

Fig. 5 パワーモデルの  $b=0.826\sim 1.15$  の場合の AUC と投与量の関係

らかの統計解析ソフトウェアを用いる必要がある。今回は SAS (Version 8.2) を用いることとする。詳細は省略するが、次のような簡単なプログラムで解析できる。

```
PROC MIXED DATA = data1;
  CLASS ID;
  MODEL logAUC = logDose/CL;
  RANDOM INT/SUB = ID;
RUN;
```

ここで、data1 という SAS データファイルに、被験者番号を表す ID、 $\log_e(Dose)$  の値である logDose、 $\log_e(AUC)$  の値である logAUC の 3 つの変数が格納してあるものとする。また、'MODEL' の行の 'CL' はパラメータの信頼区間を出力するためのオプション指定である。

解析結果は次のようになった。すなわち、傾き  $b$  は  $0.988 \pm 0.079$ 、その 95% 信頼区間は  $0.826\sim 1.15$ 。切片  $a$  の母集団平均は  $0.881 \pm 0.421$ 、個人差は  $\omega = 1.15$ 。一方、個人内誤差は  $\sigma = 0.344$ 。

前述の、個人の対応を無視した(単純な)パワーモデルでの解析と比較して、 $b$  の推定値そのものは同一(0.988)であるが、標準誤差は本解析のほうが小さい(0.266 に対して 0.079) ことがわかる。それに伴って、信頼区間の幅も非常に狭くなっている。 $b$  の信頼区間  $0.826\sim 1.15$  に対応するプロットを Fig. 5 に示す。これならば、AUC は用量比例的に増加する、と結論して構わないと考えられるが、いかがであろうか。

なお、 $b$  の信頼区間がどの範囲に入っていれば用量比例的と判断してよいか、についてのコンセンサスはまだ存在しない<sup>1)</sup>。しかし、どんなに厳しい基準を採用したとしても  $0.8\sim 1.2$  の範囲ならば文句無しに用量比例的のといって差し支えないであろう。

本解析の結果、個人差(の標準偏差)は 1.15 であるのに対

して、個人内のばらつき(の標準偏差)は 0.344 であると推定された。つまり、最初に行った単純なパワーモデル解析での誤差標準偏差(1.17)のほとんどは個人差に起因していたわけである。そのため、個人の対応を無視した解析を行うと、用量比例的かどうかの判断があいまいにしかできなかったのだが、個人の対応をきちんと考慮することによって、明確な結論が導けたことになる。

なお、今回の解析では「切片  $a$ 」、すなわち、個人差は絶対的な暴露量のみが存在し、用量との関係を表すパラメータ(「傾き  $b$ 」)は個人間(さらには被験者母集団)で一定であると仮定した。しかし、もちろん、薬物によっては  $b$  にも個人差が存在することもありうる。その場合は  $b_i$  も分散  $\omega_b^2$  の正規分布に従うと仮定して解析を行えばよい。この解析の結論は、例えば、「平均的には  $b=1$  であり、用量比例的とみなせる。ただし、 $b$  には個人差が存在し、 $b=0.5$  になる被験者も、 $b=2$  になる被験者も存在しうる」となるう。

#### 参考文献

- 1) 橋本敏夫, 笠井英史, 山田雅之, 榎秀之, 半田淳, 滝沢毅, 平田篤由(2001), 臨床薬理試験における薬物動態の線形性に関する統計学的評価. 薬物動態16: 244-252.
- 2) G. Verbeke, G. Molenberghs(訳: 松山裕, 山口拓洋)(2001), 医学統計のための線型混合モデル. サイエンスティスト社, 東京.

## レクチャーノート：母集団薬物動態解析のイロハ(7)

慶應義塾大学病院薬剤部 笠井英史

### 1. 解析バリデーションとは

今回はバリデーションの問題を取り上げる。ここでいうバリデーションとは母集団薬物動態解析(PPK 解析)の結果得られたモデルがどの程度妥当なものであるかを検証するプロセスのことであり、「解析バリデーション」とも言う。

今まで述べてきたように、PPK 解析は極めて高度な統計学的手法(非線形混合効果モデル)を用いている。そして、実は、解析結果がどの程度まで信頼できるのかという重要な点は統計学には未解決の問題なのである。従って、PPK 解析を行う際には、得られるモデルおよびパラメータ推定値の妥当性・安定性を慎重に検討する必要がある。この検討のことを解析バリデーションと呼ぶ。

解析バリデーションの手法自体はいくつか提案されているが、残念ながら、最適なバリデーション方法に関する

コンセンサスはまだない。従って、研究者がそれぞれの目的に応じた適切なバリデーション手法を選択しなければならない。今回はそのひとつの例をお示しする。

## 2. 解析バリデーションの手法

さて、モデル解析バリデーションの原則は、モデル構築の際に使用したデータ以外のデータセットを用いてその予測性を評価することである。

なぜ別のデータを用意しなければならないのだろうか？それは、得られたモデルというものは、現有のデータによくフィットするように選択されてきたものだからである。よく当てはまるように選ばれてきたモデルなのだから、そのモデルがもとのデータをどの程度予測できるのかを評価しようというのは自己言及に過ぎない。PPK解析のひとつの目的は、得られたモデルでもって、将来の(未知の)患者での薬物動態を予測しようということである。したがって、「別のデータ」での予測性を評価することが重要になってくる。

その第二のデータを用意する方法としては大きく分けて(1)外部データ法、(2)内部データ法の2種類がある。

外部データ法とは、独立な二つの試験から得られたデータがあるとして、その内の一つのデータを用いてモデル構築およびパラメータ推定を行い、残りのもう一つのデータによってバリデーションを行う方法である。解析バリデーションのそもそもの定義からして、この外部データ法がもっとも正当で望ましい方法である。しかしこの方法が利用できる状況は、少なくとも日本の臨床試験データの母集団解析においてはあまりないと考えられる。また、データの利用効率の観点からも望ましくない。

内部データ法においては独立な二つの試験データを用意する必要はない。その代わりに、ある一つのデータセットから(a)データ分割法、あるいは、(b)クロスバリデーション法もしくはブートストラップ法を用いて擬似的に「もう一つの」バリデーション用データを用意することになる。

以前はデータ分割法がよく用いられていた。データ分割法においては全データをあらかじめモデル構築用とバリデーション用の2つに分割しておくことによって、擬似的に外部データを生成する。しかしこの方法ではデータをどのように分割するかという点で恣意性が残る。また、データをどういう割合で分割するかも難しい問題である。

そのため、最近では、ブートストラップ法が比較的多く用いられるようになってきている。ブートストラップ法の一般原理を以下に示す。

- ① データを  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  とする。  $\mathbf{x}$  から推定する統計量(算術平均, 幾何平均, 相関係数, 回帰係数, 等)を  $\phi$  とする。

- ②  $\mathbf{x}$  から  $n$  個のデータを復元抽出し  $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$  を構成する。

- ③  $\mathbf{x}^*$  から目的の統計量  $\phi^*$  を算出する。

- ④ ②, ③を  $B$  回繰り返して、 $\phi_1^*, \phi_2^*, \dots, \phi_B^*$  を得る。

- ⑤  $\phi$  の標準誤差を  $\sqrt{\frac{1}{B-1} \sum_{i=1}^B (\phi_i^* - \bar{\phi}^*)^2}$  (ただし  $\bar{\phi}^* = \frac{1}{B} \sum_{i=1}^B \phi_i^*$ ) によって推定する。

PPK解析に用いるデータにおいては、一人の被験者から複数の薬物濃度測定値が得られている。この場合、②の復元抽出は被験者単位で行うのが一般的である。すなわち、 $n$  人からなるデータであるならば、そこから、 $n$  人ずつ復元抽出するわけである。

この方法によって、解析的に得ることが難しいパラメータの標準誤差等(例えば相関係数の標準誤差)を求めることができる。また、詳細は省略するが、ブートストラップ結果を用いて、パラメータの信頼区間を推定することも可能である。なお、ブートストラップ回数  $B$  は 200 ~ 1,000 回程度で十分であるといわれている。

さて、このように  $B$  回のブートストラップを行った結果をどのように評価したらよいのであろうか。この評価方法に関しても、現在のところは特に標準的といえる手法は存在しない。しかし、例えば、次のような評価を行えばよいであろう。

- (1)  $B$  回の解析における成功率を評価する。

PPK解析とは統計学的には非線形混合効果モデルによるパラメータ推定である。非線形モデルであることに起因して、モデルがデータに対してあまりに複雑であったり、あるいは、特定の測定値に強く影響を受けているような場合には、計算が収束しなくなることも多い。そこで、モデルの安定性の指標として、 $B$  回の解析のうち計算が正常に終了した割合を用いることが考えられる。

ただし、どの程度の割合で計算が成功したならば「安定」と判断できるかの基準はまだ不明である。一般的に言って、9割以上で成功しているならば「安定」と考えてよからうし、成功率が5割以下ならば「不安定」と見て差し支えないと思われる。しかし、これらの基準値に関しては今後のブートストラップバリデーションの経験の蓄積を待つしかないであろう。

- (2) 推定したパラメータそれぞれについて、平均、標準偏差を求める。

$B$  回のブートストラップのうち、計算が正常に終了した回数を  $B_1$  回とする。この  $B_1$  回での各パラメータ推定値の平均および標準偏差(標準「誤差」ではないことに注意)を算出する。この平均値が元データにおけるパラメータ推定値に「近い」ことを確認する。また、 $B_1$  回の標準偏差は元

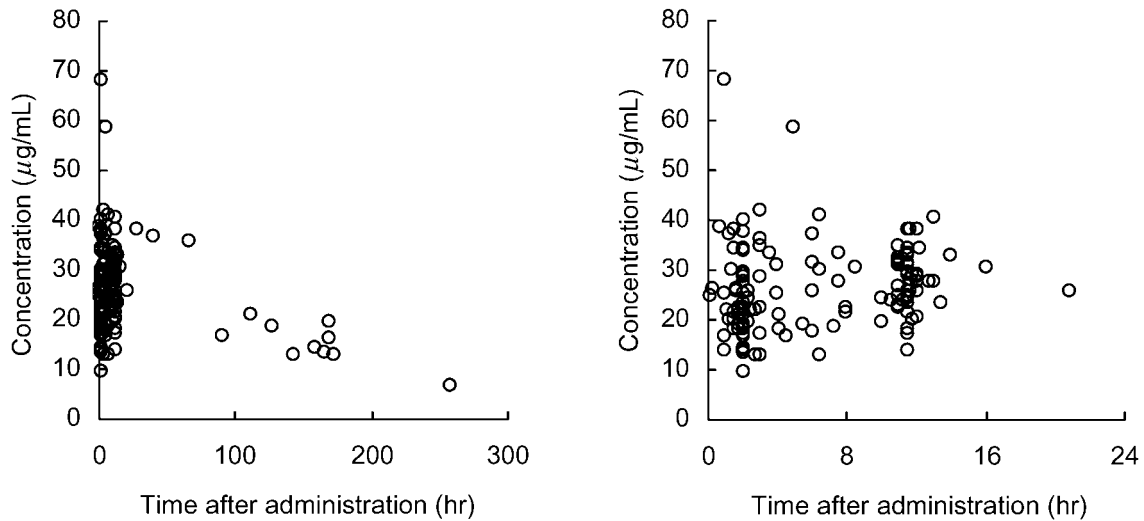


図1 新生児59人におけるフェノバルビタール血清中濃度データ  
右側の図は投与後24時間までの部分を拡大したもの

データでのパラメータ推定値の標準誤差に相当することから、この両者が「近い」値であることを確認する。ただし、どの程度ならば「近い」と判断してよいかについての基準はまだ不明である。この点に関しても、今後のブートストラップバリデーションの経験が蓄積していくことが待たれる。

### 3. フェノバルビタールのデータの解析例

Grasela らによって報告されたフェノバルビタールの血中濃度データ<sup>1)</sup>の解析例を示す。このデータはフェノバルビタールを新生児59人に投与後の血清中濃度である。濃度データは合計で155点であった(図1)。

このデータに吸収相のない線形1-コンパートメントモデルを当てはめて解析し、 $CL$  および  $Vd$  を推定した。 $CL$ 、 $Vd$  とともに体重に正比例するものと仮定した。 $CL$  および  $Vd$  の個体内変動はいずれも相対誤差モデルとした。すなわち、 $i(=1, 2, \dots, 59)$  番目の患者の  $CL_i$ 、 $Vd_i$  を次のようにモデル化した。

$$CL_i = \theta_1 \cdot WT_i \cdot \exp(\eta_{CLi})$$

$$Vd_i = \begin{cases} \theta_2 \cdot WT_i \cdot \exp(\eta_{Vdi}) & \dots\dots (Apgar \geq 5) \\ \theta_2 \cdot (1 + \theta_3) \cdot WT_i \cdot \exp(\eta_{Vdi}) & \dots\dots (Apgar < 5) \end{cases}$$

ここで、 $WT_i$  は患者  $i$  の体重である。また、出生状態が分布容積に影響を与えるかどうかを検討するために、Apgarスコアが5未満(出生状態が悪いことを意味する)の患者において、分布容積が  $1 + \theta_3$  倍になると仮定した。 $\eta_{CLi}$ 、 $\eta_{Vdi}$  はそれぞれ平均0、分散  $\omega_{CL}^2$ 、 $\omega_{Vd}^2$  の正規分布に従う独立な確率変数である。個体内誤差モデルとしても相対誤差モデルを仮定した。すなわち、患者  $i$  の  $j$  番目の血清中濃度

データを  $Y_{ij}$ 、そのときのコンパートメントモデルによる予測値を  $f_{ij}$  として、

$$Y_{ij} = f_{ij}(1 + \varepsilon_{ij})$$

とモデル化した。ここで、 $\varepsilon_{ij}$  は個体内変動であり、平均0、分散  $\sigma^2$  の正規分布に従う独立な確率変数である。すなわち、このモデルの場合、推定すべきパラメータは  $\theta_1$ 、 $\theta_2$ 、 $\theta_3$ 、 $\omega_{CL}^2$ 、 $\omega_{Vd}^2$ 、 $\sigma^2$  の6個である。

解析の結果得られた母集団パラメータを表1に示す。

### 4. ブートストラップ法による解析バリデーションの例

以上の解析に対して、ブートストラップバリデーションを試みた。ブートストラップ回数は200回とした。

200回のデータすべてについて解析は正常に終了し、本モデルが安定であることが示された。

200回のパラメータ推定値の平均値±標準偏差を表2に示す。200回バリデーション結果と元データによる解析結果はよく一致していた。さらに、各パラメータの200回の推定値のヒストグラムを図2に示した。PPK解析結果のパラメータ推定誤差(SE)および信頼区間はこの分布が正規分布であることを前提としている。図2のヒストグラムがほぼ正規分布に近い左右対称の形をしていることが視覚的に確認できよう。したがって、本モデルによって得られたパラメータ推定値が妥当であり、かつ、それに基づく統計学的議論が信頼できるものであることが示された。

ブートストラップ法に限らず、PPK解析のバリデーションを行う際には、「これを検討しておけば十分」という王道は存在しない。PPK解析結果の使用目的に応じて、もっとも適切と思われるバリデーション手法を解析者が選択ある

表 1. NONMEM 解析によって得られたフェノバルビタールの母集団パラメータ

パラメータ	推定値 ± 標準誤差
$CL$ (L/hr/kg)	0.00468 ± 0.000203
$Vd$ (L/kg)	0.965 ± 0.0249
$\theta_3$	0.152 ± 0.0729
個体間変動分散	
$\omega_{CL}^2$	0.0432 ± 0.0205
(CV%)	20.8
$\omega_{Vd}^2$	0.0271 ± 0.00629
(CV%)	16.5
個体内変動分散	
(CV%)	0.0108 ± 0.00191
	10.4

$\theta_3$ : Apgar スコアが 5 未満のときの、分布容積の増加率を表すパラメータ

表 2. ブートストラップバリデーション結果

パラメータ	200回ブートストラップ
	推定値 ± 標準誤差
$CL$ (L/hr/kg)	0.00467 ± 0.000198
$Vd$ (L/kg)	0.964 ± 0.0258
$\theta_3$	0.164 ± 0.0782
個体間変動分散	
$\omega_{CL}^2$	0.0416 ± 0.020
$\omega_{Vd}^2$	0.0268 ± 0.00662
個体内変動分散	0.0108 ± 0.00189

$\theta_3$ : Apgar スコアが 5 未満のときの、分布容積の増加率を表すパラメータ

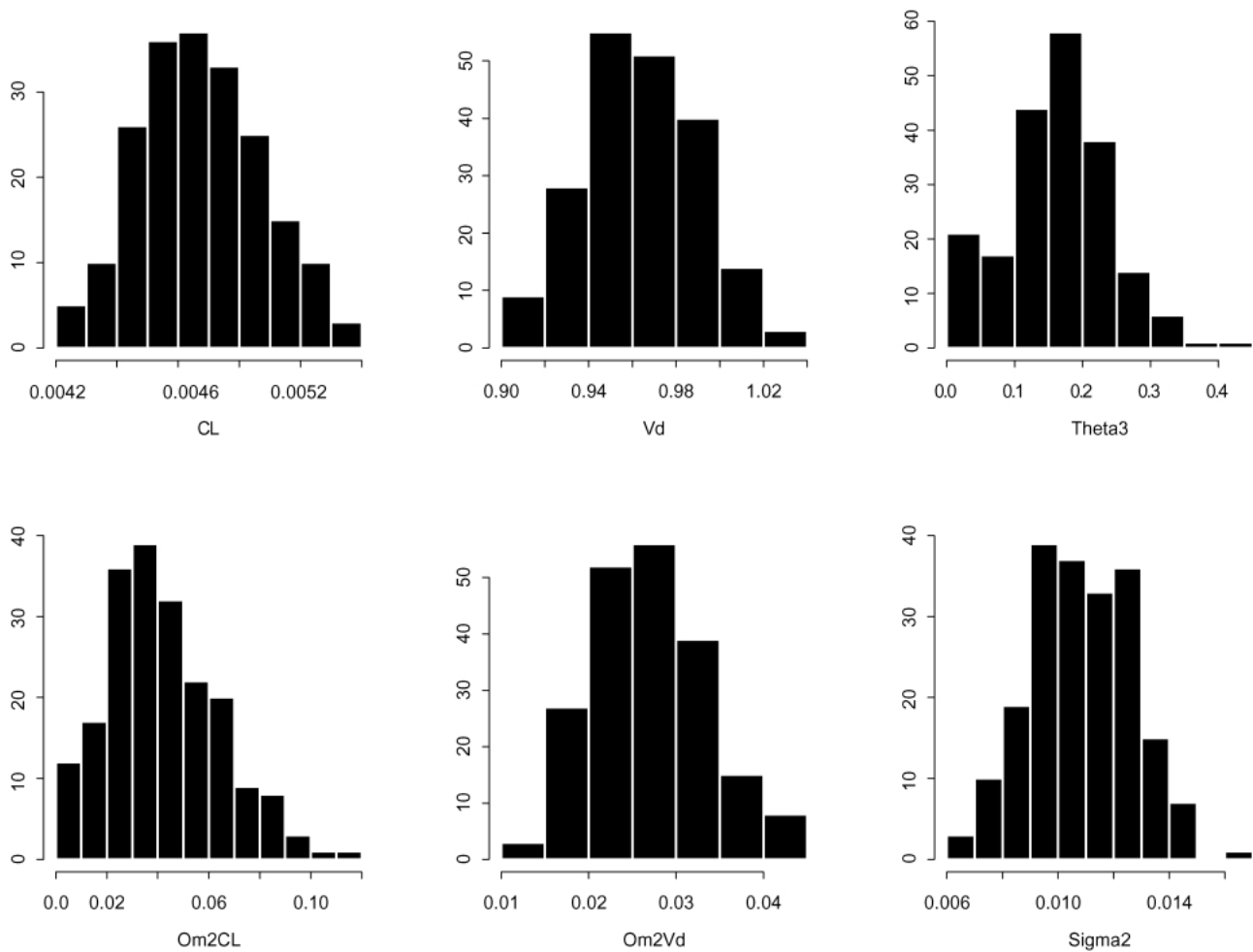


図 2 ブートストラップ結果の各パラメータのヒストグラム

上段左から、 $CL$  (L/hr/kg),  $Vd$  (L/kg),  $\theta_3$  (Apgar スコアが 5 未満のときの、分布容積の増加率を表すパラメータ). 下段は左から、 $\omega_{CL}^2$ ,  $\omega_{Vd}^2$ ,  $\sigma^2$ .

いは考案し、解析結果が妥当であることを必要十分なレベルで示すことが重要であろう。

## 参考文献

- 1) Grasela, T. H., Jr. and Donn, S. M. (1985): Neonatal population pharmacokinetics of phenobarbital derived from routine clinical data. *Dev. Pharmacol. Ther.*, 8: 374-383.

## レクチャーノート：新薬の探索段階における薬物動態試験の実際(1)

### 創薬の探索段階における代謝・動態研究

万有製薬株式会社 つくば研究所 千葉雅人

レクチャーノートの新連載シリーズとして、「新薬の探索段階における薬物動態試験の実際」を取り上げます。新薬の芽を見つけ、開発での成功がより高い化合物を選び、臨床開発候補化合物に育てていく過程で、薬物動態の側面からどのように化合物の動態特性を最適化していくのか？どのような化合物が薬物動態の面から望ましいか？・・・戦略 (Strategy-Paradigm)・試験項目・HTS システムの構築・データの解釈など・・・といった工夫や方法論および苦労話などを、製薬会社の非 (前)臨床薬物動態部門で新薬の探索に携わる創薬研究者の方々から、自由に紹介していただくコーナーにしたいと思います。また、読者の皆様からのご意見やご質問を、随時受け付け、可能である限り、執筆担当者からの回答を掲載したいと思っています。

### はじめに

1985年までの調査から、198例の新規臨床試験において開発が打ち切られた原因は、39%がヒトでの不適切な薬物動態によることが明らかとなっていた (Fig. 1) [1]。一方、ヒト薬物代謝酵素やトランスポーターに関する学術情報の蓄積やヒト実験材料の普及は、薬物探索段階での *In vitro* 実験の有用性を飛躍的に高め、より正確なヒトでの体内動態の予測に貢献してきた。実際、最近 (1992から2001年) の in-house 調査では、不適切な薬物動態による臨床開発の打ち切り (18%) は、安全性試験動物での毒性発現 (36%) について 2 位となっている (Fig. 1)。いずれにせよ、新薬の探索段階における薬物動態研究では、技術革新による探索効率の向上に支えられて、開発候補品のヒトでの体内動態が、その新薬の臨床での投与イメージ (投与量・回数・薬効持続時間など) に近づく方向に、体内動態を最適化していくことが重要である。

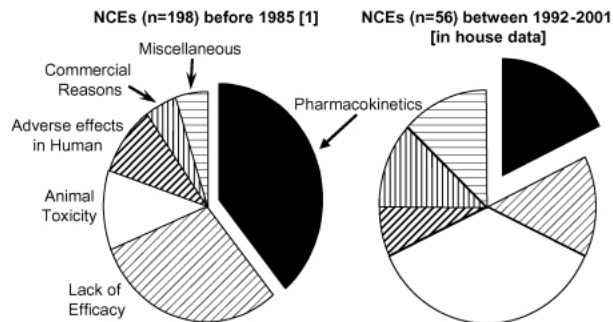


Fig. 1. Why did projects fail in clinical development?

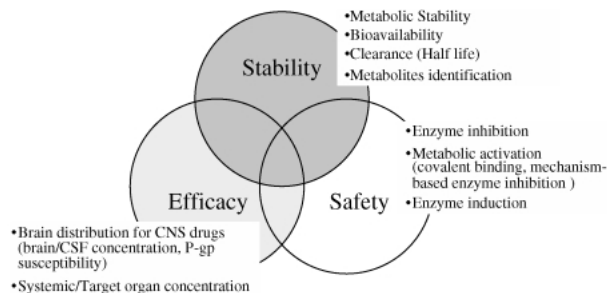


Fig. 2. Evaluations in drug metabolism at discovery stage to optimize candidate for the development.

新薬探索段階での薬物動態試験 (Fig. 2) は、Stability (代謝安定性および経口バイオアベイラビリティの評価と代謝経路の探索的同定)、Efficacy (*in vivo* での薬理効果を裏付けるための脳内・標的臓器への移行性および P-gp 基質性の評価 [CNS 薬の場合])、Safety (代謝的活性化による酵素阻害および非可逆的結合) を柱として、化合物最適化のための情報をフィードバックすることと、開発段階で (動的に) より成功する可能性の高い候補品を見極めることを目的として実施される。

### 1. より安定な化合物を目指して

新薬の探索段階では、ヒトで、より代謝的に安定な体内動態プロファイルを持つであろう開発候補品の創製を目指して、化合物の構造修飾を行う場合が多い。肝ミクロソームを用いた代謝安定性試験は、探索の早い段階で多くの化合物に HTS 方式で実施され、特に、代謝安定性が期待できない化合物のスクリーニングに有用である (Fig. 3)。

さらに、*in vivo* での肝代謝安定性を予測するために、ラット遊離肝細胞あるいはヒト凍結肝細胞を、ラットあるいはヒトの 100% 血清に懸濁して、試験化合物の代謝安定性を直接測定する方法が確立された [2,3]。Figure 4 には、ヒト血清中に懸濁されたヒト凍結肝細胞での代謝クリアランスを、使用された凍結肝細胞でのスケールアップ・ファクター