

## システム適合性試験における精度の推定

岩上 猛,<sup>a</sup> 植田泰輔,<sup>a</sup> 木村良夫,<sup>a</sup> 森本副吉,<sup>a</sup>  
松田りえ子,<sup>b</sup> 林 讓,<sup>b</sup> 今井一洋<sup>c</sup>

## Estimation of Precision in System Suitability Tests

Takeshi IWAGAMI,<sup>a</sup> Taisuke UEDA,<sup>a</sup> Yoshio KIMURA,<sup>a</sup> Noriyoshi MORIMOTO,<sup>a</sup>

Rieko MATSUDA,<sup>b</sup> Yuzuru HAYASHI,<sup>b</sup> and Kazuhiro IMAI<sup>c</sup>

Hayashi Pure Chemicals,<sup>a</sup> National Institute of Health Sciences,<sup>b</sup> Graduate School  
of Pharmaceutical Sciences, The University of Tokyo<sup>c</sup>

(Received April 10, 2002; Accepted July 10, 2002)

In system suitability tests, the daily conditions of an analytical instrument are checked with the standard deviation (SD) of measurements as a criterion. This paper examines how exactly the SD values obtained from the measurements can indicate instrumental conditions. The HPLC measurement for acetaminophen is repeated six times daily to obtain an SD estimate, and these procedures are repeated for seven days. The degree of scattering of the seven SD values corresponds well to the 95% confidence intervals calculated from the chi-square distribution ( $n=6$ ). Therefore it is concluded that the variability of the daily conditions of the HPLC system used in this study is too small to detect by referring to the SD estimates ( $n=6$ ) and the daily change in the SD estimates only represents the error of the SD estimation itself. However, if the repetition number is 40 ( $n=40$ ), the SD estimates will be a good indicator. Forty measurements are practically impossible for a slow analysis like HPLC, and this paper recommends the function of mutual information theory as a substitute for repetitive measurement.

**Key words**—precision; repeatability; system suitability test; HPLC; FUMI theory

## 緒 言

分析値の標準偏差 (SD) や相対標準偏差 (RSD) として表される精度は、分析法バリデーション、GLP, GMP などにおいて重要な概念である。<sup>1,2)</sup> 精度を基準にして、分析システムの信頼性は数値的に評価される。例えば、日本薬局方<sup>3)</sup>には、システム適合性試験におけるシステムの再現性では、「試験を6回くり返すとき、ピーク面積の相対標準偏差は1.0%以下である」などと記されている。

本論文では、実験から求めた測定値のSDの信頼性に焦点を当てる。問題点を明確にするために、まずある実験の結果を示す。Figure 1は、同じサンプルを同じHPLCの条件で続けて6回くり返し測定し、測定値のSDを求めるという操作を7日間行った結果である。SD推定値 (○) は0.088から0.26

の間を日毎に変動している。

Figure 1の解釈として、次の両極端な場合を考えよう：

(場合1) Fig. 1のSD推定値はHPLCシステム

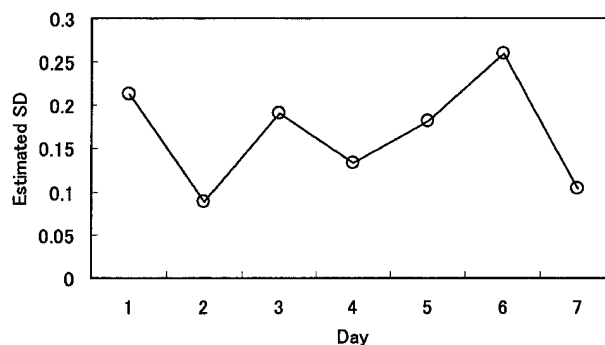


Fig. 1. Daily Change in the SD Estimates (○) of Measurements in HPLC Analysis of Acetaminophen

Samples for repetition are 10 ng/1 acetaminophen. Six measurements ( $n=6$ ) are consecutively obtained a day to estimate the SD and this series of measurements is repeated for seven days to give seven SD estimate (○). The average of the SD estimates is 0.167.

<sup>a)</sup> 林純薬工業株式会社, <sup>b)</sup> 国立医薬品食品衛生研究所,

<sup>c)</sup> 東京大学大学院薬学系研究科

e-mail: fumi@nihs.go.jp

の調子の日間変動を正確に表している；

(場合2) HPLC システムの調子はほぼ一定であり、Fig. 1 の SD 推定値の変動は、推定自体に起因する誤差を示している。

「この分析系の真の SD は変動しているのか又は一定なのか？」という問題である。場合1 が正しいとすると疑問が生じる。測定値の SD は、分析系の調子を表す指標の1つである。Figure 1 では、最も小さい SD 推定値と最も大きい SD 推定値は約3倍異なるから、SD 値の定数倍である検出限界も同様に3倍程度異なる。しかし、最近の分析機器の検出限界は、数日の間でこれほど大きく変動するのだろうか？

一方、場合2を議論するためには、6回のくり返し測定から求めた SD 推定値のバラツキを知る必要がある。幸い、このバラツキはコンピュータ実験から簡単に求められる。Figure 2 には、SD が1の正規乱数を6個発生し、その SD を計算するという操作を100回行い、得られた100個の SD 推定値(○)をプロットしてある。真の SD は1であるが、SD 推定値は0.3—1.9の間にばらついている。調子が全く変化しない分析系の測定値の真の SD は一定であるが、6回のくり返し測定から SD を求めると、その SD 推定値は求めるたびに Fig. 2 のようにばらつくのである。

現実の測定系における SD 推定値のバラツキ (Fig. 1) と調子が一定である理想的分析系の SD 推定値のバラツキ (Fig. 2) を比べれば、場合1と2を判定できる。現実系のバラツキの度合いが理想系

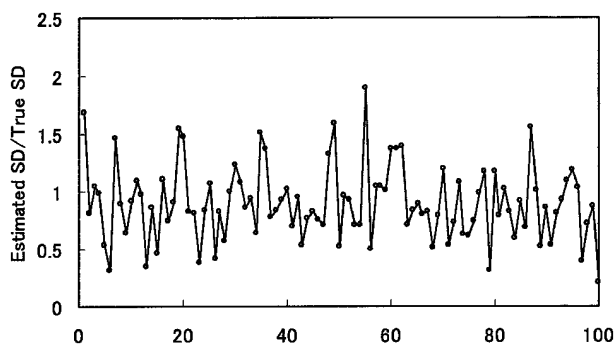


Fig. 2. Variation in the SD Estimates Obtained from Six Measurements ( $n=6$ ) in the Ideal Analytical System in which the SD of Measurements is Constant Irrespective of Time (Computer Simulation)

The variation in the SD estimates corresponds to the estimation error of the SD by repeated measurements.

のバラツキの度合いより有意に大きければ、場合1が正しく、両者のバラツキの度合いが同じくらいならば場合2が正しい。

本論文では、HPLCによるアセトアミノフェンの測定を純度試験のシステム適合性試験のモデルとして取り上げる。また、分析機器の測定値の SD は FUMI 理論 (Function of Mutual Information)<sup>4,5</sup>からも推定できるので、この SD 推定値のバラツキも調べ、場合1と2の判断の補足的な材料とする。調子が一定である分析系における SD 推定値のバラツキの度合いは、Fig. 2 のようにコンピュータ実験でも示されるが、本文中ではカイ2乗分布から95%信頼区間として求める。カイ2乗分布は分散の推定値の分布であり、統計学ではよく知られている。最後に、分析機器の SD 推定値を求めるための方法(くり返し測定と FUMI 理論)について議論する。

## 実 験

**試料** アセトアミノフェンは日本薬局方標準品を使用した。

**HPLC 測定** くり返し測定は、最初にアセトアミノフェン 1 mg/l・メタノール溶液を調製し、次いで移動相で100倍に希釈した溶液 (10 ng/l) をすべての実験に用いた。使用した HPLC システム (東ソー製 LC8020) において、カラムは TSK-GEL ODS 100 s (4.6×250 mm, 5 μm, 東ソー製)、移動相組成は 0.05 mol/l KH<sub>2</sub>PO<sub>4</sub>:CH<sub>3</sub>OH (4:1)、流速は 1.0 ml/min、温度は 40°C、検出は紫外可視吸光検出器の波長 225 nm、サンプル注入量は 20 μl でマイクロシリンジを用いて手動注入した。

**くり返しによるシステム SD の推定** 上記の測定により得られたクロマトグラムをコンピュータに記録し、ソフトウェア MAY2000 (ヤザワ) により解析し、測定値 (全面積値) とその統計量を得た。

**FUMI 理論によるシステム SD の推定** FUMI 理論の説明は文献<sup>4)</sup>に既に記載されているので、ここでは要点だけを述べる。FUMI 理論は、クロマトグラム中のシグナルとノイズから情報を得て、測定値の SD と RSD を推定する理論である。くり返し測定なしに、1回の測定 (1つのクロマトグラム) から測定精度を推定できる。分析機器のベースラインノイズが FUMI 理論のノイズモデルで近似できないと、FUMI 理論を利用できないことが短所の1つ

である。

FUMI 理論に必要な一連の数学的操作（フーリエ変換など）は MAY2000 を用いて行った。フーリエ変換は 512 データポイントの HPLC ベースラインに対して行った。

使用したアセトアミノフェンの濃度 (10 ng/l) では、シグナルとノイズの比は約 9 である。検出限界付近の濃度でシステム適合性試験や日常点検を行うこともあるので、10 ng/l を選定した。測定誤差の主原因がベースラインノイズである場合には、くり返し測定 (Fig. 1), FUMI 理論 (Fig. 3), コンピュータシミュレーション (Figs. 5, 6) を対比できることも、上の濃度を選んだ理由である。なお、高い濃度 (例えば、100 ng/l) で実験を行っても、同様に FUMI 理論で SD 推定値を得ることができ

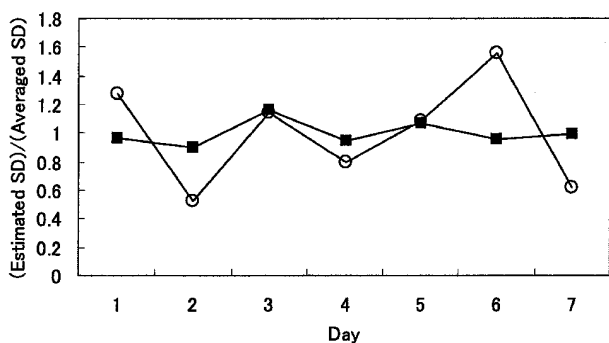


Fig. 3. Daily Changes in the Measurement SD by Different Estimation Methods in HPLC Analysis of Acetaminophen  
 ○ : SD is estimated from repeated measurements ( $n=6$ ), ■ : SD is estimated by the FUMI theory (see the text). The experiments are the same as those in Fig. 1. The SD estimates by the repetition method (○) are the same as those of Fig. 1, but the Y axes are different.

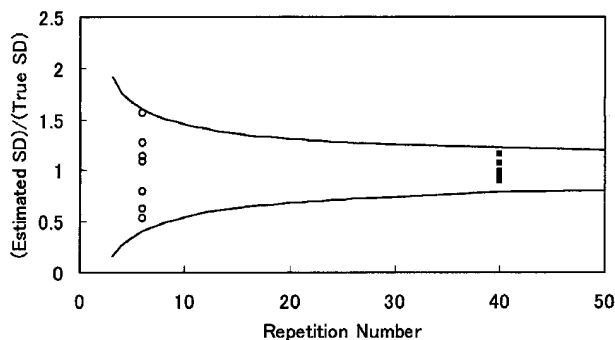


Fig. 4. 95% Confidence Intervals (—) of SD Estimates from the Chi-Square Distribution and Distributions in the Experimental SD Estimates in Fig. 3 (○ and ■)  
 In the experimental SD, the average of the seven SD estimates is substituted for the true SD in the Y-axis. Some SD estimates (■) overlap and are not spotted in the figure.

る。

ノイズの合成 Figure 5にあるノイズの合成は、正規乱数を用いて行った。Figure 5に示してあるホワイトノイズのSDと同じSDの正規乱数を合成ノイズのポイント数だけ発生し、ホワイトノイズを作る。同時に、マルコフ過程のSDと同じSDの正規乱数をホワイトノイズと同じポイント数だけ発生し、自己相関係数を考慮してマルコフ過程を作る。このようにして作ったホワイトノイズとマルコフ過程を足し合わせてモデルノイズとした。

FUMI 理論の SD の信頼区間 FUMI 理論では、1つのノイズから1つのSD推定値が得られるので、100個のノイズを作りこれから100個のSD推定値をまず得た。次に、これら100個のSD推定値のSDを計算した。このSDの1.96倍をSD推定値の

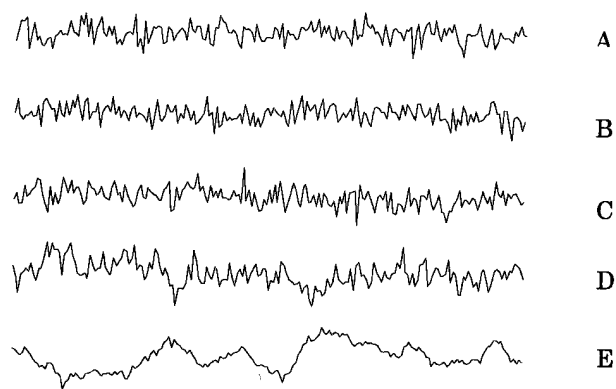


Fig. 5. Noises with Different Noise Parameters (Computer Simulation)  
 The noises are created with the following parameters (the SD of white noise, the SD of the Markov process and auto-correlation coefficient of the Markov process): (A) 1, 0, 0, (B) 1, 0.05, 0.95, (C) 1, 0.25, 0.75, (D) 1, 0.5, 0.95, (E) 0, 0.5, 0.95.

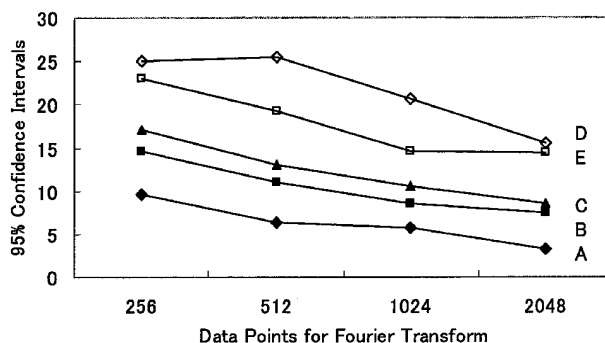


Fig. 6. 95% Confidence Intervals of the SD Estimates in the FUMI Theory (Computer Simulation)  
 The Y-axis is represented on the relative scale:  $1.96 \times [(\text{SD of 100 SD estimates}) / (\text{mean of 100 SD estimates})] \times 100$ . The SD concerns the whole area measurements (150 data points).

95% 信頼区間とし、Fig. 6 にプロットした。なお、ここでの SD 推定値は、ピークの端から端まで 150 ポイントで積分した測定値 (全面積積分) の標準偏差である。

### 結果と考察

**実験値から求めた SD** HPLC でアセトアミノフェンを 1 日 6 回くり返し測定し、SD 推定値を計算する実験を 7 日間行った。その結果 (○) を Fig. 3 に示す。これは Fig. 1 と同じであるが、調子が一定である分析系の SD 推定値のバラツキと比べるため、Fig. 3 では、縦軸を (SD 推定値)/(7 個の SD 推定値の平均) とした。このように標準化した SD 推定値は、0.5—1.6 の間で変動している。

FUMI 理論では、1 回の測定から 1 個の SD 推定値が得られるので、本研究における 7 日の実験からは計 42 個 (=6×7) の SD 推定値が得られる。しかし、Fig. 3 では、見やすくするため、1 日に 1 個の SD 推定値 (■) をプロットした。

SD 推定値の変動幅は、くり返し測定 (○) では大きく、FUMI 理論 (■) では小さいことが分かる (Fig. 3)。同じデータを用いたにも関わらず、このように異なった結果が得られた。この違いの原因を以下で考察する。

**くり返し測定で求めた SD** 調子が一定である分析系 (測定値の真の SD は一定) において、くり返し測定で求めた SD 推定値のバラツキを考察する。Figure 4 には、カイ 2 乗分布から計算した SD 推定値の 95% 信頼区間 (\_\_\_\_) をくり返し数に対してプロットしてある。95% 信頼区間は、同じ条件で SD 推定値を 100 個得たとき、95 個はこの区間内にあり、残りの 5 個はこの区間外にあることを意味する。くり返し数が 6 の時、95% 信頼区間の上限 (\_\_\_\_) は約 1.6 であり、下限 (\_\_\_\_) は約 0.4 であることが分かる。この 95% 信頼区間 (0.4, 1.6) は、真の値 1 を中心に ±60% のバラツキとも表現できる。

くり返し数が大きくなるに従って、SD 推定値の 95% 信頼区間は狭くなる。つまり、くり返し数が大きいほど、SD 推定値の信頼性は高くなる。40 回のくり返しでは、SD 推定値の 95% 信頼区間は約 ±25% である (Fig. 4)。

真の SD が一定である分析系の 95% 信頼区間

(Fig. 4) の上に、アセトアミノフェンの実際の測定結果 (Fig. 3 の○) をプロットした。Figure 4 の横軸 (くり返し数) は 6 であるが、縦軸、(SD 推定値)/(真の SD 値)、の分母は知ることができない。そこで、真の SD 値を 7 個の推定値の平均で代用し、Fig. 3 の 7 個の結果 (○) をプロットした。

実測の SD 推定値の変動 (○) は調子が一定である分析系の 95% 信頼区間 (\_\_\_\_) にちょうど含まれることが確認できる (Fig. 4)。既に議論したように、○は 0.5 から 1.6 の間で変動し、\_\_\_\_は 0.4 と 1.6 にある。

**FUMI 理論から求めた SD** くり返し測定の結果と同様に、FUMI 理論の結果 (Fig. 3 の■) も Fig. 4 にプロットする。FUMI 理論の推定の信頼性は、30—50 回のくり返し測定に相当するので (後述)、FUMI 理論の結果はくり返し数 40 にプロットしてある。7 個の測定結果 (■) もちょうど 95% 信頼区間に含まれることが確認できる (Fig. 4)。Figure 4 には示していないが、FUMI 理論の 42 個の SD 推定値すべてを同様にプロットしても、95% 信頼区間に一致するバラツキを示す。

くり返し測定と FUMI 理論の SD 推定値のプロット (○と■) の幅がカイ 2 乗分布から求めた 95% 信頼区間に一致することから、「緒言」の場合 2 が現実であると結論できる。つまり、測定機器の調子を表す真の測定値の SD は、実験期間中では検出できるほどには変動しないで、ほぼ一定である。

FUMI 理論の推定の信頼性は、30—50 回のくり返し測定に相当するとして議論を進めてきたが、この事実は以下で示すコンピュータ実験から確かめられた。FUMI 理論の SD 推定は非線型最小 2 乗法などを含み数学的に複雑であるので、SD 推定値の 95% 信頼区間を計算する理論はまだ存在しない。しかし、Fig. 2 のようなコンピュータ実験を行い、FUMI 理論による SD 推定値のバラツキから、その 95% 信頼区間を求めた。

分析機器のベースラインノイズは、機器の種類 (HPLC や GC/MS) によって異なることが知られている。FUMI 理論では、ホワイトノイズとマルコフ過程の和として実際の分析機器ノイズを近似する。Figure 5 にはホワイトノイズの SD、マルコフ過程の SD と自己相関係数の計 3 つのパラメータ (ノイズパラメータと呼ぶ) を指定することにより合成し

た典型的なベースラインノイズ (A-E) を示してある。なお、Fig. 5 では、ホワイトノイズだけのノイズ (A) とマルコフ過程だけのノイズ (E) の間に、これらの和となるノイズ (B-D) を設定した。

ここで用いた HPLC のベースラインノイズのノイズパラメータは、ホワイトノイズの SD は 0.00084, マルコフ過程の SD は 0.000485, マルコフ過程の自己相関係数は 0.914 であった。この実測ノイズは、ノイズ D に似ている。なぜならば、どちらのノイズもホワイトノイズの SD がマルコフ過程の SD の約 2 倍であり、自己相関係数が約 0.95 だからである (Fig. 5 の説明参照)。

FUMI 理論では、ベースラインノイズをフーリエ変換して実際の分析機器のノイズパラメータを求め、このノイズパラメータから測定値の精度を計算する。そのため、1つのベースラインから1つの SD 推定値が得られる。そこで、同じ3つのパラメータを持つベースライン (例えば A) を 100 個作り解析すれば、100 個の SD 推定値が得られる。この 100 個の SD 推定値から、SD 推定値の標準偏差が計算できる。荒い近似ではあるが、この SD 推定値の分布をガウス分布と見なして SD 推定値の標準偏差の 1.96 倍をその 95% 信頼区間とする。

Figure 6 は、FUMI 理論の SD 推定値の 95% 信頼区間を表している。ただし、図の縦軸は、真の SD を 100 としたときの相対的な信頼区間を示している。この Fig. 6 から、SD 推定値の信頼性は、ノイズのタイプを決定する 3つのパラメータとフーリエ変換するデータポイント数に依存することが分かる。マルコフ過程 (E) よりも、ホワイトノイズ (A) の方が SD 推定値の 95% 信頼区間が狭い。また、フーリエ変換のポイント数を増やすと SD 推定値のバラツキは小さくなることも分かる。

FUMI 理論による SD 推定値の 95% 信頼区間 (Fig. 6) とカイ 2 乗分布による SD 推定値の 95% 信頼区間 (Fig. 4) を比べれば、FUMI 理論に相当するくり返し数が分かる。Figure 6 より推定精度が良くないノイズ (D の 256 と 512 ポイント) でも、縦軸の値は約 25 である。これは、95% 信頼区間は約  $\pm 25\%$  であり、Fig. 4 の表示では、真の値 1 を中心に約 0.75—1.25 であることを意味する。Figure 4 の 95% 信頼区間が約 0.75—1.25 である時のくり返し数は 30—50 であるので、FUMI 理論の推定値

の信頼性は、30—50 回のくり返し数に相当すると判断できる。そこで、FUMI 理論に相当するくり返し数を 40 としたことは妥当と考えられる。

本研究で用いた HPLC 装置のノイズが Fig. 5 のノイズ D に似ていることから、本研究の分析機器では、FUMI 理論の SD 推定精度の 95% 信頼区間は約  $\pm 25\%$  と推測できる (Fig. 6 参照)。測定機器のベースラインノイズは様々であるが、多くの場合 SD 推定値の 95% 信頼区間は、 $\pm 25\%$  より狭い。つまり信頼性はこれより高いと考えられる。Figure 6 を参照すれば、A のようなホワイトノイズをベースラインに持つ分析機器では、FUMI 理論の解析 (フーリエ変換) に用いるデータ数が 512 ポイントならば、測定値の SD 推定値の 95% 信頼区間は 10% 以下であることが分かる。

## 結 論

アセトアミノフェンの HPLC 測定における測定精度をくり返し測定と FUMI 理論によりそれぞれ求めた。これらの SD 推定値の 95% 信頼区間をカイ 2 乗分布から導かれる 95% 信頼区間と比較した。なお、カイ 2 乗分布から求めた 95% 信頼区間は、調子の変動が全くない (真の SD は一定である) 理想的な分析系に対する SD 推定値のバラツキの度合いを示す。

その結果として

- 6 回のくり返し測定から得られる SD 推定値の日間変動は、用いた HPLC の調子の変動を表しているのではなく、SD 推定自体のバラツキを表している；
- 測定値の SD で表される分析機器の調子の日間変動は、用いた HPLC に関する限り、観測できるほど大きくない；
- FUMI 理論から求めた SD は、測定装置の正しい精度 (真の SD) に近い；

が結論された。今まで、HPLC 以外に、キャピラリー電気泳動、GC/MS、原子吸光分析などに応用され、FUMI 理論の真度は問題ないことが確かめられていることが、<sup>4)</sup> 3 番目の結論の裏付けである。

以上により、機器分析に関しては、正しい SD 推定値を求める方法としては、くり返し測定より、FUMI 理論の方が優れていると結論できる。

謝辞 本研究は、創薬等ヒューマンサイエンス総合研究事業の一環として行われた。

#### REFERENCES

- 1) Kaniwa N., Ojima Z., *Pharm. Tech. Japan*, **16**, 171-179 (2000).
- 2) Kaniwa N., Ojima Z., *Pharm. Tech. Japan*, **16**, 541-565 (2000).
- 3) The 14th Revised Pharmacopoeia.
- 4) Hayashi Y., Matsuda R., *HPLC Bunseki no Seido*, Hayashi Pure Chemicals, 1999.
- 5) Okuda H., Kojima S., Tsunakawa N., ed. by *Iyakuhin no hinshitsukakuho*, LIC, 2002.