、具体的な試験手順を考慮に入れつつ、生殖能と併せて評価を行う。子宮重量は発情周期に合わせて 3～4 倍の変動がみられ、エストロゲン分泌量の増加に反応した液体貯留が認められる発情前期で最高となる。このため、ステロイド産生と周期性を阻害する物質では子宮の委縮が認められる。またその逆に、エストロゲン様作用を有する物質では子宮の高度拡張が認められる。発情周期が保たれた動物から採取した子宮の重量については分散が大きいため、データと発情周期との相関性を検討しなければ影響の大きい物質関連作用しか実証されないことになる。同様に、子宮の組織についても発情周期および妊娠によって変化がみられる。子宮内膜はエストロゲンおよびプロゲストゲンによって変化するため、同様の作用を有する物質に反応すれば、肥厚および過形成を生じ、ステロイド産生を阻害する物質に反応すれば、発育不全および萎縮を生じる。影響が発育期に生じれば、性成熟遅延および生殖器の低形成が生じる。段落 64 で述べたように、着床痕数に関するデータは、生存産児数および吸収胚数に関するデータと併せて解釈しなければならない。

187. 試験ガイドライン 421 または 422 では、所定の投与を受けたげっ歯類の母動物は、通常の発情周期に戻っていない授乳期第 4 日に屠殺されるため、その卵巣の病理組織学的検査からは限られたデータしか得られない。ラットの卵巣重量については、発情周期に一致した変動が認められないことから、変化が認められれば、それは有害作用とみなされる。卵巣機能は発情周期に合わせて変化するため、病理組織学的検査によって、卵母細胞および卵胞の消失、持続性の多嚢胞性卵巣、黄体形成抑制、黄体嚢胞の形成など、さまざまな影響が明らかになる。組織学的変化のすべてが卵巣重量に影響を与えるわけではないため、卵巣重量に影響が認められないからといって、組織検査の必要性を除外することにはならないことを認識しておくことが重要である。同様に、組織病変の性質および程度によって、生殖能における随伴作用の有無が決まる。通常、投与に関連して有意な組織学的所見が認められた場合、いかなる所見であっても、ヒトで同様の影響が認められる可能性を示唆するものであるとみなされる（OECD, 2008）。

188. 卵胞数と妊孕能との関連性については多くの研究によって示されている。ヒトにおける卵胞数と更年期の開始との関連性は明らかになっているが、げっ歯類における生殖老化との関連性ははっきりわかっていない。卵胞数の変化について、どの程度で有害とすべきであるのかという点に関しては、一般的な合意を得た基準がない。卵胞数のデータの解釈は、組織検査データおよび生殖能と併せて行わなければならない。卵胞数における変化は、臓器重量における変化、または病理組織学的変化に先がけて分かることが多い。卵胞数減少の程度によって、生殖能における随伴作用の有無が決まる。卵胞数の減少が認められる場合、卵母細胞に対する直接毒性、または卵胞形成のパラクリン調節を変化させる顆粒膜細胞もしくは莢膜細胞に対する作用が示唆される。卵胞数に認められた用量反応性および統計学的に有意な変化からは、ヒトに対する潜在的影響が

示唆される。

189. 発情周期に合わせて認められる腟重量の変化は、程度はより小規模であるものの、子宮重量の変化と平行している。二世代生殖毒性試験で採取する腟スメアのデータからは、発情周期の長さ、発情持続時間、発情間期の長さ、偽妊娠発生率に関する情報が得られる。発情周期の長さは、発情周期の特定の段階を選び、その段階が再発するまでの日数を計測して測定する。統計解析は、正常な 4、5 日周期を示す群と周期異常を示す群との比較により行う。また、発情周期の特定の段階の持続時間の割合にも周期の変化が表れるため、発情間期または発情期の持続日数について比較を行っても良い。ただし、そのような解析を行う場合には、影響が表れる段階が 1 つとは限らないため、個体の周期パターンを考慮しなければならない。解析する発情段階を 1 つに限ってしまうと、影響が隠される例も出てくる。発情周期に対する影響は、生殖能にも変化を生じるが、発情周期に対する影響の性質および程度による。通常、発情周期の長さに統計学的に有意な変化が認められた場合、または発情期もしくは発情間期の延長が認められた場合は、有害作用とみなす。影響が発育期に生じれば、腟部の欠損、低形成および形成異常が生じる。腟の低形成は、外性器の過形成および AGD の変化と併せて認められることがある。性成熟期の腟開口は容易かつ有効な発育指標となる。

190. 雌のげっ歯類において、下垂体重量の変化は有害作用を示唆する。げっ歯類では、エストロゲン様物質への曝露により生じる下垂体重量の増加は腫瘍の形成に先がけて認められることが多く、高プロラクチン血症を併発し、腟スメアに持続的に角化上皮を検出する例もある。エストロゲン刺激の低下による下垂体重量の減少を認める頻度は比較的少ない。雄の下垂体の重量および病理組織学的検査については、雌と同様である。

生殖能（Reproductive Performance）

191. 生殖能のデータの解釈については、用いた各試験計画書に則って実施しなければならない。1 世代につき 2 回以上の交配を行う試験および 2 世代以上を育成する試験のデータは必ずしも同じではない。また、試験ガイドライン 421、422、415 および 416 では雌雄ともに投与を行うため、生殖能の特定の側面に認められた影響が雌に起因するものであるのか、雄であるのか、または両者に起因するものであるのかの判断は通常、不可能である。これまで述べてきたように、生殖能に関する各種指標の評価は、病理組織学的検査、精子パラメータ、卵胞数、発情周期など、入手可能な他の情報と併せて行うことも重要である。

192. 交尾率（mating index）によって、神経内分泌・性腺系の統合機能に関する情報が得られる。交尾率には、性的衝動、ホルモン失調、発情周期の乱れなど、多くの因子が影響する。交尾行動は神経系の損傷によっても、またホルモン変化によっても影響を受けるため、交尾に対する影響の原因を特定決定することはどの試験プロトコールでも不可能であろう。

193. 受胎率（fertility index）からは、雄および雌の妊娠を達成する能力について情報が得られる。妊孕能に関するデータの解釈に際しては、用いた試験手順、とりわけ雄の交配前投与期間を考慮しなければならない。雄の投与期間が精子形成サイクル全体を含まない試験（すなわち TG 421 および 422）では、生殖毒性が受胎率にあらわれないことがある。受胎率に影響が認められる場合、雌雄ともに投与を行う試験では、雌雄いずれの影響によるものであるのか特定することは困難である。妊孕能に関する情報は、病理組織学的検査、精子パラメータ、卵胞数、発情周期など、入手可能な他の情報と併せて検討する。既述したとおり、げっ歯類の雄の精子数は非常に多いため、受胎率の変化として表れるには精子数の大幅な減少を要する。このため、受胎率単独では、

どちらかというと評価項目としては感度に乏しい。

194. 出産率（gestation index）については、慎重に取り扱わなければならない。同腹児数に関わらず、すべての腹を同等に扱うため、特に感度の高い評価項目というわけではない。このため、妊娠率からは吸収胚率の増加に関する情報は得られない。

195. 妊娠期間および分娩時間の長さについては、出生児の体重およびその生存率に関するデータと併せて評価を行う。妊娠期間の有意な短縮によっても、出生児の体重および生存率の低下を招くことがある。また、妊娠期間の有意な延長は、母動物または児動物の健康に影響を与える難産（異常分娩）につながる。その上、子宮内発育の延長によって、出生児の体重は平均よりも重くなる。同腹児数の減少、または顕著な子宮内発育遅延がある場合、出生児の体重が対照値と同程度になるまで、妊娠期間がやや延長することがある。

196. 同腹児数は生殖能の重要な包括的指標である。同腹児数の減少要因には、排卵される卵母細胞数の減少、精子パラメータ（数、運動性、形態）の不良を原因とする卵母細胞の受精の失敗、着床前・着床後胚損失数、または流産数の増加などがある。このため、同腹児数に関するデータの解釈についても、入手可能な他の生殖データと併せて解釈しなければならない。

197. 性比に変化を生じる要因には、同腹児数を変えることなく受胎産物の性比を変える XY 精子の産生の選択的変化、同腹児数の減少を招く雌雄胎児の選択的消失など多くの因子がある。ホルモン状態の変化によっても、出生児の雌雄誤判別を招く雌性胎児の雄性化、または雄性胎児の雌性化（同腹児数への影響なし）が生じる。このため、性比データの評価に際しては、胚・胎児の損失数に関するデータおよび遺伝毒性試験のデータも併せて評価しなければならない。また、性比の変化は、化学物質が内分泌活性を有することを示唆するものであって、XY 精子の変化とはなんら関係がないことが多い。

198. 生存率（survival index）は児動物の生存能力を表す重要な評価項目である。授乳期第 14 日には児動物は摂餌ようになるため、その際、被験物質を直接摂取する可能性がある。生存率の低下要因には、出生前または授乳期間中の曝露による児動物の発育に対する作用、構造異常、母動物の育児放棄、泌乳量の減少など多くの因子がある。生存率データを解釈する際には、胚の着床前および着床後損失数、死産児数、死亡児動物の総数、影響が認められた児動物の腹ごとの数、および死亡パターンに関するデータについても考慮する。

V. データの欠落

199. 試験ガイドラインは危険有害性の同定および用量反応の評価のための試験としてデザインされており、リスク評価に有効な手段を提供する。本文書で解説した試験によって、化学物質の潜在的な発達生殖毒性に関する広範な評価が促進される。しかし、試験が影響を検出する感度は、本来試験のプロトコールとともに、それがどのように使用されるかによって制限される。その制限のいくつかについて解説していくが、試験結果の解釈時にこれらを考慮に入れることが重要である。

評価項目

200. 生殖発生毒性試験の試験ガイドラインで採用されている試験方法では、被験物質の作用機序または既知の標的器官特異性などの特異情報が、試験デザインまたは評価項目に組み込まれていない。

201. 試験デザイン、用量選択、ならびに生殖毒性データの解釈および外挿には、発達系における被験物質の薬物動態学的および薬力学的特性を理解すること、また妊娠期、授乳期および新生児期にいたる発達期間中のさまざまな経路を通じた直接および間接曝露の複雑性を理解することが必要不可欠である。現行の試験ガイドラインは、そのようなデータ収集またはデータ利用に対応していない。

202. 現行の発達および生殖毒性に関する試験ガイドラインは器官系の機能発達の限られた変化にのみ重点を絞っている（US EPA, 2002）。多世代生殖毒性試験では生殖器系機能について、発達神経毒性試験では神経系機能について、それぞれ広範な評価を行うが、他の機能的評価項目は含まれていない。たとえば、心臓系、泌尿器系、呼吸器系、免疫系および内分泌系の正常な機能的発達については評価しない。

203. 周産期曝露によって発がんの増強が疑われるような物質について、それらのために実施する第２段階試験（second tier testing）に特化した試験ガイドラインは存在しない。しかし、現在、国家毒性プログラムが自身の慢性毒性・発がん性併合試験に初期曝露系として周産期曝露を取り入れることを検討し始めている。

204. 発達神経毒性ガイドライン中の神経行動学的評価および神経病理学的評価によって、多くの機能領域にまたがる作用の末端がスクリーニングされる。しかし、発達神経毒性試験では、第一には通常試験に用いられる動物モデルが持つ機能に依存するためであるが、行動に関するいくつかの側面（社会的行動など）は対象とされておらず、また、ヒトにおいて重大な意味を有する高次の皮質機能に関する情報が得られない。加えて、より特化した方法で発達神経毒性を追跡して試験することにも対応していない。より特化した試験とは、妊娠期間中もしくは出生後、またはその両期間にわたって曝露を行った動物の児動物について、感覚機能または認知機能に対して比較的感受性の高い指標を用いて評価を行うような実験のことである。

曝露や影響発現の臨界期（Critical windows）

205. 現行の試験ガイドラインに基づいて試験を行った場合、発育期の多段階にわたって曝露を行うことになるが、その試験デザインでは曝露の臨界期を決定するように設定されていない。そのような情報があれば、リスク評価の対象を、高感受性を示す下位集団に絞る上で有用となると思われる。

206. 発達神経毒性試験には、試験終了時の児動物の神経行動学的評価および神経病理学的評価に先行して未投与期間（約 40 日間）が設けられているが、それ以外の試験では、発達期曝露による長期的な結果については調べられない。多世代生殖毒性試験では、各世代の動物を性成熟期まで維持するが、各個体に対して試験期間中を通して投与を行うため、出生後期間および成熟期に投与に関連した有害作用が認められても、それが、全身毒性と比較して、どれくらい発達毒性に起因するものであるのかを判断するには限界がある。

207. また、現行の試験ガイドラインの中には、老齢動物における発達期曝露の影響について着手している例はない。出生前発生毒性試験では、投与を受けた児動物を出生直前に屠殺する。一世代または多世代生殖毒性試験では、投与を受ける親動物については、生殖期間中維持するだけで、約 6～9 カ月齢で屠殺し、繁殖用に維持しない児動物については、離乳時に屠殺する。また、発達神経毒性試験では、投与を受けた児動物は生後約 70 日までしか維持しない。通常、げっ歯類を用いた慢性毒性・発がん性組合せ試験の試験ガイドラインでは、若齢成熟期（5～6 週齢）に投与を開始する。このため、これらの試験で 18～24 カ月間の投与後に行われる老齢動物の検査は、出生前または出生後の発達期曝露による潜在的影響の評価とは別物である。老齢動物にまで範囲を拡大したリスク評価の必要性は広く認識されており、そのようなリスク評価における諸問題についても特定されている（Geller and Zenick, 2005）。標準試験デザインは発達段階中の曝露が、老齢時における有害な作用を生じる潜在的影響には対応していないが、そのような問題に対処するために、多世代生殖毒性試験および発達神経毒性試験のデザインを改変すること（発育期に投与を受けた F1 動物群を維持して、老齢に達した段階で毒性評価を行うなど）も可能である。そのような試験の実施と結果の解釈には重要な課題が伴うが、その試験の実施を通じて得られる知識と見識は多大であり、リスク評価にとってかけがえのないものとなるはずである。

VI. 参考文献

Adams, J. (1986) Methods in behavioral teratology. Handbook of behavioral teratology, Riley EP, Vorhees CV (eds): New York: Plenum Press, 67-97.

Akaike, M., Tanaka, K., Goto, M. and T. Sakaguchi, (1988) Impaired Biel and Radial arm maze learning in rats with methylnitrosurea-induced microcephaly. Neurotoxicol. Teratol. 10: 327-332.

Akaike, M., Ohno, H., Tsutsumi, S. and M. Omusu, (1994) Comparison of four spatial maze learning tests with methylnitrosourea-induced microcephaly rats. Teratology 49: 83-89.

Altman, J. and K. Sudarshan, (1975) Postnatal development of locomotion in the laboratory rat. Anim. Behav. 23: 896-920.

Annau, Z. and V. Cuomo, (1988) Mechanisms of neurotoxicity and their relationship to behavioral changes. Toxicology 49: 219-225.

Anon (1967) European Commission Directive 67/548/EEC (Annex VI).

Anon (2001) European Commission Directive 67/548/EEC, updated by Commission Directive 2001/59/EC.

ATLA (2002) Alternative non-animal methods for chemical testing: current status and future prospects. 10. Reproductive toxicity. (Eds.) Worth, A.P., Balls, M., 30 (Supplement 1), 95-102.

Bammer, G. (1982) Pharmacological investigations of neurotransmitter involvement in passive avoidance responding: A review and some new results. Neurosci. Behav. Rev., 6:247-296.

Bancroft, J.D., Gamble, M. (2002) *Theory and Practice of Histological Techniques*, 5th edition, Churchill Livingstone, London.

Barlow, S. M. and F. M. Sullivan, (1975) 6. Behavioural teratology. Teratology trends and applications, Berry C. L., Poswillo D. E. (eds): Springer Verlag, 103-120.

Barlow, S. M. and F. M. Sullivan (1982) Reproductive toxicity testing in animals in Reproductive Hazards of Industrial Chemicals. An evaluation of animal and human data. Academic Press INC. (London LTD)).

Blanchard, B. A., Riley, E. P. and J. H. Hannigan, (1987) Deficits on a Spatial Navigation Task Following Prenatal Exposure to Ethanol. Neurotoxicol Teratol. 9: 253-258.

Bogo, V., Hill, T. A. and R. W. Young, (1981) Comparison of accelerod and rotarod sensitivity in detecting ethanol- and acrylamide-induced performance decrements in rats: Review of experimental considerations of rotating rod systems. Neurotoxicol. 2: 765-787.

Brandeis, R., Brandys, Y., Yehuda, S. (1989) The use of the Morris water maze in the study of memory and learning. Int. J. Neurosci., 48:29-69.

Brian JV, Harris CA, Scholze M, Backhaus T, Booy P, Lamoree M, Pojana G, Jonkers N, Runnalls T, Bonfà A, Marcomini A, Sumpter JP (2005). Accurate prediction of the response of freshwater fish to a mixture of estrogenic chemicals. Environ Health Perspect. 113(6):721-8.

付録 1 （APPENDIX 1）

用語の定義

Abortion: The premature expulsion from the uterus of the products of conception: of the embryo or of a nonviable foetus (OECD, 2001a)
妊娠中絶：胎芽あるいは生存不可能な胎児およびその付属物を、子宮内から単独では生命を維持することのできない時期に排出すること。

Anomalies: structural alterations in development that include both malformations and variations (Wise et al., 1997).
奇形（異常）：奇形と変異の双方を含む発達上の形態的変化。

Conceptus: the sum of derivatives of a fertilized ovum at any stage of development from fertilization until birth including the extra-embryonic membranes as well as the embryo or foetus (OECD, 2001a).
受胎産物：胎芽、胎児あるいは胚膜等を含む、受精から出生までの発展段階における妊娠による派生物すべて。胎児と胎盤の総称。

Developmental toxicology: the study of adverse effects on the developing organism that may result from exposure prior to conception, during prenatal development, or postnatally to the time of sexual maturation. The major manifestations of developmental toxicity include 1) death of the organism, 2) structural abnormality, 3) altered growth, and 4) functional deficiency (OECD, 2001a).
発生（発達）毒性学：受胎前、胎内での発達、あるいは出生後から性成熟期までの期間に、曝露で生じた発達中の生体における望ましくない影響に関する研究。発達毒性の発現結果としてしられている主なものには、1) 生体の死、2) 形態の構造異常、3) 成長における変化、4) 機能的障害が挙げられる。

Embryo: the early or developing stage of any organism, especially the developing product of fertilization of an egg after the long axis appears and until all major structures are present (OECD, 2001a).
胚（胎芽）:初期の発生段階にある生体、特に受精した卵に由来する発生途上の生成体で、長軸（頭尾軸）が出現してから主たる生体構造が出現するまでのもの。

Embyotoxicity: detrimental to the normal structure, development, growth, and/or viability of an embryo (OECD, 2001a).
胚毒性：胎芽の正常な形態構造、発生、成長、生存能に対して害をなす性質。

Fertility: The capacity to become pregnant or to impregnate. In humans, the capacity for producing viable offspring (IPCS, 2001).
妊孕能（にんようのう）：受胎することのできる能力あるいは妊娠させる能力。ヒトにおいては、生存児を産する能力。（訳者注：受胎能、生殖能力、受精能）

Fertilization: The fusion of the sperm and ovum resulting in the restoration of the diploid number of chromosomes (IPCS, 2001).
受精：染色体の 2 倍体への復元をもたらす精子と卵子の融合。

Foetus: the unborn offspring in the post-embryonic period (OECD, 2001a).
胎児：後胚期にある誕生前の児。

Foetotoxicity: detrimental to the normal structure, development, growth, and/or viability of a foetus (OECD, 2001).
胎児毒性：胎児の正常な形態構造、発生、成長、生存能に対して害をなす性質。

Gestation: Length of time between conception and birth (IPCS, 2001).
妊娠期：受精から出生の期間。

Hazard characterization: Involves determining whether or not an agent poses a hazard, at what doses and under what conditions of exposure (IPCS, 2001).
ハザードキャラクタライゼーション：物質等が危険有害性をもつかどうか、どの程度の用量で、どのような曝露条件で生じるか、等を決定すること。

Hazard identification: The identification of the inherent capability of a substance to cause adverse effects (IPCS, 2001).
危険有害性の同定：ある物質が有害な影響を生ずることが可能な属性をもっているかどうか同定すること。

Implantation (nidation): attachment of the blastocyst to the epithelial lining of the uterus, including its penetration through the uterine epithelium, and its embedding in the endometrium (OECD, 2001a).

着床：受精卵胞胚が子宮内膜上皮に接着すること、子宮内膜上皮へ貫入し子宮内膜内に着床する過程も含む。

Malformation/ Major Abnormality: Structural change considered detrimental to the animal (may also be lethal) and is usually rare (Wise eta al., 1997).
形態異常（奇形）：動物にとって害を及ぼす形態構造の変化（致死を含む）であり、通常は稀な現象。

No-Observed-Adverse-Effects-Level (NOAEL): Highest concentration or amount of a substance, found by experiment or observation, that causes no detectable adverse alteration of morphology, functional capacity, growth, development or life span of the target organism under defined conditions of exposure (Nordiberg et al., 2004).
無毒性量：当該物質に一定の条件で曝露をしたときに、標的器官の形態、機能、成長、発達、寿命に対して、検出可能な望ましくない変化を引き起こさない最高濃度あるいは投与量で、実験あるいは観察によって見だされる。

Perinatal: The period around the time of birth.
周産期：出生周辺の時期。

Pregnancy: The condition of having an implanted embryo or fetus in the body, after fusion of an ovum and spermatozoon (IPCS, 2001).
妊娠：精子と卵が融合したのち、その胚が着床している状態、または胎児を胎内に保有する状態。

Reproductive toxicity: includes adverse effects on sexual function and fertility in adult males and females as well as developmental toxicity in the offspring (GHS).
生殖毒性：成体の雄あるいは雌の性機能および妊孕能に対する望ましくない影響と児における発生（発達）毒性。

Resorption: a conceptus which, having implanted in the uterus, subsequently dies and is being, or has been resorbed (OECD, 2001a).

Early resorption: evidence of implantation without recognizable embryo/foetus.

Late resorption: dead embryo or foetus with external degenerative changes.

吸収：子宮に着床した受胎産物が、その後死亡し吸収されることまたは吸収されてしまった状態。

初期吸収：胚や胎児が認められない着床の証拠。

後期吸収：死亡した胚や胎児で外的な変性変化があるもの。

Risk assessment: An empirical based paradigm that estimates the risk of adverse effect(s) from exposure of an individual or population to a chemical, physical or biological agent. It includes the components of hazard identification, assessment of dose-response relationships, exposure assessment and risk characterization (IPCS, 2001).
リスク評価：個人あるいは集団が、化学的、物理的、生物学的な因子に曝露されることで生じる望ましくない影響の可能性と程度（リスク）をみつもる経験的なパラダイム。危険有害性の同定、量 - 反応関係の評価、ばく露の評価、リスクの総合判定から構成される。

Sexual maturation: Achievement of full development of sexual function and reproductive system (IPCS, 2001).
性成熟：性機能および生殖系の十分な発達が完了すること。

Teratogenicity: Induction of structural abnormality (IPCS, 2001).
催奇形性：形態異常を引き起こす能力。

Variation/Minor Abnormality : Structural change considered to have little or no detrimental effect on the animal; may be transient and may occur relatively frequently in the control population (Wise, 1997).
変異／微小な異常：動物にほとんど有害な影響をおよぼさないか全く害がないと考えられる程度の形態構造の変化；一過性であったり対照となる母集団において比較的頻繁に起こったりするもの。