

## *Thèse*

*Présentée en vue d'obtenir le diplôme de*

### *DOCTORAT*

*en Sciences Pharmaceutiques*

*Par*

*Mr. Abderrahim Bouabidi*

## **Etude Critique des Différentes Approches de Validation des Méthodes Analytiques**

*Soutenue le Vendredi 26 Avril 2013 à 10 h00 devant le Jury composé de :*

<b>A. TAIRI</b>	Faculté des sciences Ben M'Sik, Université Hassan II Mohammedia Casablanca	Président
<b>A. NICOLAS</b>	Faculté de Pharmacie de Nancy	Rapporteur
<b>Y. CHERRAH</b>	Faculté de Médecine et de Pharmacie Université Mohammed V Souissi Rabat	Rapporteur
<b>M. EL KOUALI</b>	Faculté des sciences Ben M'Sik Université Hassan II Mohammedia Casablanca	Rapporteur
<b>A. BOUKLOUZE</b>	Faculté de Médecine et de Pharmacie Université Mohammed V Souissi Rabat	Examineur
<b>A. ABOURRICHE</b>	Faculté des sciences Ben M'Sik Université Hassan II Mohammedia Casablanca	Examineur
<b>E. ROZET</b>	Faculté de Médecine Université de Liège ULG	Examineur
<b>P. HUBERT</b>	Faculté de Médecine Université de Liège ULG	Directeur
<b>M. TALBI</b>	Faculté des sciences Ben M'Sik Université Hassan II Mohammedia Casablanca	Directeur

*A la mémoire de Yassine*

*A ma Mère et mon Père*

*A Malika, Mohamed Amine et Nada*

## Remerciements

---

*Cette thèse a été réalisée en cotutelle de deux universités :*

- *L'Université de Liège – Faculté de Médecine Département de Pharmacie – Laboratoire de Chimie Analytique*  
*et*
- *l'Université Hassan II Mohammedia – Casablanca - Faculté des sciences Ben M'sik - Unité de Formation et de Recherche Sciences et Procédés d'Analyse (UFR SPA) - Laboratoire de Chimie Analytique et Physicochimie des Matériaux,*

*Ce projet de recherche n'aurait pas pu aboutir sans la riche collaboration et l'aide précieuse et le soutien infaillible de nombreuses personnes. Je désire leur témoigner ma gratitude via ces quelques lignes.*

*Je tiens tout d'abord à remercier Mes Deux Directeurs de thèse :*

*Monsieur le Professeur Philippe Hubert, pour avoir accepté d'encadrer cette thèse et pour m'avoir donné l'opportunité de réaliser un travail de recherche passionnant. Je désire lui exprimer ma profonde gratitude pour son aide constante, ses conseils toujours judicieux et ses préceptes qui se sont constamment révélés être d'une profonde sagesse.*

*Et*

*Monsieur le Professeur Mohammed Talbi sans qui cette aventure n'aurait pu commencer. Je tiens à lui exprimer mes sincères remerciements pour m'avoir accueilli au sein de son laboratoire, son suivi, sa disponibilité, les idées scientifiques et les précieux conseils apportés tout au long de ce travail.*

*Ce travail n'aurait probablement pas pu être entamé et atteindre de telles dimensions sans la contribution de Mr Eric Rozet qui tout au long de ces années, a été d'une très grande disponibilité. Ses conseils, son soutien, son encadrement et sa confiance m'ont permis de mener ce travail à bien. Je lui témoigne de ma profonde reconnaissance.*

*Je souhaite également adresser mes plus vifs remerciements à Madame Houda Bourichi et Mr Miloud El Karbane pour leurs aides précieuses et leurs soutiens.*

*Je souhaiterais remercier mes rapporteurs pour le temps qu'ils ont accordé à la lecture de cette thèse et à l'élaboration de leur rapport :*

*Je remercie Monsieur Alain Nicolas., Professeur Faculté de Pharmacie de Nancy, de m'avoir fait l'honneur d'être le rapporteur de ma thèse, pour sa disponibilité, c'est avec joie que je le remercie l'intérêt qu'il a porté à mes travaux.*

*C'est également avec plaisir que je remercie Monsieur Yahia Cherrah., Professeur à la Faculté de Médecine et de Pharmacie Rabat, pour sa disponibilité. Je lui suis reconnaissant du temps qu'il m'a accordé et de m'avoir fait l'honneur de rapporter ce travail.*

## Remerciements

---

*Un grand merci à Monsieur M'hamed El Kouali, professeur à la Faculté des sciences Ben M'Sik, d'avoir accepté la charge d'être le rapporteur de ma thèse. Veuillez trouver la manifestation de mes remerciements et de mon profond respect. Je me dois également de le remercier pour ces multiples discussions reflétant parfaitement sa disponibilité et son plaisir de partage.*

*Mes sincères remerciements s'adressent aux autres Membres du Jury: Abdelmjid Abourriche Professeur à Faculté des sciences Ben M'Sik, Abdelaziz Bouklouze, Professeur à Faculté de Médecine et de Pharmacie Rabat, Abdelmjid TAIRI Professeur à Faculté des sciences Ben M'Sik, qui me font l'honneur de juger ce travail.*

*Mes sincères remerciements s'adressent aussi à tous les membres du laboratoire de chimie analytique à Liège et du laboratoire de Chimie Analytique et Physicochimie des matériaux à Casablanca, pour leur bonne humeur, leur amitié et les captivantes discussions partagées.*

*Je tiens aussi à remercier Mr Jaouad Cheikh Lahlou PDG des laboratoires Cooper Pharma, Mr Ayman Cheikh Lahlou DG des laboratoires Cooper Pharma pour leurs soutiens moral et matériel apportés à ce travail,*

*J'adresse mes sincères remerciements à Mr Mounir El Guezzar Directeur Développement des laboratoires Cooper Pharma, à tous les membres du laboratoire de contrôle qualité Cooper Pharma et à tous mes anciens collègues du travail pour leur aide, leur soutien moral et leurs conseils. Qu'ils trouvent ici, le respect que j'éprouve pour eux. Je tiens à remercier chaleureusement tous ceux qui ont contribué de près ou de loin à la réalisation scientifique et matérielle de cette thèse.*

*Je remercie aussi Mr Jouad Mufti Zada qui a eu la gentillesse de corriger les fautes d'orthographe et d'expression de ce manuscrit, merci pour sa disponibilité et son soutien.*

*Je tiens également à remercier mes amis, mes proches ainsi que les personnes qui, de près ou de loin, m'ont soutenu par leurs paroles et leurs encouragements.*

*Enfin je voudrais dédier ce travail à tous les membres de ma famille et en particulier à mes parents, ma femme Malika et mes deux enfants Mohamed Amine & Nada : La gentillesse, la patience et le soutien de tous les instants dont ils ont fait preuve à mon égard durant la longue durée de réalisation et la rédaction de cette thèse ont été pour moi un encouragement constant.*

Merci à tous

## CHAPITRE I : Objectifs et structure de la thèse

1. Objectifs
2. Structure de la thèse

## 1. Objectifs

Dans de nombreux domaines, les analyses chimiques sont importantes en raison des décisions prises à partir des résultats fournis, par exemple le contrôle de la qualité des produits, qui est indispensable à effectuer pour les industries chimiques, pharmaceutiques et alimentaires ; pour les problèmes d'environnement (contrôle du taux et de la nature des substances, polluants ou non polluants, que contiennent les fumées, les eaux usées, les déchets industriels ou autres, les résidus d'engrais, d'insecticides ou de défoliants dans les sols, dans les produits agricoles) ; dans le diagnostic médical (taux de glucose, d'urée, de fer, de calcium ou de potassium contenus dans le sang ou l'urine, pH, gaz du sang...) ; pour la connaissance d'un produit de base (minerai, pétrole brut) ou pour l'amélioration des performances d'un produit transformé, qui ne peut se faire que si l'on connaît la composition chimique exacte de ce produit ; dans l'expertise légale (identification des traces de poisons, d'explosifs ou de liquides inflammables, contrôles antidopage, etc.).

Le coût des "résultats erronés" peut être énorme dans tous les domaines. Par exemple, la nécessité de faire des répétitions d'analyses, les procès devant la justice, l'évaluation erronée de la qualité environnementale ou des produits alimentaires, Ils entraînent une perte de confiance dans la fiabilité des résultats analytiques. La fiabilité des résultats analytiques est importante: ils doivent répondre aux besoins des clients (et de toutes les autres parties prenantes et , par là ,acquérir leur confiance. Pour l'industrie pharmaceutique, toutes les décisions critiques sont prises sur la base des résultats obtenus par le laboratoire de contrôle, telles que la libération des lots d'un produit pharmaceutique, l'établissement et la vérification de la durée de validité d'un produit, la bioéquivalence entre deux médicaments, les études de pharmacocinétique ou le diagnostic d'une maladie, etc. La fiabilité de ces décisions ne dépend que de la fiabilité et de la cohérence des résultats obtenus par les méthodes d'analyse.

Pour assurer cette fiabilité, la validation des méthodes analytiques est donc d'une importance primordiale. La validation des méthodes analytiques constitue une étape essentielle dans le processus intégral de démonstration de la fiabilité des résultats analytiques qui seront générés par ces méthodes dans leurs utilisations quotidiennes. En

outre, l'interprétation et l'évaluation des résultats d'analyse au cours de cette étape-clé doivent être fondées sur des bases statistiques. Sinon, l'importance et l'utilité des résultats d'analyse obtenus pour la prise de décision demeurent vagues.

On est confronté, dans la littérature et dans la pratique, à plusieurs approches de validation analytique.

L'objectif principal de cette thèse est de comparer et critiquer ces différentes approches et de statuer sur la fiabilité des décisions prises par chaque approche, au moyen des résultats obtenus à partir des méthodes d'analyse quantitative validées selon ces différentes approches.

Pour atteindre cet objectif, le but des méthodes d'analyse quantitative et de leur validation devrait être rappelé et une étude comparative des différents guidelines existants sera faite. En effet, certaines lacunes dans les schémas de validation existants fournis par les différents documents réglementaires, sont identifiées, ce qui laisse place à un écart entre les critères de validation et l'utilisation prévue de la méthode. En outre, des recommandations concernant l'utilisation de chacun de ces critères ont été présentées en accord avec l'objectif de la méthode analytique et de celui de la validation analytique qui est de prédire la fiabilité future des résultats de routine. L'objectif suivant sera de réaliser une étude comparative entre les différentes approches existantes dans le domaine analytique à savoir les approches classiques et une approche basée sur une méthodologie statistique spécifique: intervalle de tolérance  $\beta$ -expectation. Puis une illustration de l'universalité potentielle de cette méthodologie sera faite à travers l'application de différentes méthodes d'analyses rencontrées dans le domaine pharmaceutique ainsi qu'une étude de capabilité des méthodes d'analyse, validées par cette approche, en les considérant comme des procédés à part entière et en proposant de nouveaux indices de capabilité adaptés à ce genre de procédés. Enfin une étude de certaines pratiques rencontrées dans les laboratoires d'analyse sera faite, à savoir : l'utilisation des placebos en tant que standards de validation au lieu des échantillons réels issus de la fabrication de routine.

## 2. Structure de la thèse

**Le Chapitre II** est consacré à la définition des différentes étapes de cycle de vie des méthodes d'analyse, à savoir ; la sélection, le développement, la validation et l'utilisation de routine de la méthode, après avoir examiné l'étape de validation , avec une revue du concept d'"Erreur totale" comme outil d'évaluation de la fiabilité des méthodes d'analyse ainsi que le rôle de la méthodologie statistique spécifique liée à ce concept: "intervalles de tolérance" et le profil d'exactitude.

Cependant, même si les documents réglementaires publiés définissent les critères de validation exigés pour démontrer que la méthode est adaptée à l'usage pour lequel elle est prévue, il y a une grande marge d'interprétation laissée aux analystes menant à différentes approches pour la validation de la méthode. Par ailleurs, l'adéquation de ces interprétations à l'objectif de la validation de la méthode n'est pas toujours vérifiée.

**Le chapitre III** est, donc consacrée à une étude comparative et critique des critères de validation, à travers les différentes références réglementaires existantes.

L'objectif est de démontrer le grand manque, existant dans ces lignes directrices, d'une méthodologie pratique pour prendre une décision quant à la validité ou non d'une méthode analytique.

En outre, différentes approches pratiques ont été développées par les praticiens et sont présentes dans la littérature et le choix d'une approche ou d'une autre est laissé à l'appréciation de chaque laboratoire.

Cependant, les paramètres statistiques à calculer, avec ces approches, ne sont parfois pas adéquats pour évaluer la performance de la méthode analytique, ex : coefficient de corrélation, pour l'évaluation de linéarité. D'autre part, il y a un manque d'orientation concernant la façon d'interpréter globalement les résultats de la validation bien que certaines de ces approches proposent des limites pour certains critères, au-delà desquels ce critère de performance n'est pas acceptable.



**Le chapitre IV** est une analyse critique des différentes stratégies classiques de validation analytique qui existent dans la littérature et qui sont appliquées dans les laboratoires d'analyse. Ensuite, une comparaison de ces différentes approches sera faite avec la dernière approche de validation basée sur le concept d'erreur totale, ceci en appliquant toutes ces approches à des cas pratiques.

Ainsi, deux exemples pratiques ont été traités avec les différentes approches et différentes conclusions ont été faites concernant la validité de chaque méthode. Par ailleurs, des simulations ont été réalisées pour mettre en évidence le risque producteur et consommateur lié à chaque méthodologie.

Ensuite, une évaluation de l'universalité et de la flexibilité de l'approche de l'erreur totale a été réalisée, **dans le chapitre V**, en l'appliquant à différentes méthodes analytiques et avec différentes matrices rencontrées dans l'industrie pharmaceutique. Des critères d'acceptation ont aussi été proposés pour les différents types des méthodes analytiques rencontrées.

Enfin, les différentes étapes d'une méthode analytique, à savoir : l'échantillonnage, le prélèvement, la préparation de prélèvement, la mesure, l'étalonnage etc., nous font considérer qu'un résultat d'analyse s'obtient à la suite d'un grand nombre d'étapes et nous laissent par là considérer la méthode d'analyse comme un processus similaire aux processus de production.

Afin de tenter de mesurer la performance de ce processus (méthodes d'analyse), une étude de capabilité à été réalisée dans **le chapitre VI** en adaptant l'indice de capabilité Cpk, utilisé couramment pour les procédés de fabrication, au contexte des méthodes analytique, en proposant un nouvel indice spécifique à ces procédés analytiques.

Cet indice développé ainsi que sa proportion de non-conformité ont été appliqués à plusieurs cas couvrant les applications couramment rencontrés dans l'industrie pharmaceutique et qui ont été validées par l'approche de l'erreur totale.

**Le chapitre VII** est une étude comparative entre les résultats de validation réalisée avec des échantillons authentiques et ceux générés par une étude de validation réalisée à partir du placebo. L'objectif est de vérifier la transférabilité des performances quantitatives des

standards de validation à des échantillons authentiques du médicament, puisque, la plupart du temps, seuls des échantillons préparés artificiellement sont utilisés comme des standards de validation dopés.

Enfin, des conclusions générales de cette thèse ainsi que des perspectives pour de futures recherches non seulement dans le domaine de la prédiction de la fiabilité des résultats d'analyse, mais en particulier dans l'étude des différentes pratiques, parfois erronées, dans le domaine de validation analytique rencontrées dans les laboratoires d'analyse sont donnés dans le **chapitre VIII**.

## CHAPITRE II : Introduction

### 1. Génération des résultats quantitatifs : Cycle de vie d'une méthode d'analyse

1.1. Sélection

1.2. Développement et Optimisation

1.3. Validation

1.4. Utilisation en routine

### 2. Concept d'erreur totale et approche du profil d'exactitude

2.1 Concept d'erreur totale

2.2 Rôle des intervalles de tolérance

2.2.1 Intervalle de confiance de la moyenne

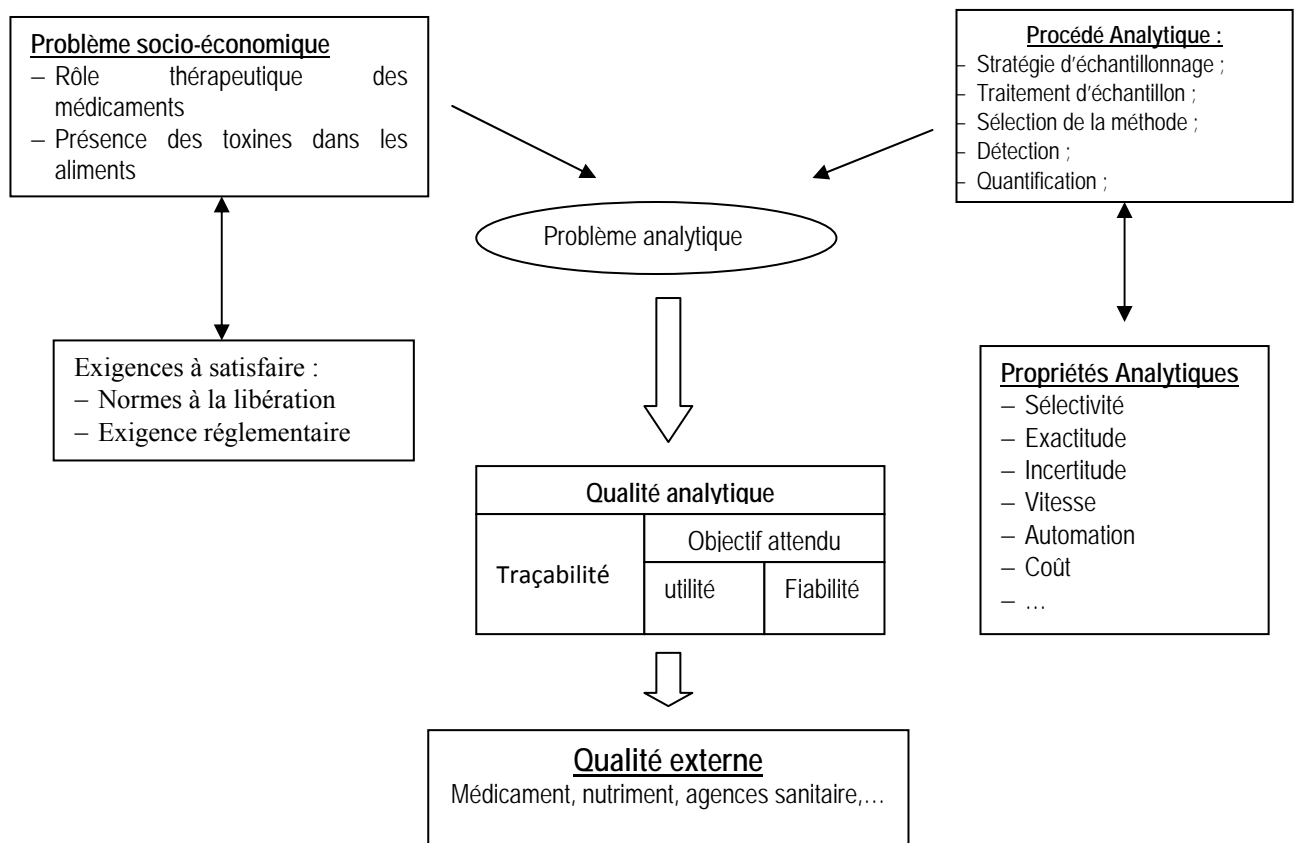
2.2.2 Relation entre intervalle de tolérance et erreur totale

### 3. Bibliographie

**1. Génération des résultats quantitatifs: Cycle de vie d’une méthode d’analyse**

Les méthodes d’analyse sont souvent décrites comme des procédures immuables et figées. C’est un peu l’impression que donnent les manuels et les autres recueils de normes techniques [1]. Or, comme tout procédé de production, les méthodes d’analyse naissent, évoluent et meurent. Pour comprendre clairement le rôle et la place de la validation dans la vie d’une méthode d’analyse, il est intéressant de décrire son cycle de vie depuis le moment où elle est choisie jusqu’au moment où on l’abandonne. [1]

Les domaines d'application des méthodes d'analyse sont très divers, allant du contrôle des médicaments, de la bio-analyse dans le cadre des études cliniques et des études de bioéquivalence jusqu’aux études environnementales et agro-alimentaires... Quelle que soit la méthode d'analyse utilisée et quel que soit le domaine d’application, chaque laboratoire doit être en mesure de produire des résultats fiables lors de l'exécution de l'analyse pour un client ou pour des fins réglementaires, afin de répondre au problème analytique et donc à celui d’ordre socio-économique comme le montre la Figure 1. [2]



**Figure II-1 : Problème analytique (adapté par Valcarcel et al 1997) [2]**

Dans le domaine analytique, deux types de méthodes sont mentionnés : les méthodes qualitatives et les méthodes quantitatives. Dans cette dernière catégorie, la mise en œuvre de ces méthodes (le cycle de vie) peut s'articuler en quatre grandes phases généralement successives telles que l'illustre la Figure 2 [3] :

1. une phase de Sélection où des objectifs et des conditions opératoires initiales sont définis ;
2. une phase de Développement, avec ou sans optimisation au moyen de plans d'expériences;
3. une phase de Validation (Validation Interne/Externe) précédée, selon les cas, d'une phase de prévalidation;
4. une phase d'application en routine (usage en routine), incluant le plus souvent une validation en routine et parfois une validation partielle ou une revalidation.

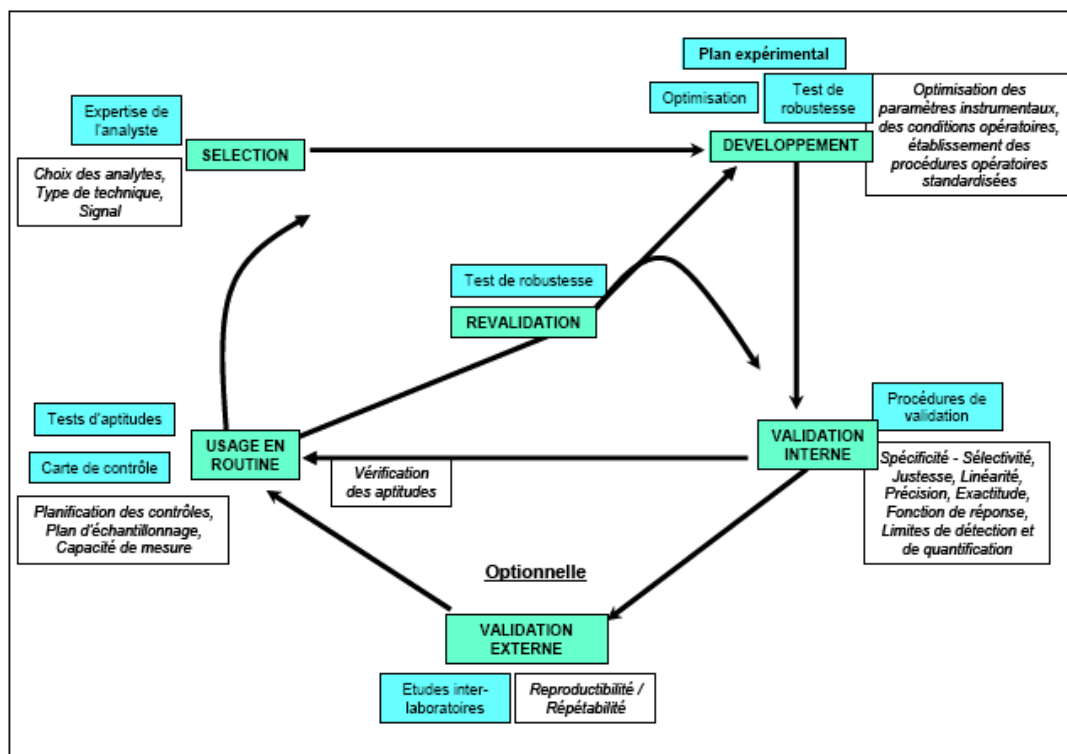


Figure II-2 : Cycle de vie d'une méthode analytique [4]

### 1.1 Sélection d'une méthode

D'abord, il faut sélectionner la technique analytique, c'est-à-dire choisir parmi les diverses méthodes physico-chimiques publiées dans la littérature ou maîtrisées par le laboratoire celle qui doit permettre de déterminer le ou les analytes représentatifs du problème analytique à traiter. Cette démarche repose entièrement sur le savoir-faire et l'expertise de l'analyste [1].

Le choix de la méthode est le point de départ véritable de la validation. Il résulte d'un compromis entre les possibilités instrumentales du laboratoire, le coût d'une mesure et les performances requises. Pour estimer le poids de chacun de ces éléments, on procède de façon assez subjective. Pour être rigoureux, il faudrait pouvoir calculer un rapport coût - bénéfice qui intégrerait les performances métrologiques et le coût de la méthode et, si plusieurs méthodes sont concurrentes, sélectionner celle qui a le rapport coût-bénéfice le plus favorable [1].

Malheureusement, ces données économiques et techniques ne sont pas facilement disponibles. Souvent, elles deviennent accessibles lorsque la méthode est parfaitement connue et que le problème de sa validation ne se pose plus vraiment [1].

Toutefois et afin de sélectionner la méthode adéquate, des questions basiques, comme les suivantes, doivent être posées [2]:

- Quel est l'analyte (s) à chercher?
- Quelle est la nature de l'échantillon : les propriétés physiques et chimiques de l'échantillon?
- Quel est le niveau (s) de concentration de l'analyte (s) à chercher?
- Quel est le niveau d'exactitude nécessaire?
- Quelles sont les limites d'acceptation des performances de la méthode?
- Quels sont les autres constituants de l'échantillon qui sont généralement présents?
- Combien d'échantillons doit-on être analysés?

En effet, il est primordial de connaître le plus d'informations possibles sur l'échantillon qui sera analysé. Les points simples et évidents à résoudre peuvent être les éléments attendus dans l'échantillon (les excipients, les contaminants, les produits secondaires, les produits de dégradation, les métabolites, ...), les composants d'intérêt, les caractéristiques telles que la volatilité de l'échantillon, les solvants possibles, la présence d'acides et / ou de bases,

l'historique de l'échantillon, comme des détails sur la synthèse. Les réponses à ces questions aident à sélectionner la méthode appropriée. Les propriétés chimiques du composé en cours d'analyse sont aussi extrêmement importantes pour la sélection de la méthode. Celles-ci peuvent inclure, mais ne sont pas limitées à, la présence de groupes fonctionnels (chromophores), ionisation, et la polarité [2].

La méthode de détection la plus couramment utilisée dans le domaine pharmaceutique pour la détection des substances médicamenteuses contenant un ou plusieurs chromophores comme groupes fonctionnels est la spectrophotométrie UV/Visible. Les techniques de détection avec une plus grande sélectivité et sensibilité comme les techniques électrochimiques, de fluorescence et la détection par spectrométrie de masse peuvent également être considérées quand la détection directe est jugée inappropriée. Parfois, une modification chimique ou une dérivation peut être utilisée pour l'analyse d'un composé médicamenteux dépourvu de tout chromophore [2].

Si l'analyse quantitative est souhaitée, des critères importants tels que la spécificité/sélectivité de la méthode concernant les interférences possibles, la limite de détection, l'intervalle de dosage, le seuil de quantification, la fidélité, la justesse, l'exactitude, les méthodes d'échantillonnage (gaz, liquide, solide), la préparation des échantillons (extraction en phase solide, la digestion, etc.), la vitesse, l'automatisation, la facilité d'utilisation, l'efficacité, le coût, la résolution temporelle et spatiale, la réglementation (FDA, EPA, BPL, ISO) peuvent être utiles pour orienter la sélection de la méthode [2].

## 1.2 Développement et optimisation

Une fois la méthode sélectionnée, il est souvent nécessaire d'effectuer plusieurs expériences pour l'optimiser. Cette optimisation de la méthode peut être très complexe avec une procédure originale. En effet, en raison du nombre important des paramètres impliqués dans le développement des méthodes d'analyse, il est difficile de trouver l'ensemble des niveaux optimaux des paramètres de la méthode analytique.

Généralement, pour optimiser les conditions opératoires d'une méthode analytique, les plans d'expériences, notamment la méthodologie des surfaces de réponse, sont utilisés [5].

Les plans d'expériences font partie de la chimiométrie [1]. Ils ont pour objectif de fournir à l'analyste, moyennant un nombre d'expériences restreint, un maximum d'informations pertinentes pouvant expliquer certains effets et prédire les réponses étudiées.

Dans la méthodologie des plans d'expériences, tous les facteurs considérés sont étudiés simultanément. La stratégie consiste à réaliser les expériences de manière programmée et raisonnée, en faisant varier les niveaux de tous les facteurs à la fois [6-8]. Le meilleur moyen pour y parvenir, tout en expliquant les phénomènes observés et en exploitant les résultats obtenus, est l'utilisation des outils mathématiques combinés aux outils statistiques ; ce qui rend complexe leur mise en œuvre. Cela n'est possible que par l'utilisation des logiciels développés qui rendent le travail accessible, facile et rapide [4].

Plusieurs types de plans sont définis pour réaliser les expériences et sont décrits dans la littérature [9-12]. Cependant, pour le choix d'un plan adéquat, il y a un certain nombre de prérequis, tels que l'objectif poursuivi par l'analyste, le type et le nombre de facteurs à étudier, les informations à recueillir, et même les exigences de l'expérimentateur [4]. Le Tableau 1 donne une classification des plans d'expériences.



Classe	Type de plan sélectionné
Détection des facteurs influents	Plans de Plackett-Burman
	Plans Factoriels
	Plans de Taguchi...
Optimisation d'un processus (procédures)	Plans Central composite
	Plans de Box-Behnken...
	Plans de Doehlert
Optimisation d'un mélange	Plans de Scheffé
	Extreme vertices designs
Comparer un grand nombre de produit	Blocs incomplets équilibrés...

**Tableau II-1- Classification des plans d'expériences**

L'optimisation des méthodes chromatographiques est le plus souvent réalisée par la méthodologie des surfaces des réponses.

Nous nous contenterons ici d'expliquer brièvement les plans de la surface de réponse les plus utilisées en développement des méthodes chromatographiques : Plans central composite, plans de Doehlert et plan de Box Behnken

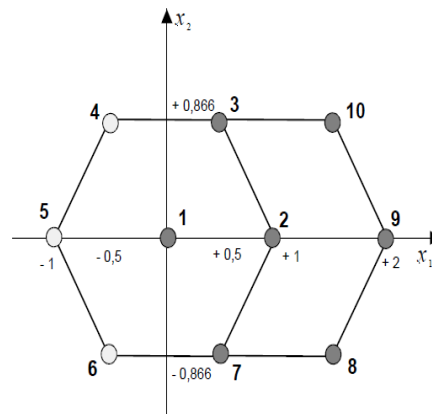
#### **a- Les plans composites**

Un plan composite est constitué de trois parties :

1. Un plan factoriel dont les facteurs prennent deux niveaux.
2. Au moins un point expérimental situé au centre du domaine d'étude.
3. Des points axiaux en étoile. Ces points expérimentaux sont situés sur les axes de chacun des facteurs.

#### **b- Les plans de Doehlert**

La caractéristique principale des plans de Doehlert [11] est d'avoir une répartition uniforme des points expérimentaux dans l'espace expérimental. La Figure 3 donne la disposition de ces points pour un plan à deux facteurs (essais 1 à 7). Tous les points sont à la même distance du centre du domaine d'étude et sont situés sur le cercle trigonométrique. Ils forment un hexagone régulier.

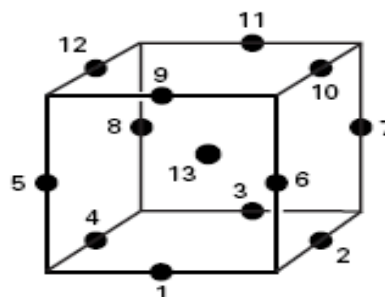


**Figure II-3 : Extension d'un plan de Doehlert à deux facteurs. Trois points suffisent pour retrouver un nouveau plan de Doehlert [13]**

### c- Plans de Box-Behnken

Box et Behnken ont proposé en 1960 ces plans qui permettent d'établir directement des modèles du second degré. Tous les facteurs ont trois niveaux : -1, 0 et +1. Ces plans sont faciles à mettre en œuvre et possèdent la propriété de séquentialité. On peut entreprendre l'étude des  $k$  premiers facteurs en se réservant la possibilité d'en rajouter de nouveaux sans perdre les résultats des essais déjà effectués [13-15].

Le plus connu des plans de Box-Behnken est celui qui permet d'étudier trois facteurs. Les points expérimentaux sont au milieu des arêtes de chacun des côtés du cube (Figure 4). Ce plan comporte douze essais auxquels on peut ajouter un (ou plusieurs) point(s) central (aux) [14-17]



**Figure 4 –Plan de box-Behnken pour trois facteurs**

Généralement, la réalisation d'un plan d'expérience exige une préparation soignée et s'effectue en respectant les étapes suivantes [18] :

- Description de l'étude
- Définition des objectifs de l'étude
- Choix des réponses
- Etablissement de la liste des facteurs influents
- Choix du domaine de variation des facteurs
- Etablir la liste de l'ensemble des contraintes si possible
- Choix et construction du plan d'expériences
- Expérience aux bornes du domaine d'étude
- Eviter les confusions d'effet en recherchant l'orthogonalité entre des facteurs
- Utiliser les logiciels de construction de plans d'expériences quand cela est possible
- Validation statistique du modèle postulé
- Interprétation des résultats
- Conclusion de l'étude

En conclusion, L'application des plans d'expérience pour l'optimisation des méthodes permettra non seulement de trouver les niveaux de facteurs qui répondront le plus adéquatement au problème analytique rencontré, mais Il fournira également un outil pour obtenir l'optimisation robuste de la méthode analytique. Cela correspond adéquatement à la demande de l'CH Q2R1 [19]: "L'évaluation de la robustesse se fait durant la phase de mise au point; elle dépend du type de méthode considérée. Cette évaluation doit démontrer que la méthode demeure fiable lorsqu'on introduit des variations planifiées de paramètres ". En effet, la surface de réponse permettra de trouver une région dans le domaine expérimental où de légères modifications délibérées des paramètres n'affectent pas les réponses étudiées. Ce type d'optimisation de méthode permet d'éviter l'étude de vérification de la robustesse qui est parfois exigée par les agences sanitaires d'enregistrement des médicaments après le processus de validation dans le cadre d'une nouvelle demande d'enregistrement ou lors de dépôt des dossiers d'Autorisation de mise sur le marché (AMM). [2]

### 1.3 Validation

Comme l'indique très explicitement la Figure 5, la validation ne doit intervenir qu'après la mise au point complète de la méthode. L'objectif de la validation est de démontrer que la méthode employée permet d'atteindre des objectifs techniques définis à l'avance. La clause 5.4.5.1 de la norme ISO 17025 : 2005 [20] donne une définition de la validation : « La validation est la confirmation par examen et l'apport de preuves objectives du fait que les exigences particulières en vue d'une utilisation prévue déterminée sont remplies » [1].

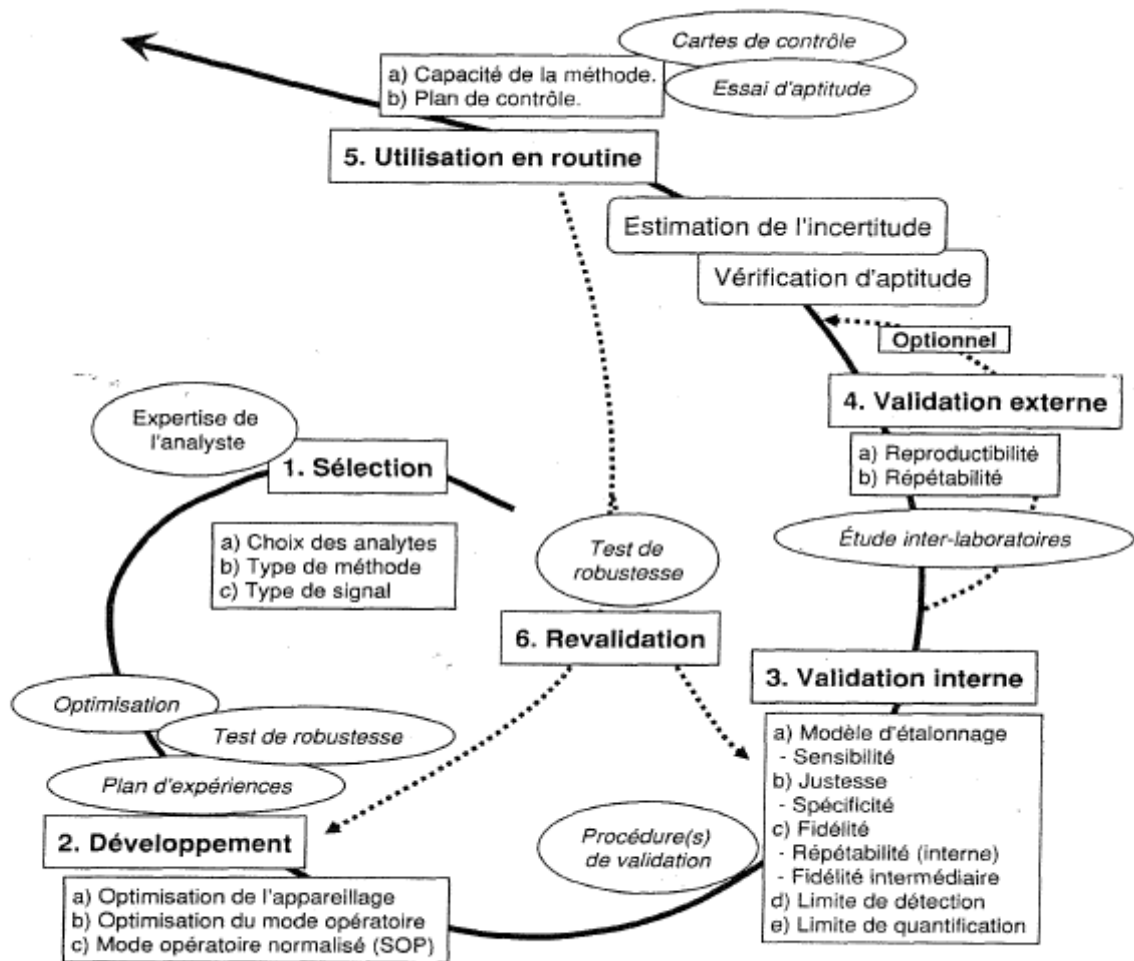


Figure II-5 : Cycle de vie d'une méthode d'analyse [1]

Une distinction entre validation intra-laboratoire (ou in-house) et inter-laboratoires (ou en collaboration) peut être faite. La première est universelle et requise pour toutes les méthodes, la seconde est surtout applicable aux méthodes qui seront utilisées par plusieurs laboratoires ou dont les résultats servent lors d'échanges commerciaux ou de contrôles

officiels. Par exemple, dans l'industrie pharmaceutique, il est inutile (voire impossible) de procéder à la validation interlaboratoire d'une méthode qui ne sert en interne qu'à l'étude d'une molécule non encore mise sur le marché. Par contre, dans les industries agroalimentaires, il faut toujours procéder à une validation inter-laboratoires pour une méthode qui sert à mesurer la conformité d'une denrée lors d'un échange commercial [3]. L'objectif spécifique de la méthode d'analyse doit être défini avant de commencer toute validation. En effet, comme indiqué précédemment, une méthode doit être adaptée à un but donné [1-3].

L'étude de validation devrait inclure les sources de variabilité qui peuvent être couramment rencontrés lors de l'application de la méthode en routine et au moins différents types d'équipement et différents analystes qui vont appliquer la méthode tous les jours. Si l'analyse est effectuée toujours avec le même équipement et par le même opérateur, d'autres équipements ou opérateurs ne devront donc pas être pris en compte. Enfin, avant que l'équipement ne soit utilisé pour l'étape de validation ainsi que pour des analyses de routine, sa performance doit être vérifiée [2].

Le domaine de la validation des méthodes analytiques est défini par plusieurs textes réglementaires ou des guidelines tels que l'ICH [20], la FDA [21], EURACHEM [22], IUPAC [23], AOAC [24]. Ces derniers définissent les critères de validation à tester, mais ne proposent pas des approches expérimentales et se limitent surtout à des notions générales [2]. Nous rappelons qu'au cours de ce travail, nous avons réalisé une étude comparative et critique entre ces différents guidelines afin de relever les différentes défaillances et contradictions présentes dans ces documents (cf. chapitre 2)

Pour ces raisons, plusieurs méthodologies pratiques ont été proposées [25-28] afin d'aider les analystes à valider leurs procédures d'analyse. Ces guides ont grandement contribué à faire progresser la validation des procédures de dosage. Néanmoins, le premier guide de la Société Française des Sciences et Techniques Pharmaceutiques (SFSTP) [25] est exclusivement dédié aux spécialités pharmaceutiques. Il a montré ses faiblesses concernant l'objectif de la validation des méthodes analytiques. Ainsi, par exemple, l'analyste peut être pénalisé, s'il développe une procédure qui est trop fidèle. En outre, l'analyste est confronté

à un grand nombre de tests statistiques qui complique parfois sa décision, au lieu de la rendre aisée. Cette confusion entre les règles de diagnostic et de décision se trouve également dans le deuxième guide SFSTP [27] consacré aux procédures de bioanalyse. Pour cette raison, le récent guide SFSTP [28] visait à aligner les objectifs de la validation en fonction des objectifs de la procédure analytique. Il insiste également sur la validation de la procédure analytique tel qu'elle sera utilisée en routine. Enfin, il fournit une règle de décision simple en utilisant les intervalles de tolérance (cf. section 2, chap. 1). Ce guide a également pour but de proposer un consensus sur les normes généralement reconnues, tout en intégrant largement la terminologie de l'ISO [29]. Il a aussi présenté une stratégie expérimentale pour la validation des procédures de dosage, indépendante du secteur industriel, utilisant au mieux les expériences réalisées afin d'extraire un maximum d'informations à partir des résultats et de minimiser les risques de ré-analyse des échantillons de routine [30]. Cette approche devrait donc réduire considérablement le risque d'accepter une procédure qui ne donnerait pas des résultats suffisamment exacts ou, au contraire, de rejeter une procédure qui peut être appropriée. La notion de profil d'exactitude proposée dans ce guide SFSTP peut être étendue aux autres secteurs d'activité, tels que l'environnement ou l'agroalimentaire.

### 1.3.1 Critères de validation

Les principaux critères de validation sont ceux qui sont largement reconnus et couramment utilisés dans les laboratoires d'analyse [19,21, 22] et s'articulent comme suit :

- Spécificité-sélectivité,
- Fonction de réponse (courbe d'étalonnage),
- Linéarité (des résultats),
- Fidélité (Répétabilité et Fidélité intermédiaire),
- Justesse,
- Exactitude
- Limite de détection (LD),
- Limite de quantification (LQ),
- Intervalle de dosage,

Des définitions plus détaillées de ces critères peuvent être trouvées dans le chapitre 2.

En outre, d'autres critères spécifiques peuvent être requis, selon les domaines concernés, comme par exemple [21]:

- Stabilité de l'analyse,
- Recouvrement d'extraction,
- Effet de dilution,
- etc.

Il est également important de préciser qu'à l'heure actuelle, il n'y a toujours pas de convergence entre les différents documents réglementaires (ISO, ICH, Afnor, Sanco, FDA, Conférence de Washington, etc.) quant à la définition des critères de validation à tester. C'est ainsi que la notion de linéarité apparaît ou non et que son interprétation peut varier d'un document à l'autre. Il en va de même avec la justesse qui, selon les documents, est confondue avec l'exactitude. Pour éviter cela, la nomenclature issue des normes ISO a été choisie, dans le dernier guide SFSTP [28], comme référentiel principal pour la définition des critères de validation.

Il convient de souligner que les critères de validation évoqués ci-dessus doivent être évalués, autant que possible, dans la même matrice que celle des échantillons à analyser [21]. Chaque nouvelle procédure d'analyse devra par ailleurs être validée pour chaque type de matrice (par exemple pour chaque type de fluide biologique envisagé et pour chaque espèce animale).

### **1.3.2 Objectifs de la procédure d'analyse**

L'objectif d'une procédure analytique quantitative est-il de démontrer qu'en moyenne la réponse évolue linéairement en fonction de la concentration introduite, de démontrer que le biais moyen de la procédure est inférieur à x%, de démontrer que la fidélité observée sur un nombre donné de mesures est inférieure à x% ? Bien sûr que non. L'objectif d'une procédure analytique quantitative est de pouvoir quantifier le plus exactement possible chacune des quantités inconnues que le laboratoire aura à déterminer. En d'autres termes,

ce que tout analyste attend d'une procédure analytique, c'est que la différence entre le résultat rendu ( $x$ ) et la « vraie valeur » inconnue de l'échantillon ( $\mu_T$ ), qui par ailleurs restera toujours inconnue, soit petite ou du moins inférieure à une limite d'acceptation, c'est-à-dire [28] :

$$-\lambda < x - \mu_T < \lambda \Leftrightarrow |x - \mu_T| < \lambda \quad \text{Eq. II-1}$$

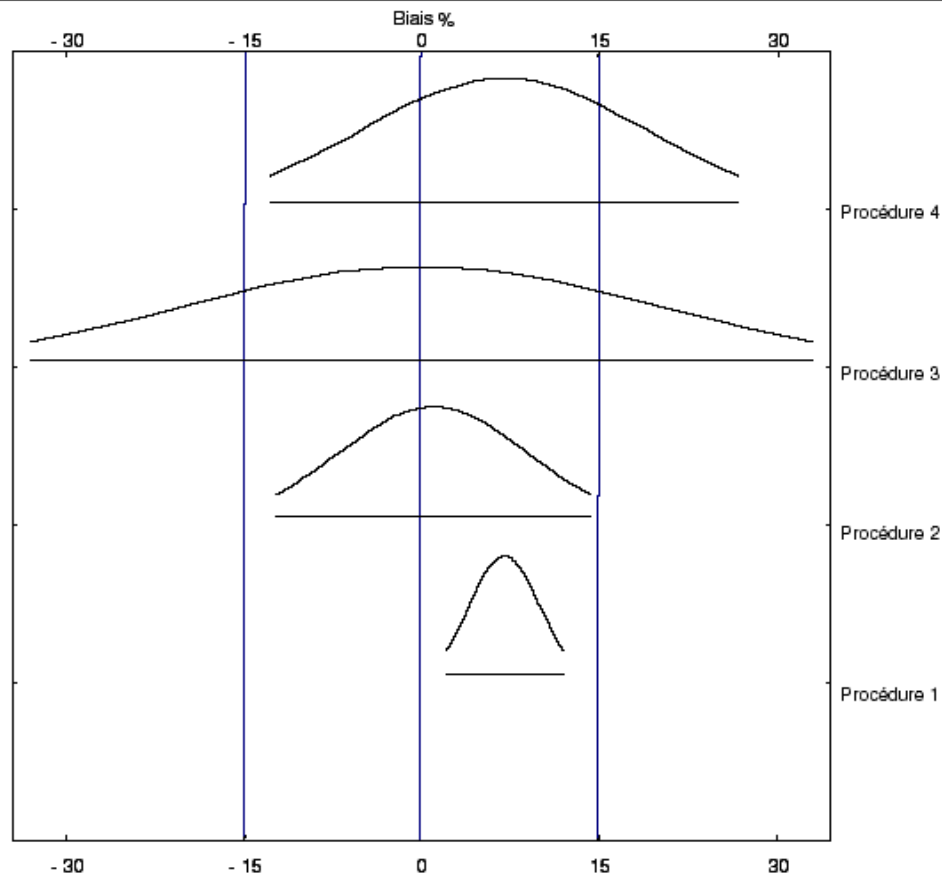
Avec  $\lambda$  la limite d'acceptation, qui peut être variable selon les exigences de l'analyste ou la finalité de la procédure analytique, laquelle est liée aux exigences communément admises par la pratique professionnelle (par exemple : 1 ou 2% sur des matières premières, 5% sur des spécialités pharmaceutiques, 15% en bioanalyse, environnement, etc.). Deux principes fondamentaux apparaissent: d'une part, une notion de limite d'acceptation des performances d'une procédure analytique, d'autre part, et de manière implicite, celle de la responsabilité de l'analyste dans la décision d'accepter ou non une procédure en fonction de ses performances et de l'usage pour lequel elle est prévue.

Par ailleurs, toute procédure analytique (méthode de dosage) se caractérise par un "vrai biais"  $\mu_M$  (erreur systématique), et une "vraie fidélité"  $\sigma_M^2$  (erreur aléatoire mesurée par un écart-type ou une variance). Ces deux paramètres sont propres à toute procédure analytique et sont tout aussi inconnus que la "vraie valeur" de l'échantillon à déterminer. En fait, les expériences réalisées en phase de validation permettent d'obtenir des estimateurs de ce biais et de cette variance (fidélité) qui seront d'autant plus fiables que les expériences effectuées sur des échantillons connus – les standards de validation (SV) – seront adéquates, c'est-à-dire que le plan d'expériences et le nombre d'essais seront appropriés. Ces estimateurs de biais et de variance ne sont pas une fin en soi, ils constituent une étape intermédiaire obligatoire pour pouvoir évaluer si la procédure analytique va vraisemblablement pouvoir ou non quantifier avec une exactitude suffisante chacune des quantités inconnues, c'est-à-dire remplir son objectif.

La Figure 6 illustre graphiquement les concepts précédents ainsi que l'équation (Eq.1). Cette figure présente la répartition de 95% des mesures fournies par quatre procédures analytiques différentes ayant chacune un vrai biais  $\mu_M$  et une vraie fidélité  $\sigma_M^2$  ainsi qu'une



limite d'acceptation  $\lambda$  commune. Dans cette illustration, cette limite vaut 15%, ce qui est un choix classique pour les procédures bio-analytiques [21].



**Figure II-6 : Exemples de procédures ayant la même limite d'acceptation  $\lambda = 15\%$ . Le biais est exprimé en % d'écart à la valeur vraie et la fidélité comme un coefficient de variation.**

Comme l'illustre la Figure 6, la procédure 3 (0% de biais, 20% CV de fidélité) ne remplit pas son objectif puisqu'elle fournit une trop grande proportion de mesures qui s'écarte de plus de 15% de la véritable valeur des échantillons, c'est-à-dire une trop grande proportion de mesures en dehors des limites d'acceptation de  $\lambda = \pm 15\%$ . Cette dernière procédure se caractérise par un biais  $\mu_M$  négligeable et par une fidélité  $\sigma_M^2$  insatisfaisante. La procédure 4 ne remplit pas non plus son objectif. Une proportion trop importante de mesures sort des limites d'acceptation fixées. Notons cependant que la procédure 4 se caractérise par un biais de +7% et une fidélité de 12% qui sont individuellement inférieurs à 15%, , comme l'exige la FDA [21].

En revanche, les procédures 1 et 2 atteignent les objectifs fixés. Nous pouvons donc les retenir et conclure à leur validité en fonction de l'objectif fixé a priori. En effet, elles

permettent de garantir la présence d'au moins 95% de résultats corrects à l'intérieur des limites d'acceptation. Pourtant, en examinant plus attentivement la Figure 6, on notera que la procédure 1 présente un biais de +7% , inférieur à la limite d'acceptation, mais très fidèle (3% CV). La procédure 2, au contraire, se caractérise par un biais négligeable (+1%) mais moins fidèle (8% CV).

Les différences entre ces deux procédures importent peu car nous avons bien dans les deux cas la garantie que les résultats fournis ne seront jamais trop éloignés des vraies valeurs à quantifier. La qualité des résultats est de loin plus importante que les caractéristiques intrinsèques de la procédure en termes de biais ou de fidélité.

Mais vouloir atteindre cet objectif très exigeant a un coût considérable, généralement trop élevé pour être une stratégie acceptable pour un analyste qui ne dispose généralement que de peu de temps. Dès lors, pour réduire ce coût, il va devoir prendre des risques qu'on voudra minimum (ou tout au moins compatibles avec les objectifs analytiques). Pour maîtriser ce risque, on peut inverser le raisonnement et fixer d'emblée une proportion maximale acceptable de mesures qui pourront sortir des limites  $\pm \lambda$ , comme par exemple 5 % des mesures ou 20 % des mesures au maximum en dehors des limites d'acceptation. Cette proportion représente donc le risque que l'analyste est prêt à prendre.

Ainsi, tel que l'illustre la Figure 7 sur la base du "vrai biais"  $\mu_M$  et de la "vraie fidélité"  $\sigma_M^2$ , paramètres qui, rappelons-le, caractérisent toute procédure analytique et qui sont tout aussi inconnus que la "vraie valeur"  $\mu_T$  de l'échantillon à déterminer, les limites d'acceptation des performances de la procédure en fonction de son utilisation permettent de définir un espace des procédures analytiques, pour un domaine d'activité donné, où se trouvent toutes les procédures analytiques acceptables : par exemple  $\pm 15\%$  en bio-analyse suite à la convention de Washington. C'est dans cet espace donc qu'il faudra chercher la "bonne procédure analytique" et ce, en fonction de la proportion de mesures que l'analyste désire avoir dans les limites d'acceptation (par ex : 95%, 80% or 66%). Dans le cas présenté dans la Figure 7, le triangle intérieur représente la zone de toutes les procédures analytiques pour lesquelles on souhaite que 95 fois sur 100 du résultat  $x$  soit compris dans les limites d'acceptation que l'analyste s'est fixé en fonction des contraintes de son domaine d'activité.

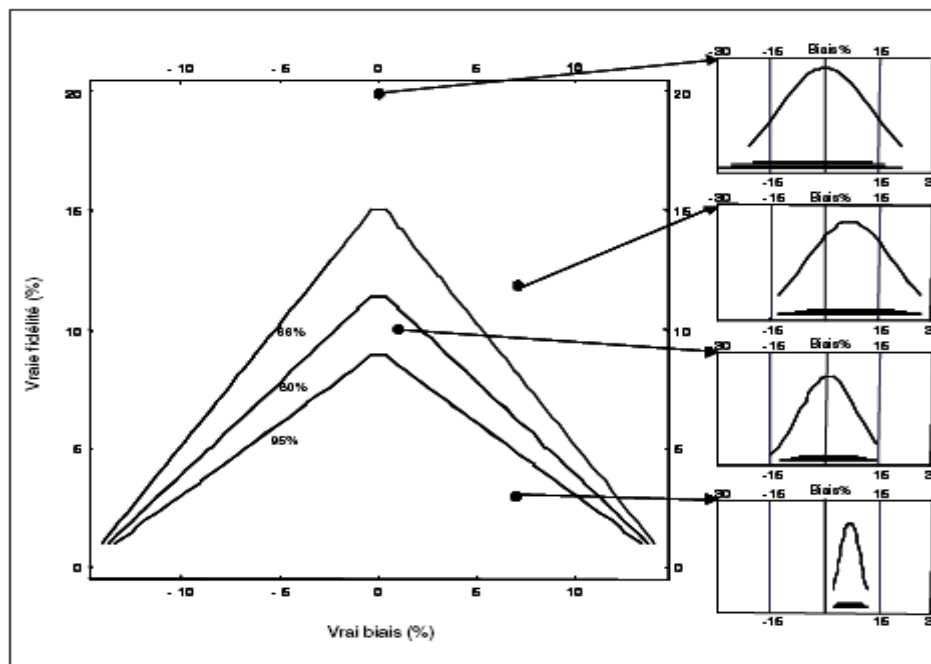


Figure II-7. Limites d'acceptation des performances d'une procédure en fonction de son « vrai biais » (%) et de sa « vraie fidélité » (CV, %) [28].

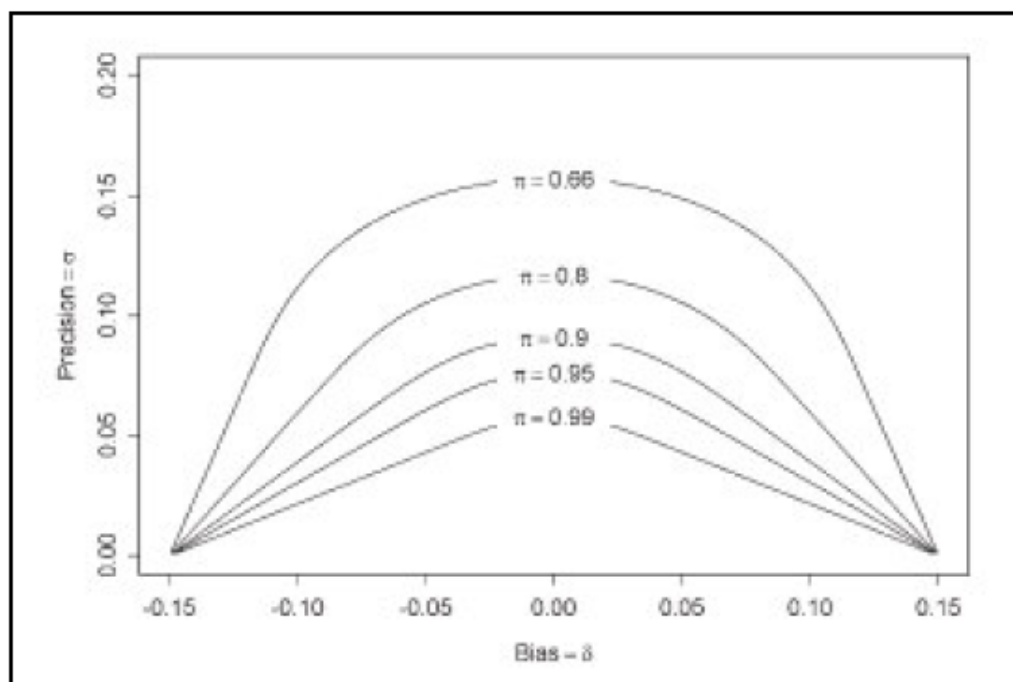


Figure II-8: Régions des limites d'acceptation des procédures analytiques en fonction de leur « vrai biais » ( $\delta$ ) et leur "vrai fidélité " (écart-type,  $\sigma$ ) pour les limites d'acceptation fixées à  $\pm 15\%$ .  $\beta$  représente la proportion de résultats inclus à l'intérieur de la région délimitée par les courbes en forme de cloche [28].

La Figure 8 montre quatre autres zones en forme de cloche correspondant respectivement à des proportions de 95%, 90%, 80% et 66% des résultats compris dans les limites d'acceptation. La proportion de **99%** est donnée à titre indicatif et ne correspond à aucune exigence réglementaire connue. Notons toutefois que pour une procédure se caractérisant par un biais nul et une fidélité de 15%, on n'obtiendra guère plus de 66% des mesures « proches » des vraies valeurs à quantifier.

Toutefois, Boulanger et Al.[31] ont montré que, pour respecter la règle 4-6-15 de la FDA [21] statuant sur la validité d'une séquence d'analyse en routine, une proportion minimale de 80% doit être prise lors de l'étude de validation. La règle 4-6-15 recommandée par la Conférence de Washington (1990) [31] et par le guide de la FDA [21] pour les contrôles de qualité en routine (QC), veut dire qu'au moins 4 échantillons contrôles de qualité sur 6 doivent se trouver dans les limites d'acceptation de 15%. Cette règle de décision revient donc à accepter que seulement 2/3 -ou 66,7%- des mesures soient comprises entre les limites d'acceptation ( $\pm 15\%$  dans le cas présent) [27]. En effet, Boulanger et Al. [31] ont montré qu'en prenant une proportion de 80% à la phase de validation, la probabilité de passer correctement la règle 4-6-15 en routine est d'au moins 90%. C'est en contradiction avec la proposition (intuitive) fréquemment rencontrée dans la littérature [32-33] que les 4 / 6 ou 66,7% des résultats doivent se situer dans les limites d'acceptation de  $\pm 15\%$ . L'adoption de 66,7% comme proportion minimale des résultats qui seront inclus dans les limites d'acceptation lors de la phase de validation peut entraîner jusqu'à 32% de séries acceptables de routine qui vont être rejetées avec la règle 4-6-15 [31].

Dès lors, une procédure peut être qualifiée d'acceptable et donc de valide si grande est la probabilité ("garantie") que la différence entre chaque valeur mesurée ( $x$ ) d'un échantillon et sa "vraie valeur" ( $\mu_T$ ) soit comprise dans les limites d'acceptation que l'analyste s'est fixées. La notion de "bonne procédure analytique" avec un risque connu peut se traduire par la relation suivante:

$$P(|x - \mu_T| < \lambda) \geq \beta \quad \text{Eq. II-2}$$

Avec  $\beta$ , la probabilité d'obtenir des mesures dans les limites d'acceptation, et  $\lambda$  la grandeur définissant les limites d'acceptation fixées a priori en fonction des contraintes du secteur d'activité. Le risque associé d'une procédure s'évalue par la probabilité attendue d'obtenir des mesures en dehors des limites d'acceptation ( $1 - \beta$ ).

### 1.3.3 Objectif de la validation des méthodes

Sachant que les notions de "vrai biais" et de "vraie fidélité" sont des paramètres qui resteront toujours inconnus, mais qu'ils pourront être estimés par les mesures qui seront réalisées en phase de validation, quel est donc l'objectif de la validation ?

Il paraît raisonnable de prétendre que l'objectif de la validation est de donner aux laboratoires ainsi qu'aux autorités compétentes les garanties que chaque mesure qui sera réalisée ultérieurement en routine, une fois la procédure validée, sera suffisamment proche de la "vraie valeur" inconnue de l'échantillon ou du moins comprise dans une limite acceptable, en fonction de la finalité de la procédure. Le but est de minimiser le risque tant au niveau du producteur que du futur consommateur. Par conséquent, et contrairement à ce que la pratique courante peut laisser penser, l'objectif de la validation n'est pas simplement, d'obtenir des estimateurs du biais et de la fidélité. Il n'en demeure toutefois pas moins que ces paramètres, sont, comme nous l'avons déjà signalé, des éléments indispensables à l'établissement des garanties précitées.

A la lecture de cet objectif, deux concepts fondamentaux se dégagent, à savoir :

"suffisamment proche", signifiant, par exemple, que la mesure réalisée en routine sera à moins de  $\lambda\%$  de sa "vraie valeur" inconnue (cf. éq. 1)

"garanties", signifiant qu'il est très probable que quelle que soit la mesure, elle sera suffisamment proche de la vraie valeur inconnue (cf. éq. 2).

Dans ces conditions, l'exactitude, la fidélité, la linéarité, ... ne sont plus que des "paramètres statistiques" ou "éléments de calcul" permettant de chiffrer ces garanties. En effet, on attend d'une procédure de dosage qu'elle soit capable de quantifier, et non pas tellement qu'elle soit fidèle, même si, admettons-le, cette fidélité constitue un gage de réussite. C'est dans ce contexte que l'analyse statistique des résultats de validation trouve

sa véritable dimension. Toutefois, dans cette perspective, il faut encore faire la différence entre les statistiques qui permettent de prendre une **décision** (quant à l'aptitude d'une procédure à correspondre à l'usage pour lequel elle est prévue) de celles qui aident à poser un **diagnostic**, c'est-à-dire nous renseigner sur un point particulier de ses performances, comme par exemple la linéarité ou le passage de l'ordonnée par 0.

En fait, ce dont on a réellement besoin, c'est d'un bon outil de décision c'est-à-dire un outil offrant la **garantie** que la plupart des futures mesures seront dans les limites d'acceptation. Si les garanties offertes par la règle de décision ne sont pas satisfaisantes, mais dans ce cas seulement, les outils de diagnostic aideront alors l'analyste à identifier les causes possibles du problème.

### 1.3.4 Les règles de décision

Puisque les documents réglementaires relatifs à la validation des méthodes sont à caractère global, ils laissent une place à l'interprétation des analystes qui pourront alors choisir la règle de décision qui permettra de déclarer valide une méthode d'analyse. Selon le processus de décision adopté, nous pouvons distinguer entre 3 approches statistiques conventionnelles, se caractérisant toutes par le fait qu'ils comparent séparément la justesse et la fidélité, c'est-à-dire respectivement les erreurs systématiques et les erreurs aléatoires de la méthode. Nous distinguons aussi une quatrième approche qui s'intéresse à la combinaison des erreurs systématique et aléatoire dans le concept d'erreur totale.

1. L'approche descriptive [34],
2. L'approche de différence [26, 34, 35],
3. L'approche d'équivalence [26; 34-36].
4. L'approche erreur totale ou profil d'exactitude

#### a. L'approche descriptive :

L'approche descriptive utilise seulement des estimations des paramètres statistiques : Biais et fidélité intermédiaire [34].

Les valeurs estimées de chaque critère sont calculées à chaque niveau de concentration des standards de validation et sont comparées aux limites d'acceptation fixées a priori. Les limites d'acceptation rencontrées dans l'industrie pharmaceutique pour des méthodes de dosages de principes actifs dans des produits finis sont:

- $\pm 2$  % pour le biais relatif
- 3% pour le  $CV_{FI}$

Cela signifie que si le biais estimé est inclus dans l'intervalle de [-2 %, +2 %], **la justesse** de la méthode pour le niveau de concentration étudié est acceptée.

**Pour le critère de fidélité**, si le  $CV_{FI}$  est plus petit que 3 %, la fidélité de la méthode pour le niveau de concentration étudié est acceptée.

### b. L'approche de différence

L'approche de différence est basée sur les tests d'hypothèses comme le test de Student [26, 34, 35].

Les tests des hypothèses statistiques sont composés de deux hypothèses mutuellement exclusives, à savoir l'hypothèse nulle  $H_0$  et l'hypothèse alternative  $H_1$ .

Le critère de justesse (ou le biais de la méthode) est évalué en utilisant un test bilatéral de Student dont les hypothèses nulles et alternatives sont données à l'Eq. 3. Ceci peut être vérifié en comparant l'intervalle de confiance à 95 % du biais global estimé à la valeur 0% de biais relatif.

#### Hypothèse nulle :

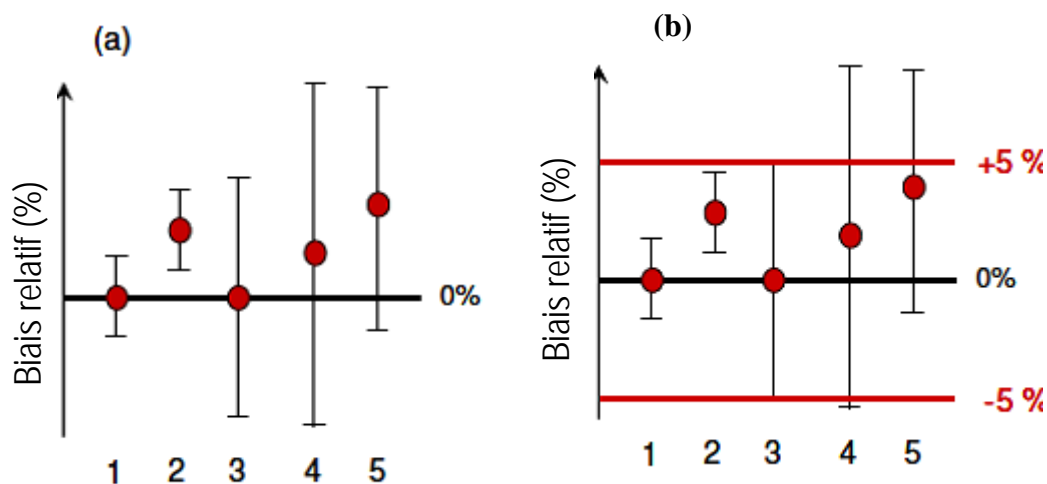
$H_0$ : biais = 0  $\leftrightarrow$   $H_0$ : biais relatif = 0 %  $\leftrightarrow$   $H_0$ : recouvrement = 100 %

#### Hypothèse alternative :

$H_0$ : biais  $\neq$  0  $\leftrightarrow$   $H_0$ : biais relatif  $\neq$  0 %  $\leftrightarrow$   $H_0$ : recouvrement  $\neq$  100 % **Eq. II-3**

Avec  $\text{biais} = x_i - \mu_T$ , le biais relatif =  $(x_i - \mu_T) / \mu_T * 100$  et le recouvrement =  $(x_i / \mu_T) * 100$

Si cet intervalle contient 0 % de biais, la justesse de méthode est acceptée, sinon, elle devrait être rejetée comme illustré dans la Figure 9a.



**Figure II-9 : Les règles de décisions de validité des méthodes analytiques pour 5 situations différentes (1 à 5) selon le biais relatif de la méthode ; (a) l'approche de différence du biais relatif, (b) l'approche d'équivalence du biais relatif avec les limites de décision fixées à  $\pm$  5%**



### c. L'approche d'équivalence

L'approche d'équivalence diffère des précédentes par le fait qu'au lieu de vérifier si le biais relatif estimé ou le  $CV_{FI}$  estimé est inclus dans les limites d'acceptation, elle vérifie si la vraie valeur de ces paramètres est incluse dans leurs limites d'acceptation respectives, cela pour chaque niveau de concentration des standards de validation étudiées comme le montre la Figure 9 (b)

Cela consiste à comparer des intervalles de confiance des paramètres étudiés aux limites d'acceptation [33-36].

Pour le critère de justesse, l'intervalle de confiance à 95 % du biais relatif est calculé pour chaque niveau de concentration et comparé par exemple aux limites d'acceptation de  $\pm 2$  %.

L'hypothèse statistique de ce test est maintenant :

$$H_{01} : \text{Biais relatif} \leq -\Delta \text{ vs } H_{11} : \text{biais relatif} > -\Delta$$

et

$$H_{02} : \text{Biais relatif} \geq \Delta \text{ vs } H_{12} : \text{biais relatif} < \Delta \quad \text{Eq. II-4}$$

où  $\Delta(\%)$  est la limite de décision, à savoir la différence maximale tolérée pour le biais relatif de la méthode (par exemple 2%).

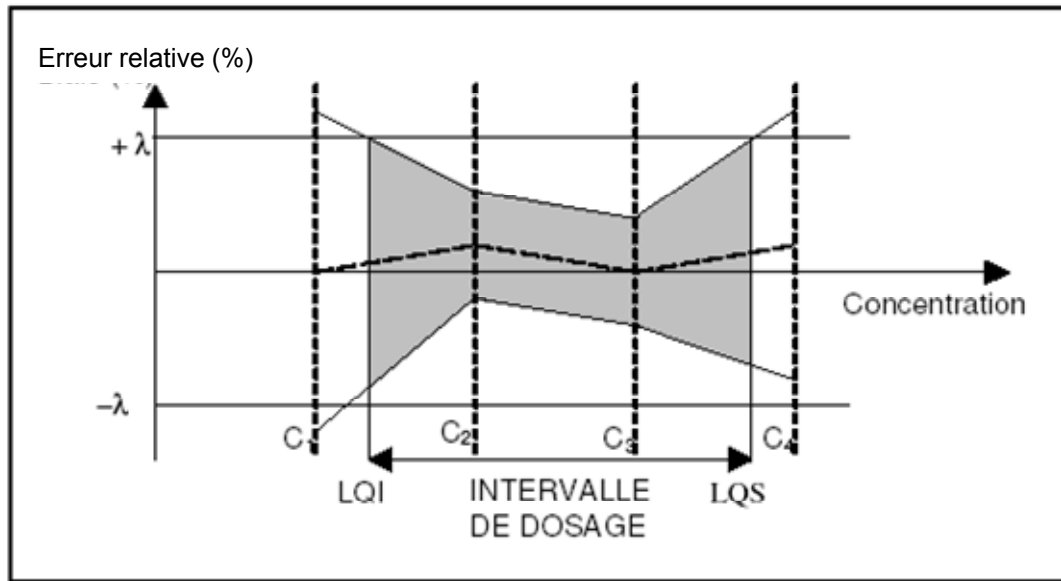
Comme illustré par les deux lignes horizontales représentant les limites de décision de  $\pm 5\%$  au niveau de la Figure 9b, la règle de décision a changé. Avec cette nouvelle règle, les procédures 4 et 5 sont rejetées, alors que la procédure 2 est maintenant considérée comme valide.

**Pour le critère de Fidélité**, c'est la limite supérieure de l'intervalle de confiance à 95% du  $CV_{FI}$  qui est comparée à, par exemple, la limite de 3% et qui doit être plus petite que cette valeur maximale.

**d. L'approche de l'erreur totale ou Profil d'exactitude**

A côté de ces approches classiques, une approche originale, basée sur le profil d'exactitude en utilisant les intervalles de tolérance statistique et l'erreur totale de mesure comme outil de décision unique a été proposé [28, 30,37]. Face à la disponibilité de ces différentes approches, il est du devoir de l'analyste de choisir l'approche la plus appropriée.

Toutefois, cette dernière approche est une règle de décision, à la fois pratique et visuelle. Elle repose sur l'intégration du profil d'exactitude dans des limites d'acceptation ( $\pm\lambda$ ) [28, 30]. Le profil d'exactitude, construit à partir des intervalles de tolérance (voir la section 2.2 de ce chapitre « Rôle des intervalles de Tolérance ») permet donc, comme illustré dans la Figure 10, de décider de la capacité ou non d'une procédure à fournir des résultats dans les limites d'acceptation. La zone en grisé montre l'intervalle de dosage dans lequel la procédure est capable de quantifier avec une exactitude connue et un risque fixé par l'analyste. C'est ainsi que dans ces conditions, si l'analyste est prêt à assumer par exemple un risque de 5%, il pourra au terme de la validation de sa procédure garantir que 95 fois sur 100 les futures mesures fournies par celle-ci seront comprises dans les limites d'acceptation fixées en fonction des contraintes de son secteur d'activité (ex: 1% ou 2% sur les matières premières, 5% sur les spécialités pharmaceutiques, 15% en bio-analyse, environnement, etc.).



**Figure II-10: Illustration du profil d'exactitude comme outil de décision.**  
**LQI: limite de quantification inférieure; LQS: limite de quantification supérieure**

Comme les vraies valeurs de biais et de fidélité de la procédure sont inconnues, le profil d'exactitude par niveau de concentration ( $C_1$ ,  $C_2$ , ...) s'obtient en calculant l'intervalle de tolérance qui permet d'évaluer la probabilité d'obtenir des résultats futurs dans des limites d'acceptation et ce, à partir des estimations disponibles du biais et de la fidélité de la procédure après la phase de validation au niveau de concentration en question. A noter que les estimations du biais et de la variance sont des éléments essentiels pour calculer les intervalles de tolérance, mais la décision n'est pas faite sur la base de ces estimations de biais et de variance [2].

Le profil d'exactitude, représenté dans la Figure 10, s'obtient en reliant les limites de tolérance basses et hautes estimées pour chaque niveau de concentration. Si, sur une partie de l'intervalle de dosage sous épreuve, les limites de tolérance devaient sortir des limites d'acceptation, comme dans l'exemple de la Figure 10 pour les niveaux de concentration  $C_1$  et  $C_4$ , de nouvelles limites de quantification, et par là-même un nouvel intervalle de dosage seraient définis.

La Figure 10 représente ces nouvelles limites LQS (limite de quantification supérieure) et LQI (limite de quantification inférieure) qui sont en parfait accord avec la définition de ce critère,

à savoir la plus petite quantité de substance à analyser qui peut être dosée avec une exactitude (justesse + fidélité) définie.

L'utilisation du profil d'exactitude, comme seul outil de décision, permet non seulement de réconcilier les objectifs de la procédure avec ceux de la validation mais aussi de visualiser rapidement la capacité de la procédure à répondre de façon fiable à son objectif analytique [28, 30,37].

Une étude comparative et critique a été réalisée dans le cadre de ce travail en appliquant ces 4 différentes approches sur deux exemples de dosage des principes actifs dans le produit fini (Voir Chapitre IV).

#### 1.4 L'utilisation en routine

Si la validation s'avère conforme, la vie de la méthode se poursuit par son utilisation en routine. L'obligation de maîtriser qualité implique un contrôle permanent des performances dans le temps. La norme ISO 17025 (2000) , par exemple, [38] exige que les méthodes d'analyse soient contrôlées lors de l'utilisation de routine afin de garantir la fiabilité des résultats analytiques qui sont produits pendant les séries de tous les jours et donc des décisions subséquentes. De même, la FDA [21] précise que, une fois la méthode d'analyse validée, son exactitude doit être contrôlée régulièrement afin de s'assurer que la méthode continue à fonctionner de façon satisfaisante. Pour effectuer ce contrôle, on s'appuie, bien sûr, sur les connaissances acquises lors du développement et de la validation de la méthode mais aussi sur des outils spécifiques qui sont les cartes de contrôle et les essais d'aptitude [1,2].

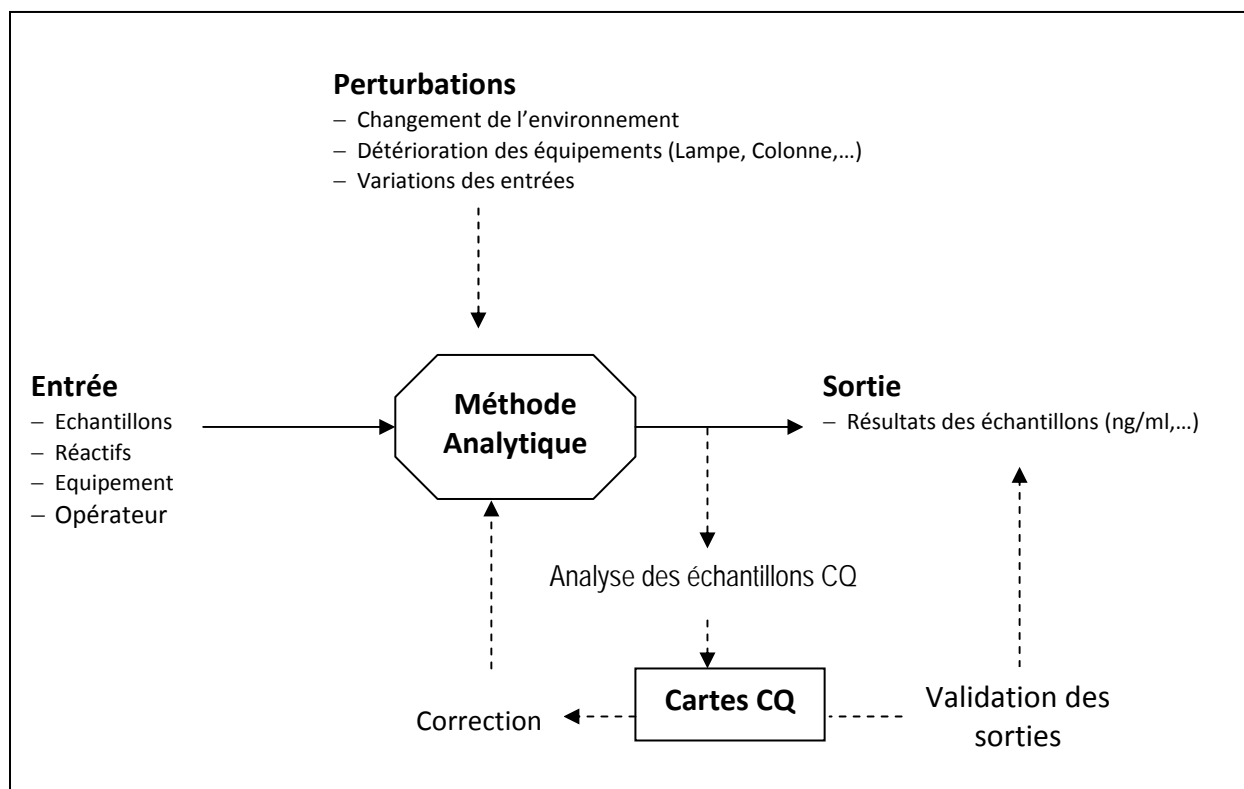
La méthode classique pour vérifier la fiabilité d'un système est de répéter l'analyse d'un échantillon de référence à intervalles réguliers et de le surveiller avec les techniques de maîtrise statistique des procédés (MSP) [2, 39-42].

EURACHEM (1998) [22] encourage l'utilisation des cartes de contrôle de la qualité (CQ) pour assurer la fiabilité quotidienne des méthodes et résultats obtenus. En outre, ils proposent d'utiliser au moins 5% d'échantillons de CQ de routine dans les essais de routine (un sur vingt échantillons analysés devrait être un échantillon CQ). Si un écart dans les résultats se trouve, le système doit être ajusté, dans la mesure du possible. En général, la variation dans les résultats de tout système d'analyse peut être interprétée comme la somme de deux composantes: une composante aléatoire qui est due à la variabilité inévitable de la méthode analytique, et une composante systématique ou non aléatoire qui pourrait être fonction du temps et qui peut être causée par l'utilisation des instruments, par des changements des conditions environnementales, etc.

La plupart des systèmes d'analyse réelle contiennent des pièces mécaniques, électriques et / ou optiques, dans lesquelles le facteur temps peut avoir une importance déterminante : le vieillissement de la lampe du détecteur photométrique, une diminution de la capacité de

séparation de la colonne chromatographique, ... D'autres systèmes d'analyse peuvent être affectés par des changements dans les conditions environnementales, comme la température, l'humidité, .... Ces changements peuvent provoquer des dérives dans les résultats, c'est-à-dire que des tendances systématiques des résultats sont fonction du temps, comme le montre la Figure 11. Le but de la MSP est de séparer les variabilités communes (causes aléatoires) des causes spéciales (systématique). Les cartes de contrôle sont les principaux outils utilisés pour atteindre cet objectif.

Le but de cartes de contrôle est de tester les deux hypothèses suivantes ; celle qui stipule que ces variations proviennent des causes communes contre l'hypothèse alternative qui suppose que ces variations sont d'origine spéciale. Dans le premier cas, le système est dit travailler sous contrôle statistique ou être dans un état contrôlé, et dans le second, le système est dit hors contrôle, devant donc être ajusté.



**Figure II-11- Les procès d'évaluation pour contrôler les méthodes d'analyse durant l'utilisation en routine [39]**

Le Guide de la validation des méthodes de bioanalyse de la FDA (2001) [21] propose une règle spécifique pour surveiller les méthodes d'analyse lors de l'utilisation de routine. Pour

atteindre cet objectif, un certain nombre d'échantillons de CQ doivent être analysés avec les échantillons inconnus.

Le guide précise que deux répétitions de QC à trois niveaux de concentration devraient être intégrées dans chaque série de dosage : un près de la limite inférieure de quantification (LQI;  $3 \times \text{LQI}$ ), un autre en milieu de gamme, et un près de l'extrémité supérieure de la fourchette. Les résultats des échantillons CQ fourniront la base d'acceptation ou de rejet de la série. Au moins quatre échantillons sur les six CQ devraient être dans les 15% de leur valeur nominale respective. Cette règle est appelée règle de 4-6-15. Deux des six échantillons QC peuvent être en dehors des 15% de leur valeur nominale respective, mais pas les deux à la même concentration.

Cette règle de 4-6-15 est l'une des nombreuses méthodologies qui permettent d'évaluer si un processus ou une méthode d'analyse est stable dans le temps. Concernant cette règle particulière, des critiques ont été faites: les risques producteur et consommateur sont très mal contrôlés. Pour renforcer le contrôle de ces risques, il est nécessaire d'analyser un plus grand nombre total d'échantillons (par exemple: 8-12-15 , 12-18-15), et afin de contrôler adéquatement les risques pour le consommateur, le nombre total des échantillons CQ à analyser doit être supérieure à 80 (pour avoir au plus 10% de chances d'observer au moins 80% des résultats dans les limites d'acceptation, la règle doit être : 74-86-15) [2;42; 31].

Toutefois, d'autres types de méthodes de contrôle statistique existent ; les plus courantes sont les cartes de Shewhart, les cartes des sommes cumulées (CUSUM), les cartes des moyennes pondérées exponentielles (EWMA) [2,43].

## 2. Concept d'erreur totale et Approche du Profil d'exactitude

### 2.1. Concept d'erreur totale

#### 2.1.1. Introduction

Durant les années 50 et 60, les analystes chimistes qualifiaient respectivement la performance des méthodes d'analyse en terme de justesse et de fidélité, ou d'erreurs systématiques et aléatoires.

La fidélité et la justesse (ou les erreurs aléatoires et systématiques) sont des caractéristiques de la performance d'une procédure d'analyse quantitative. L'Erreur totale décrit la qualité d'un résultat de mesure (test) et fournit une estimation de l'erreur maximale qui peut survenir pour une seule mesure. L'erreur totale est essentiellement utilisée dans le laboratoire pour juger de l'acceptabilité d'une procédure d'analyse pour son utilisation prévue.

L'Erreur Totale (ET) représente l'erreur globale ou totale qui peut survenir dans le résultat d'une mesure (test) en raison à la fois de problèmes de fidélité (erreur aléatoire) et de justesse (erreur systématique) de la procédure de mesure. Elle est communément définie comme  $ET = \text{Biais} + Z * SD$ , où le biais est l'estimation de l'erreur systématique, SD (écart type) est l'estimation de l'erreur aléatoire, et Z est un multiplicateur qui représente le niveau de confiance souhaité.

L'objectif de l'utilisation de l'erreur totale est d'estimer (décrire) l'erreur maximale qui pourrait se produire dans un résultat d'analyse obtenu à partir d'une procédure de mesure. Dans la validation analytique d'une méthode, l'erreur totale mesure la qualité d'une mesure qui peut être comparée à une qualité attendue de cette même mesure et, qui peut être exprimée en terme d'erreur totale acceptable (ETA) ou de limite d'acceptation nommée souvent  $\lambda$ .



Bref, l'erreur totale est une caractéristique de qualité des résultats individuels alors que la fidélité et la justesse sont des caractéristiques de performance de la méthode analytique qui contribuent à la qualité d'un résultat de mesure. Différentes combinaisons de fidélité et de justesse peuvent produire la même qualité pour un résultat de test, et il est donc préférable de fixer des objectifs pour l'erreur totale acceptable, plutôt que de fixer des objectifs individuels pour l'écart type et le biais

La norme ISO 15189 [44] reconnaît indirectement cette exigence importante dans sa déclaration à la section 5.5:

- Les spécifications de performance de chaque procédure d'analyse utilisée doivent être en relation avec l'utilisation prévue de cette procédure

Cela implique que l'estimation de la fidélité et de la justesse d'une méthode d'analyse doit être adaptée à son utilisation prévue. Pour ce faire, le laboratoire doit définir quelle est la qualité du test recherché en relation avec cette utilisation (contrôle de matière première ou du produit fini ; recherche des impuretés ou dosage du principe actif...), ce qui conduit à la nécessité de définir une exigence qualité en appliquant des limites d'acceptation pour chaque type de test et en considérant la notion d'erreur totale qui est la somme de la fidélité et de la justesse en phase de validation[45].

Notons cependant que l'erreur totale n'est pas un concept reconnu dans le Vocabulaire international de métrologie (VIM) [46] ou le GUM [47].

Pour les laboratoires pharmaceutiques et biopharmaceutiques, une prise de conscience de la notion d'erreur totale (appelée aussi erreur totale d'analyse ou erreur de mesure totale) est récemment apparue. Par exemple, en Europe, un groupe de travail qui comprend des industriels, des universitaires et les autorités réglementaires de l'industrie pharmaceutique et agro-alimentaire, a promu l'utilisation de l'erreur totale pour évaluer la validité des méthodes analytiques [28, 30,37]. De même, les industriels et les autorités réglementaires [par exemple, la US Food and Drug Administration (FDA)], dans un récent résumé de l'atelier bioanalytique de 2006, conclu avec un nouveau critère, à savoir l'erreur totale, afin d'évaluer la validité des dosages immunologiques (Ligand Binding Assay ou LBA [48]). Cette

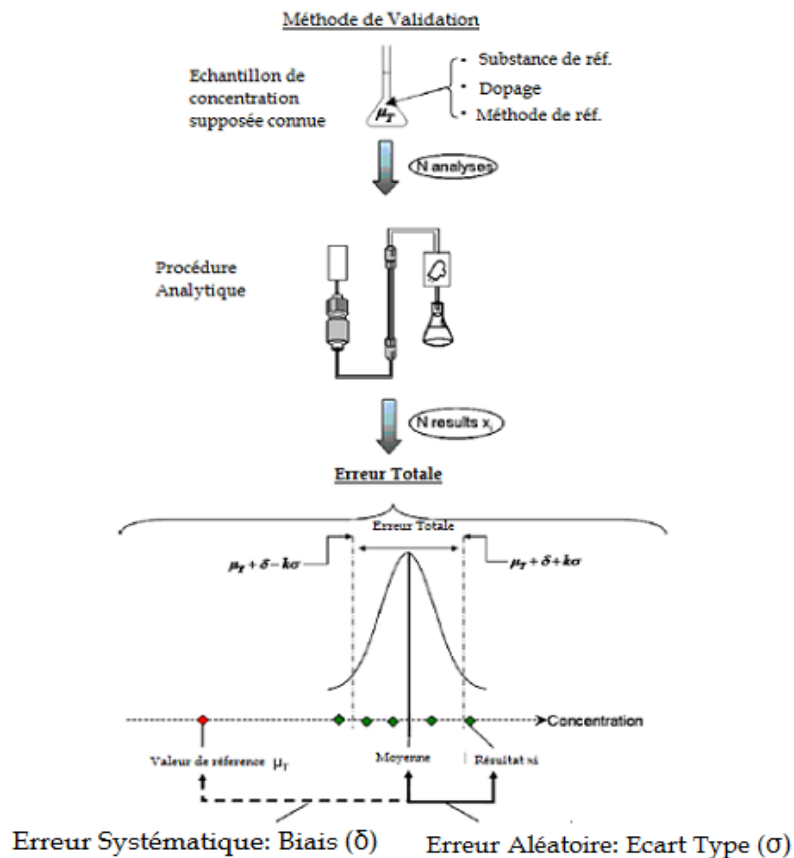
application récente de l'erreur totale dans ces secteurs a été faite afin de satisfaire les demandes les plus exigeantes des autorités réglementaires sur la gestion des risques, en particulier, le risque consommateur/client [49,50] (par exemple, le risque de déclarer acceptable un résultat inacceptable). [51] Enfin cette notion a été introduite dans le dernier guide de l'EMEA relatif à la validation des méthodes bioanalytiques (Guideline on bioanalytical method validation)[52]

### 2.1.2. Types d'erreur

En règle générale, deux principaux types d'erreur sont reconnus en chimie analytique: l'erreur systématique et l'erreur aléatoire, comme le montre la Figure 12 [20, 46, 53]. Pour estimer l'erreur aléatoire, les variances, les écarts-types ou les coefficients de variation sont calculés sur la base des mesures répétées du même échantillon. Pour estimer l'erreur systématique, la première étape consiste à calculer la moyenne de plusieurs répétitions de mesures de même échantillon, puis l'écart à une valeur de référence est calculé afin d'estimer le biais.

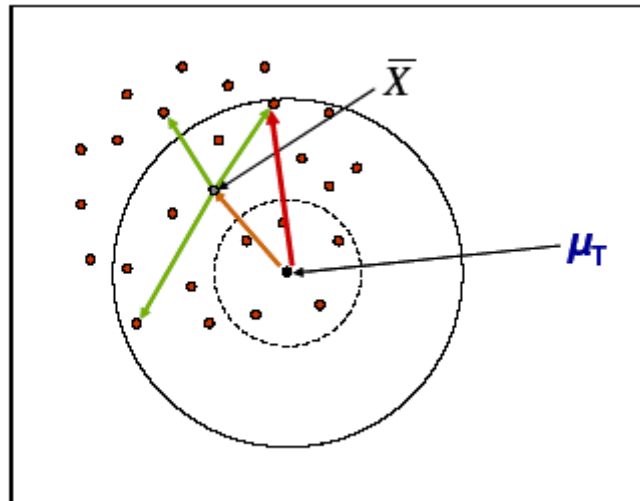
Ces deux composantes d'erreur sont faciles à estimer lorsque plusieurs répétitions d'un échantillon avec une concentration de référence ou une valeur conventionnellement vraie sont analysées. L'évaluation séparée de ces deux parties est souvent appliquée dans la validation des méthodes, dans le transfert, dans la comparaison des méthodes ou dans les études inter-laboratoires [2].

Enfin, certains auteurs ont proposé de subdiviser ces deux grandes catégories d'erreurs afin de diagnostiquer les sources les plus importantes d'erreurs [19, 20,45, 54-57]. La détermination de la quantité relative de chaque source est essentielle, si l'on vise à améliorer les performances analytiques d'une méthode (par exemple, en réduisant le biais dont l'origine est due à des facteurs contrôlables, tels que les opérateurs ou les équipements). Ce point est d'un intérêt surtout pour les laboratoires de R & D ou les fabricants des kits de dosage.



**Figure II-12. L'erreur totale pour la validation des méthodes. \* La concentration de l'échantillon n'est jamais connue parfaitement, c'est, en général, la valeur de référence ou valeur vraie par convention qui est utilisée [51].**

Cependant, quand seulement une seule mesure est effectuée, comme lors de l'analyse en routine, il est impossible de différencier ces deux types d'erreur. Comme l'illustre la Figure 13, une seule mesure est soumise à l'ensemble des erreurs d'analyse, c'est-à-dire de la combinaison simultanée des composantes systématiques et aléatoires de l'erreur [2].



**Figure II-13- Représentation schématique des composantes de l'erreur; flèche orange: Erreur systématique ou biais, flèche verte : erreur aléatoire ou écart type ; Flèche rouge : erreur totale.  $\bar{X}$  est la moyenne des résultats (points rouges) et  $\mu_T$  est la vraie valeur.**

Donc, l'erreur totale est une mesure de la distance entre tout résultat individuel et la valeur de référence de la concentration de l'analyte présent dans l'échantillon. L'erreur d'analyse totale ou l'erreur totale est donc la combinaison simultanée des parties systématiques et aléatoires de l'erreur.

### 2.1.3. Calcul de l'erreur totale

Comme on a vu précédemment, l'objectif principal de l'erreur totale est de fournir la quantité globale d'erreur d'analyse liée à n'importe quel résultat de mesure. Une étape principale dans le cycle de vie des méthodes d'analyse où l'estimation de l'erreur totale est importante est l'étape de validation analytique. En effet, des échantillons de concentration considérée comme connue (par exemple, la valeur réelle conventionnelle ou la valeur de référence) sont analysés afin de satisfaire aux critères de validation exigés par la plupart des lignes directrices réglementaires [19,21,22,44,48,58-61]. L'objectif final de cette étape est de vérifier si la méthode d'analyse fournira des résultats fiables et exacts. Le calcul de l'erreur totale permet d'atteindre cet objectif et ce en estimant la quantité globale de l'erreur d'analyse à chaque niveau de concentration étudié et en la comparant à une limite de mesure acceptable [28,46, 51,62,68].

Cette limite acceptable est généralement définie par les autorités réglementaires, par des spécifications provenant d'un processus de production industrielle ou par les exigences du client. Ainsi, il y a une distribution d'erreur autour de la valeur de référence, telle que donnée par l'équation suivante:

$$\begin{aligned} x_i &= \mu_T \pm f(\text{Erreur Totale}) \\ &\Downarrow \\ x_i &= \mu_T \pm f(\delta, \sigma) \end{aligned} \quad \text{Eq. II-5}$$

où  $\delta$  et  $\sigma$  sont, respectivement, le biais et l'écart-type,  $\mu_T$  est la valeur de référence ou la valeur conventionnellement « vraie » et  $x_i$  est le résultat de l'analyse.

L'équation (Eq. 5) donne la région où chaque résultat d'analyse peut être trouvé autour de la concentration de référence d'analyte présente dans l'échantillon et est illustré dans la Figure 12. En d'autres termes, il définit les distributions prédictives des résultats d'analyse  $x_i$ , pour une concentration de référence définie  $\mu_T$ .

Mandel (1964) a montré [63] comment la différence entre un résultat d'une mesure et la valeur de référence peut être décrite comme la combinaison d'erreurs aléatoires et systématiques, définie comme l'erreur totale:

$$ET = x_i - \mu_T = (x_i - \hat{\mu}) + (\hat{\mu} - \mu_T) \quad \text{Eq. II-6}$$

Avec ET est l'erreur totale du  $i^{\text{ème}}$  résultat  $x_i$ ,  $\mu_T$  est le résultat de référence et  $\hat{\mu}$  est la moyenne estimée de la population des résultats.

Cette équation peut être étendue pour le cas de  $n$  analyses répétées d'échantillons [64]:

$$ET = \sum_{i=1}^n (x_i - \mu_T) = \sum_{i=1}^n (x_i - \hat{\mu}) + (\hat{\mu} - \mu_T) \quad \text{Eq. II-7}$$

Cependant, avec cette formule, le deuxième terme va toujours sous-estimer la partie aléatoire puisque la distance positive entre chaque valeur et la moyenne sera toujours compensée par son équivalente négative. Par conséquent, ce terme devrait être remplacé par la différence des sommes des carrées, c'est à dire qu'elle représente une mesure de la fidélité ou erreur aléatoire:

$$ET = \sum_{i=1}^n (x_i - \mu_T) = \sum_{i=1}^n (x_i - \hat{\mu})^2 + (\hat{\mu} - \mu_T) \quad \text{Eq. II-8}$$

A partir de cette définition initiale, différents modèles pour combiner les deux parties d'erreur ont été proposées.

Pour les études de validation, nous avons besoin d'obtenir la région où les résultats d'analyse sont obtenus autour de valeurs de référence, et ces modèles d'erreur totale sont donc écrits comme suit :

$$\text{Modèle ET1 : } x_i = \mu_i \pm k\sqrt{\sigma^2 + \delta^2}$$

$$\text{Modèle ET2 : } x_i = \mu_T \pm \sqrt{k^2\sigma^2 + \delta^2}$$

$$\text{Modèle ET3 : } x_i = \mu_T \pm k\sigma + \delta$$

La valeur de k est fixée en fonction de la garantie qu'un laboratoire est disposé à fournir et dépend de l'hypothèse de la distribution des résultats. Habituellement, k est fixé à 2, exprimant ainsi une garantie de 95% en supposant une distribution normale des résultats avec une moyenne connue exprimée par le biais connu ( $\delta$ ) et d'écart-type connu ( $\sigma$ ).

Le modèle le plus utilisé est le modèle linéaire, où les erreurs aléatoires et systématiques sont combinées linéairement selon le modèle ET3. z est un facteur de probabilité, généralement égal à 1,65 ou 1,96 [65,66]. Ce modèle a été principalement utilisé dans la chimie clinique [68-70].

Toutefois, dans les secteurs pharmaceutiques et biopharmaceutiques, nous remarquons un début de tendance vers l'utilisation de l'erreur totale comme critère d'évaluation de validation des méthodes. En effet, une commission SFSTP a , d'une part , fourni un guide sur

la validation des méthodes bioanalytiques [27] où le modèle pour déclarer une méthode valide est construit comme suit:

$$\hat{\delta} \pm Q_{t(\alpha/2; \nu)} \hat{\sigma}_{I.P.}$$

avec  $\hat{\delta}$  est le biais estimé,  $\hat{\sigma}_{I.P.}$  est l'écart type estimé de la fidélité intermédiaire et  $Q_{t(\alpha/2; \nu)}$  est le quantile de la distribution de Student avec  $\nu$  degré de liberté

Il s'agit donc d'une estimation de l'erreur totale d'analyse puisque l'erreur systématique s'additionne à l'erreur aléatoire. Deuxièmement, comme nous l'avons vu à la section 1.3 Validation, une récente commission de la SFSTP [28] a également proposé d'utiliser le concept d'erreur totale pour évaluer la validité des méthodes analytiques. Enfin, à la fin d'un atelier commun, la FDA et l'AAPS ont convenu de proposer une règle pour évaluer la validité des méthodes bioanalytiques fondée sur une erreur totale maximale de 30% [48]. Cette règle a d'ailleurs été introduite dans le récent guide de validation des méthodes bioanalytiques de l'EMA [52]. Egalement, dans un récent éditorial du Journal of Chromatography B, l'implication du critère d'erreur totale proposé par la FDA / AAPS [48] est discutée en relation avec la révision des manuscrits soumis à cette revue [67].

Même si les méthodologies présentées sont très diverses et imparfaite, la principale préoccupation est que l'erreur totale est donc liée à la fiabilité globale des résultats générés par la méthode d'analyse. En effet, si l'erreur totale est élevée, les résultats générés par le procédé peuvent être de qualité insuffisante pour l'objectif initial de la méthode d'analyse. Par conséquent, l'erreur totale maximale ou le niveau de fiabilité acceptable doit être stipulé par l'utilisateur final des résultats et est liée à l'objectif spécifique de l'analyse.

En outre, il n'est généralement pas important que la méthode d'analyse ait une faible erreur systématique ou une faible erreur aléatoire, tant que la combinaison des deux composantes de l'erreur est acceptable. En effet l'utilisateur final des résultats pense en termes d'erreur totale. Les façons de combiner simultanément d'une manière statistique correcte les erreurs systématiques et aléatoires doivent donc être trouvées. Cette question est examinée dans la section suivante traitant des intervalles de tolérance statistiques.

## 2.2. Rôle des intervalles de tolérance

Pour comprendre le rôle potentiel des intervalles de tolérance dans l'industrie pharmaceutique et en particulier durant les étapes du cycle de vie des méthodes d'analyse, il est important d'examiner d'abord les principaux types d'intervalles statistiques qui existent.

### 2.2.1. Intervalle de confiance de la moyenne d'un échantillon

#### a. Intervalle de confiance

L'intervalle de confiance de la moyenne d'un échantillon est probablement l'intervalle statistique le plus connu. Au sens courant, c'est une "fourchette".

Un intervalle de confiance bilatéral à 95% de la moyenne  $m$  d'une variable est un couple de variables aléatoires  $(L, U)$  telles que  $P(L \leq \mu \leq U) = 0.95$ . L'intervalle est défini par une probabilité de recouvrement : celle que la vraie valeur de  $m$  (inconnue) soit contenue dans l'intervalle aléatoire  $[L, U]$ . [68]".

Soit par exemple  $x_1, x_2, x_3, \dots, x_k$   $k$  résultats obtenus à partir d'une méthode d'analyse quantitative. La valeur moyenne de cet échantillon de résultats peut être facilement calculée par:

$$\hat{\mu} = \frac{\sum_{i=1}^k x_i}{k} \quad \text{Eq. II-9}$$

Et la formule de l'intervalle de confiance de la moyenne avec  $1-\alpha$  comme niveau de confiance:

$$\hat{\mu} \pm Q_{(k-1, \alpha/2)} \frac{\hat{\sigma}}{\sqrt{k}} \quad \text{Eq. II-10}$$

où  $\hat{\sigma}$  est l'écart type estimé à partir de l'échantillon et  $Q_{(k-1, \alpha/2)}$  est la quantile  $\alpha/2$  de la distribution de Student avec  $k-1$  degré de liberté.

Cet intervalle exprime que la vraie valeur  $\mu$  de la moyenne de l'ensemble de la population est comprise dans cet intervalle avec une probabilité de  $1-\alpha$ . Ainsi, si les expériences sont



répétées cent fois, 95 fois sur 100, la vraie moyenne  $\mu$  sera en moyenne comprise dans les intervalles de confiance calculés.

D'après ce que dit la théorie statistique, la valeur  $\mu$  ne pourrait être connue qu'après avoir effectué un nombre infini de mesures. C'est donc un paramètre théorique qui sert à une modélisation mais qui n'est jamais connu. Pour remédier à cet inconvénient, on estime ce paramètre à partir d'un nombre réduit de mesures : la moyenne n'est donc qu'une estimation de la valeur vraie et, si on effectuait une autre série de mesures, on n'obtiendrait pas tout à fait la même moyenne. Il y a donc un risque que le résultat final retenu soit différent de  $\mu$  [1]

L'intervalle de confiance de la moyenne permet non pas de supprimer ce risque, mais de le quantifier. Pour construire l'intervalle de confiance de la moyenne, on définit un ensemble de valeurs parmi lesquelles il y a un faible risque que la valeur vraie  $\mu$  ne se trouve pas. Ce risque d'erreur, compris entre 0 et 1 (ou entre 0 % et 100 %) est appelé risque d'erreur de première espèce et sera noté  $\alpha$ . Son complément  $1 - \alpha$  est le niveau de confiance de l'intervalle [1].

Si  $\alpha$  est très petit, 1% par exemple, on peut avoir une confiance élevée dans l'intervalle choisi et il y a peu de chances que la valeur vraie ne s'y trouve pas, l'intervalle étant grand. Inversement, si le risque est grand, 10 % par exemple, on a un risque plus élevé que la valeur vraie ne s'y trouve pas et l'intervalle est alors plus petit [1].

L'idée derrière ce calcul de l'intervalle de confiance est facile à comprendre pour un analyste car elle permet de répondre à une question classique :

«combien de répétitions dois-je faire? ». En effet, c'est un statisticien connu sous le nom de Student qui a proposé une réponse à cette question en essayant de quantifier dans quelle mesure l'estimation de la moyenne est améliorée lorsqu'on augmente le nombre de répétitions: en d'autres termes, il a proposé une méthode pour évaluer dans quelle proportion on connaît mieux l'intervalle de confiance d'une moyenne si on l'estime avec 2, 3, 4 ... 100 répétitions ou plus.

'(rd(r(dr(d

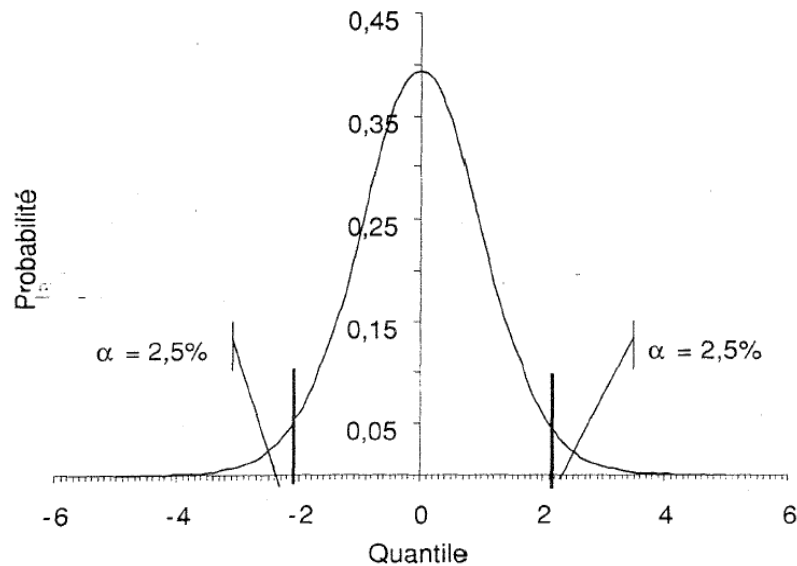


Figure II-14- Densité de probabilité d'une loi de Student à 19 degrés de liberté [1]

#### b. Intervalle de tolérance

L'**intervalle de tolérance**, également appelé **intervalle de prédiction**, est de son côté moins connu. C'est un intervalle dans lequel on est capable de prédire où va se trouver en moyenne une proportion connue des mesures. Il diffère donc de l'intervalle de confiance en cela qu'il s'intéresse à la population des mesures alors que l'IC ne s'intéresse qu'à un paramètre, comme la moyenne ou l'écart-type. En termes simples, l'intervalle de confiance caractérise le comportement de la moyenne tandis que l'intervalle de tolérance caractérise le comportement de l'ensemble des mesures. Il est donc très intéressant dans l'optique d'un contrôle de procédé et il est dommage que la notion d'intervalle de tolérance ne soit pas plus souvent évoquée. C'est sans doute parce qu'elle est un peu plus délicate à définir.

Cette notion a été introduite pour le contrôle de fabrication mais n'est pas limitée à ce cas. En effet, dans un contrôle de la qualité d'un processus de production, comme la production d'un résultat d'analyse, ce n'est pas tant le résultat moyen de l'échantillon qui importe, mais plutôt entre quelles valeurs va se situer une forte proportion des mesures. Connaître la localisation de la moyenne vraie est d'autant moins intéressant que pour des raisons économiques, il est souvent avantageux de ne pas faire de répétitions, tout en essayant de définir une stratégie qui permette de rejeter au mieux des lots défectueux ou hors spécification.

Si on connaissait de manière très précise (certaine) la moyenne  $\mu$  et l'écart-type  $\sigma$  du lot et que les mesures suivaient effectivement une loi normale, alors on déterminerait des limites grâce à la formule  $[\mu - u_{95\%}\sigma; \mu + u_{95\%}\sigma]$ . Par exemple, avec  $u_{95\%} = 1,96$  correspondant au quantile 95 % de la loi normale. On pourrait alors affirmer que 95 % des mesures obtenues se situent entre ces deux limites.

C'est la démarche utilisée pour définir un intervalle de répartition. Mais, comme la moyenne et l'écart-type sont estimés à partir d'un échantillon de mesures et il n'est plus possible de procéder ainsi. La moyenne et variance empiriques sont des variables aléatoires et non plus des paramètres fixes. On obtiendrait, en faisant différentes séries de mesures, des limites contenant tantôt, soit, plus de 95 % des valeurs, soit moins de 95 %. Ce qui ne résout pas le problème.

Pour calculer l'intervalle de tolérance, on va rechercher de manière indirecte des limites qui permettent d'affirmer qu'en moyenne une proportion, notée  $\beta$ , des mesures se situent entre ces limites bilatérales. En gardant les notations proposées, l'intervalle de tolérance s'exprime ainsi :

$$\left[ \bar{X} - K_{IT} \times S \times \sqrt{1 + \frac{1}{I}}; \bar{X} + K_{IT} \times S \times \sqrt{1 + \frac{1}{I}} \right] \text{ Intervalle de tolérance (Eq. II-11)}$$

où  $k_{IT}$  est le facteur de couverture de l'intervalle de tolérance et  $I$  le nombre de données

Cette formule ressemble à celle de l'intervalle de confiance, fourni par l'équation (Eq. 10), à deux différences près. D'abord, l'écart-type  $s$  est multiplié par un coefficient qui tient compte du nombre de mesures. Ensuite, l'intervalle est supposé contenir une proportion donnée de futures mesures  $X$  et non plus la moyenne théorique  $\mu$  de la distribution.

Pour une loi normale,  $k_{IT}$  est fournie par la formule suivante [73]:

$$K_{IT} = t_{I-1; \frac{1-\beta}{2}} \text{ Facteur de couverture de l'intervalle de tolérance (Eq. II-12)}$$

Dans l'équation (Eq 12) la quantité  $t_{I-1; \frac{1-\beta}{2}}$  représente le quantile de la loi de Student pour  $I-1$  degré de liberté et la probabilité  $(1-\beta)/2$ .

### 2.2.2. Relation entre l'Intervalle de Tolérance et Erreur Totale

Suite à cette définition de l'intervalle de tolérance, on peut établir une relation entre celui-ci et l'erreur totale. En effet, cet intervalle est toujours construit sur la base de la moyenne et l'écart-type d'un échantillon d'un produit, par exemple. S'il est exprimé par rapport à la cible de la production d'un médicament ou par rapport à la valeur réelle des analytes dans un échantillon, l'intervalle de tolérance combine donc simultanément les erreurs aléatoires et systématiques d'une procédure analytique. Dans ce cas, la valeur moyenne estimée est remplacée par le biais estimé de la méthode d'analyse ou du processus de production. C'est donc un moyen statistique commode d'exprimer l'erreur totale [30, 62].

Les intervalles de tolérance peuvent être extrêmement utiles si on veut comparer les performances réelles d'un processus de production ou des méthodes d'analyse. En effet, pour les produits au moins, les spécifications sont établies afin d'évaluer leur qualité et leur donner l'autorisation d'être commercialisés. De l'examen d'un échantillon des produits, il est possible pour les producteurs de construire des intervalles statistiques de tolérance qui peuvent être comparés aux spécifications du produit et donc décider de sa libération sur le marché.

La même méthodologie peut être appliquée pour évaluer la validité des méthodes analytiques.

Hoffman et al., (2005 et 2007) [33, 62] ont utilisé un intervalle de tolérance de type «  $\beta$  – content » pour évaluer la validité des méthodes bio-analytiques, alors que, une commission SFSTP [30] a proposé d'utiliser l'intervalle de tolérance «  $\beta$  – expectation » pour la validation des méthodes d'analyse. Brown et al. [69] ont proposé une méthodologie basée sur des intervalles de tolérance pour évaluer la bioéquivalence individuelle de deux formulations pharmaceutiques. Les intervalles de tolérance ont également été utilisés pour les tests de contrôle de qualité de l'uniformité de teneur d'un lot avant la libération du lot sur le marché [70-73]. Les intervalles de tolérance pourraient également être utiles pour estimer la limite thérapeutique d'un médicament en dose études de recherche en veillant à ne pas tomber dans le niveau de toxicité ou d'inefficacité [79], ou en évaluant la sécurité des essais cliniques [80]

### 3. Bibliographie

- [1] Feinberg. M -LABO-STAT - Guide de validation des méthodes d'analyse; Lavoisier Edition Tec & Doc 2009
- [2] Rozet E., Improvement of the predictive character of test results issued from analytical methods life cycle, Thèse présentée en vue de l'obtention du grade de Docteur en Sciences Pharmaceutiques, Université de Liège, 2007-2008
- [3] Feinberg M., Boulanger B., Dewe W., Hubert Ph., New advances in method validation and measurement uncertainty aimed at improving the quality of chemical data, *Analytical and Bioanalytical chemistry*, 380, 502 – 514 (2004)
- [4] Marini R., Evaluation des différentes approches pour l'estimation de l'incertitude des Mesures analytiques, Thèse présentée en vue de l'obtention du grade de Docteur en Sciences Pharmaceutiques, Université de Liège, 2005-2006
- [5] Duarte A.C., Capelo S., 2006, Application of Chemometrics in Separation Science, *Journal of Liquid Chromatography & Related Technologies*, 29, 114-1176.
- [4] Lantéri P., Longerey R. Chimométrie : le mariage réussi entre les sciences analytiques et l'informatique, *Analisis*, 24, M17 – M27 (1996)
- [6] Goupy J., Stratégie de recherche, Définition et objectifs des méthodes des plans d'expériences, Dunod, Paris, France (1996)
- [7] Goupy J., Plans factoriels à deux niveaux, La méthode des plans d'expériences, Dunod, Paris, France (1996)
- [8] Albert A., Analyses statistiques multivariées, Edition de l'ULg, Liège (2006)
- [9] Sado G., Sado M.-C., De l'expérimentation à l'assurance de qualité in Les plans d'expériences, AFNOR, Paris, France (1991)
- [10] Fotsing L., Développement et validation de méthodes de dosage des différents constituants de formulations polyvitaminées par électrophorèse capillaire, Thèse présentée en vue de l'obtention du grade de Docteur en Sciences Pharmaceutiques, Université de Liège, 1998 – 1999
- [11] Chiap P., Association de la dialyse à la chromatographie liquide pour le dosage automatisé de médicaments dans le plasma, Thèse présentée en vue de l'obtention du grade de Docteur en Sciences Pharmaceutiques, Université de Liège, 1999 – 2000.
- [12] Servais A.-C., séparation et dosage de substances médicamenteuses par électrophorèse capillaire en milieux non aqueux : influence de l'utilisation simultanée d'une cyclodextrine et d'un agent d'appariement d'ions, Thèse présentée en vue de l'obtention du grade de Docteur en Sciences Pharmaceutiques, Université de Liège, 2004 – 2005.
- [13] Goupy J., Creighton L., Introduction aux plans d'expériences, Dunod, Paris, France (2001)
- [14] Dean A., Voss D., Design and analysis of experiments, Springer, New York (1999)
- [15] Mason R.L., Statistical design and analysis of experiments, Wiley, New Jersey (2003)
- [16] Govaerts B., Simar L., Planification expérimentale, Eli-Lilly, Mont-Saint-Guibert, Belgique, (1996)
- [17] Feinberg M., Optimisation de la réponse, La validation des méthodes d'analyse – Une approche chimométrique de l'assurance qualité du laboratoire, Masson, Paris, France (1996)
- [18] Goupy J., plans d'expériences, Techniques de l'ingénieur, PE230 (1997)

- [19] International Conference on Harmonization (ICH) of Technical Requirements for registration of Pharmaceuticals for Human Use, Topic Q2 (R1): Validation of Analytical Procedures: Text and Methodology, Geneva, Switzerland, 2005.
- [20] Application of the statistics-accuracy (trueness and precision) of the results and methods of measurement – Parts 1 to 6, International Organization of Standardization (ISO), ISO 5725, ISO, Geneva, Switzerland, 1994.
- [21] US Federal Government, Guidance for industry: Bioanalytical Method Validation, US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Rockville, MD, USA, May 2001.
- [22] The Fitness for Purpose of Analytical Methods A Laboratory guide to method validation and related Topics, Eurachem, Teddington, Middlesex, UK, 1998.
- [23] IUPAC Technical Report Harmonized guidelines for single laboratory validation of methods of analysis 2002
- [24] AOAC, Official Methods of Analysis, Vol. 1, Association of Official Analytical Chemists (AOAC), Arlington, VA, 15th ed., pp. 673-674, 1990,
- [25] Caporal-Gautier J., Nivet J.M., Algranti P., Guilloteau M., Histe M., Lallier L., N'Guyen-Huu J.J., Russotto R., Guide de validation analytique. Rapport d'une commission SFSTP. I : Méthodologie, *STP Pharma Pratiques*, 2, 205-226 (1992)
- [26] Hartmann C., Smeyers-Verbeke J., Massart D.L., McDowall R.D., Validation of bioanalytical chromatographic methods, *Journal of Pharmaceutical and Biomedical Analysis*, 17, 193-218 (1998)
- [27] Hubert Ph., Chiap P., Crommen J., Boulanger B., Chapuzet E., Mercier N., Bervoas-Martin S., Chevalier P., Grandjean D., Lagorce P., Lallier M., Laparra M.C., Laurentie M., Nivet J.C., The SFSTP guide on the validation of chromatographic methods for drug bioanalysis: from the Washington Conference to the laboratory, *Analytica Chimica Acta*, 391, 135-148 (1999)
- [28] Hubert Ph., Nguyen-Huu J.-J., Boulanger B., Chapuzet E., Chiap P., Cohen N., Compagnon P.-A., Dewe W., Feinberg M., Lallier M., Laurentie M., Mercier N., Muzard G., Nivet C., Valat L., Harmonization of strategies for the validation of quantitative analytical procedures, A SFSTP proposal - Part I, *Journal of Pharmaceutical and Biomedical Analysis*, 36, 579-586 (2004)
- [29] ISO 5725-1, International Organization for Standardization (ISO), Statistical methods for quality control, accuracy (trueness and precision) of measurement methods and results-part 1: General principles and definitions, Geneva, ISO vol. 1, 4<sup>th</sup> ed. (1994)
- [30] Hubert Ph., Nguyen-Huu J.-J., Boulanger B., Chapuzet E., Cohen N., Compagnon P.-A., Dewé W., Feinberg M., Laurentie M., Mercier N., Muzard G., Valat L., Rozet E., Harmonization of strategies for the validation of quantitative analytical procedures A SFSTP proposal–Part II, *Journal of Pharmaceutical and Biomedical Analysis*, 45, 70–81 (2007)
- [31] Boulanger B., Dewé W., Gilbert A., Govaerts B., Maumy-Bertrand M., Risk management for analytical methods based on the total error concept: Conciliating the objectives of the pre-study and in-study validation phases, *Chemometrics and Intelligent Laboratory Systems*, 86, 198–207 (2007)
- [32] Desilva B., Smith W., Weiner R., Kelley M., Smolec J., Lee B., Khan M., Tacey R., Hill H., Celniker A., Recommendations for bioanalytical method validation of ligand-binding

- assays to support pharmacokinetic assessments of macromolecules, *Pharmaceutical Research*, 20, 1885–1900 (2003)
- [33] Hoffman D., Kringle R., Two-Sided Tolerance Intervals for Balanced and Unbalanced Random Effects Models, *Journal of Biopharmaceutical Statistics*, 15, 283-293 (2005)
- [34] Boulanger B., Chiap P., Dewe W., Crommen J., Hubert Ph., An analysis of the SFSTP guide on validation of chromatographic bioanalytical methods: progresses and limitations, *Journal of Pharmaceutical and Biomedical Analysis*, 32, 753-765 (2003)
- [35] Hartmann C., Smeyers-Verbeke J., Penninckx W., Vander Heyden Y., Vankeerberghen P., Massart D., Reappraisal of hypothesis testing for method validation : detection of systematic error by comparing the means of two methods of two laboratories, *Analytical Chemistry*, 67, 4491-4499 (1995)
- [36] Schuirmann D.J., A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, 15, 657-680 (1987)
- [37] Hubert Ph., Nguyen-Huu J.-J., Boulanger B., Chapuzet E., Cohen N., Compagnon P.- A., Dewé W., Feinberg M., Laurentie M., Mercier N., Muzard G., Valat L., Rozet E., Harmonization of strategies for the validation of quantitative analytical procedures A SFSTP proposal–Part III, *Journal of Pharmaceutical and Biomedical Analysis*, 45, 82–96 (2007)
- [38] ISO 17025, General requirements for the competence of testing and calibration laboratories. ISO, Geneva, Switzerland (2000)
- [39] Rius A., Ruisanchez I., Callao M.P., Rius F.X., Reliability of analytical systems: use of control charts, time series models and recurrent neural networks (RNN), *Chemometrics and Intelligent Laboratory Systems*, 40, 1–18 (1998)
- [40] Doerffel K., Herfurth G., Liebich V., Wendlandt E., The shape of CUSUM – an indicator for tendencies in a time series, *Fresenius Journal of Analytical Chemistry*, 341, 519–523 (1991)
- [41] Doerffel K., Assuring trueness of analytical results, *Fresenius Journal of Analytical Chemistry*, 348, 183–187 (1994)
- [42] Kringle R.O., An assessment of the 4-6-20 rule for acceptance of analytical runs in bioavailability, bioequivalence, and pharmacokinetic studies, *Pharmaceutical Research*, 11, 556-560 (1994)
- [43] Montgomery D.C., *Design and Analysis of Experiments*, John Wiley & Sons, USA, 643 p (2005)
- [44] ISO/FDIS 15189 Medical laboratories – Particular requirements for quality and competence. International Organization for Standards, Geneva Switz (2002)
- [45] Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method development and evaluation. *Clin Chem*;20:825 33 (1974)
- [46] Joint Committee for Guides in Metrology, JCGM 200:2008, International Vocabulary of Metrology – Basic and general concepts and associated terms (VIM), 2008 (available from the website: [www.bipm.org](http://www.bipm.org)).
- [47] Joint Committee for Guides in Metrology, JCGM 100:2008, Evaluation of measurement data – Guide to the expression of uncertainty in measurement (GUM), 2008 (available from the website: [www.bipm.org](http://www.bipm.org)).
- [48] Viswanathan C.T., Bansal S., Booth B., De Stefano A.J., Rose M.J., Sailstad J., Shah V.P., Skelly J.P., Swann P.G., Weiner R., — *Quantitative Bioanalytical Methods Validation*

- and Implementation: Best Practices for Chromatographic and Ligand Binding Assays, AAPS J 9 E30 (2007)
- [49] International Conference on Harmonization (ICH): Quality Risk Management Q9, Geneva, Switzerland, 2005 (available from the website: <http://www.emea.eu.int/Inspections/docs/ICHQ9-Step4QRM.pdf>).
- [50] US Food and Drugs Administration (US FDA), Process Analytical Technology (PAT) Initiative, US FDA, Rockville, MD, USA, 2004 (available from the website: <http://www.fda.gov/cder/OPS/PAT.htm> ICH Q9).
- [51] Rozet E., Marini R.D., Ziemons E., Dewe W., Rudaz S., Boulanger B., Hubert Ph., Trends in Analytical Chemistry, Total error and uncertainty: Friends or foes? Vol. 30, No. 5, (2011)
- [52] Guideline on bioanalytical method validation Committee for Medicinal Products for Human Use (CHMP) EMEA/CHMP/EWP/192217/2009 21 July 2011
- [53] Linsinger T.P.J., Use of recovery and bias information in analytical chemistry and estimation of its uncertainty contribution ?, Trends Anal. Chem. 27, 916 (2008)
- [54] Krouwer J.S., Estimating total analytical error and its sources. Techniques to improve method evaluation Arch. Path. Lab. Med. 116,726 (1992)
- [55] Krouwer J.S., Setting performance goals and evaluating total analytical error for diagnostic assays, Clin. Chem. 48, 919 (2002)
- [56] Rico's C., Baadenhuijsen H., Libeer JC., Petersen P.H., Stockl D., Thienpont L., Fraser CC. External quality assessment: currently used criteria for evaluating performance in European countries and criteria for future harmonization., Eur. J. Clin. Chem. Biochem. 34,159 (1996)
- [57] Hund E., Massart D.L., Smeyers-Verbeke J., Operational definitions of uncertainty, Trends Anal. Chem. 20, 394 (2001)
- [58] Thompson M., Ellison S.L.R., Wood R., Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC Technical Report), Pure Appl. Chem. 74, 835 (2002)
- [59] International Organization of Standardization (ISO), ISO/CEI 17025: General requirements for the competence of testing and calibration laboratories, ISO, Geneva, Switzerland, 2005.
- [60] European Commission, Method Validation and Quality Control procedures for Pesticide Residues, in: Doc. SANCO/2007/3131, 31 October 2007 (available at the website: <http://www.crlpesticides.eu/library/docs/srm/AqcGuidance.pdf>).
- [61] European Commission, 2002/657/EC: Commission Decision of 12 August 2002 implementing Council Directive 96/23/EC Concerning the Performance of Analytical Methods and the Interpretation of Results. Off. J. Eur. Commun. L221 (2002) 8.
- [62] Hoffman D., Kringle R., A total error approach for the validation of quantitative analytical methods, Pharm. Res. 24, 1157 (2007)
- [63] Mandel J., The statistical analysis of experimental data, New York, Dover, 104-105 (1964)
- [64] Krouwer J.S., Setting performance goals and evaluating total analytical error for diagnostic assays, *Clinical Chemistry*, 48, 919-927 (2002)
- [65] McFarren E.F., Lishka R.J., Parker J.H., Criterion for judging acceptability of analytical methods. *Analytical Chemistry*, 42, 358-65 (1970)



- [66] Petersen P. H., Stöckl D., Westgard J.O., Sandberg S., Linnet K., Thienpont L., Models for combining random and systematic errors. Assumptions and consequences for different models, *Clinical Chemistry and Laboratory Medicine*, 39, 589-595 (2001)
- [67] Bischoff R., Hopfgartner G., Karnes , H.T., Lloyd D.K., Phillips T.M., Tsikas D., Xu G., Summary of a recent workshop/conference report on validation and implementation of bioanalytical methods: Implications on manuscript review in the Journal of Chromatography B, *Journal of Chromatography B*, 860, 1–3 (2007)
- [68] Acton F. S., Analysis of straight-line data J.Wiley pages 47-51 (1959)
- [69] Brown, E.B., Iyer H.K., Wang C.M., Tolerance Intervals for Assessing Individual Bioequivalence, *Statistics in Medicine*, 16, 803-820 (1997)
- [70] Katori N., Aoyagi N., Kojima S., A proposal for revision of the content uniformity test and weight variation test. *Pharm. Forum*, 23, 5325–5333 (1997)
- [71] Hauck W.W., Shaik R., Sample sizes for batch acceptance from single- and multistage designs using two-sided normal tolerance intervals with specified content, *Journal of Biopharmaceutical Statistics*, 11, 335-346 (2001)
- [72] Williams R.L., Adams W.P., Poochikian G., Hauck W.W., Content Uniformity and Dose Uniformity: Current Approaches, Statistical Analyses, and Presentation of an Alternative Approach, with Special Reference to Oral Inhalation and Nasal Drug Products, *Pharmaceutical Research*, 19, 359-366 (2002)
- [73] Banfai B., Ganzler K., Kemeny S, Content uniformity and assay requirements in current regulations, *Journal of Chromatography A*, 1156, 206-212 .(2007)
- [74] Petzold M., A comparative study of some approaches for estimating tolerance limits, *Drug Information Journal*, 35, 1013–1018 (2001)
- [75] Nickens D.J., Using tolerance limits to evaluate laboratory data, *Drug Information Journal*, 32, 261–269 (1998)

**CHAPITRE III : Etude comparative des critères de validation à travers les différentes références réglementaires**

- 1. Introduction**
- 2. Spécificité & sélectivité**
- 3. Linéarité & fonction de réponse**
- 4. Fidélité**
- 5. Justesse**
- 6. Exactitude**
- 7. Limite de quantification**
- 8. Limite de détection**
- 9. Bibliographie**
- 10. Tableaux**

## 1. Introduction

Avant de procéder au contrôle-qualité de tout produit dans un laboratoire, la validation analytique constitue une étape obligatoire. Cependant, depuis plus de 20 ans, de nombreux guides ou recommandations plus ou moins officielles ont été publiées, définissant les critères de validation à évaluer pour démontrer la performance d'une méthode analytique. Toutefois, rien n'a été dit sur le niveau de garantie à atteindre pour l'estimation, en termes de nombre de mesures, de ces critères ni ce qu'on doit en faire par la suite. Que on, particulièrement, faire pour atteindre un niveau de qualité donnée [28] ou un objectif spécifié ? Il en ressort en outre les points faibles suivants [1] :

(A) Comment interpréter les définitions réglementaires de ces critères?

(B) Absence d'une méthodologie spécifique à suivre pour évaluer un critère particulier

(C) Absence de critères d'acceptation appropriés pour un critère donné

En outre, la validation des méthodes n'est pas spécifique à l'industrie pharmaceutique, mais elle est indispensable pour la plupart des industries impliquant des tests biologiques ou chimiques. Même si chaque domaine d'activité a ses propres caractéristiques et enjeux, les principaux critères à évaluer sont similaires ou devraient être semblables puisque la validation d'une méthode d'analyse est indépendante du secteur industriel, de la matrice des échantillons ou de la technique d'analyse utilisée [2].

Une terminologie harmonisée de validation devrait être adoptée pour permettre des discussions et des comparaisons des problèmes de validation entre les scientifiques de différents domaines. Ce consensus sur la terminologie n'est pas encore disponible, même si une tentative a été faite [2,3].

La terminologie utilisée entre les différents documents officiels telles que celle de la Food and Drug Administration (FDA) [4], de l'ICH [5], de l'IUPAC [6], de l'AOAC [7], de l'EURACHEM [8] ou de la SANCO [9] est différente.

Cependant, s'il est souhaitable d'avoir une harmonisation entre les différents domaines concernés par la validation analytique, il est intéressant de noter que, même dans le domaine pharmaceutique, les laboratoires n'utilisent pas la même terminologie alors qu'ils devraient adopter des définitions similaires pour décrire les critères de validation [1]

En outre, dans certains cas, comme nous allons le voir par la suite, une terminologie différente peut être trouvée dans un même document pour définir un critère et ce selon la section où elle est mentionnée.

Par conséquent, la connaissance et la compréhension de ces différences essentielles dans la terminologie et les définitions est essentielle car les méthodologies proposées pour répondre à la définition du critère peuvent prêter à confusion lors de la préparation du protocole de validation et du plan expérimental. L'interprétation ultérieure des résultats statistiques obtenus et la décision finale quant à la validité de la procédure d'analyse dépendent de la définition cohérente et adéquate des critères à évaluer.

Cela entraîne des conséquences très critiques puisque la méthode analytique validée sera utilisée quotidiennement dans les analyses de routine (libération des lots, établissement de la durée de validité, études pharmacocinétiques ou de bioéquivalence, etc) pour prendre des décisions avec des conséquences tout aussi importantes pour l'entreprise que pour la santé publique [1].

A cet effet, l'objectif principal de cette étude est de comparer les principaux guidelines relatifs à la validation analytique et de révéler les incohérences entre les définitions des critères de validation et les procédures expérimentales proposées par chaque document pour évaluer les critères de validation ainsi que les outils statistiques utilisés pour aider à la prise de décision concernant la validité ou la non-validité d'une procédure analytique.

Cette étude sera limitée aux critères de validation suivants : spécificité, linéarité & intervalle de dosage, fidélité, justesse, exactitude, limite de quantification (LQ) et limite de détection (LD).

Pour chacune de ces caractéristiques de performance nous commençons par une discussion du concept et la définition du critère. Ensuite, nous étudions les plans expérimentaux proposés par chaque document, puis les critères d'acceptation et les tests statistiques utilisés pour l'estimation de chaque critère. Le cas échéant, nous illustrons notre étude sous forme graphique. Pour une synthèse, un tableau est fourni pour chaque critère. À la fin de chaque section, nous recommandons des propositions concernant le plan expérimental à adopter en harmonie avec l'approche du profil d'exactitude

Les documents étudiés sont les suivants :

1. International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use, Topic Q2 (R1): Validation of Analytical Procedures: Text and Methodology June 1995
2. Guidance for Industry Bioanalytical Method Validation U.S. Department of Health and Human Services Food and Drug Administration - Center for Drug Evaluation and Research (CDER) - Center for Veterinary Medicine (CVM) - May 2001.
3. AOAC (Association of Official Analytical Chemists) Guidelines for Single Laboratory Validation of Chemical Methods for Dietary Supplements and Botanicals.
4. IUPAC Technical Report Harmonized guidelines for single laboratory validation of methods of analysis 2002
5. Method validation and quality control procedures for pesticide residues analysis in food and feed document No. SANCO/10684/2009
6. EURACHEM Guide The Fitness for Purpose of Analytical Methods A Laboratory Guide to Method Validation and Related Topics December 1998

## **2. Spécificité & sélectivité**

### **a. Concept & Définition**

Le premier critère pour évaluer une méthode analytique consiste à vérifier sa capacité à établir de manière univoque l'existence de la substance à analyser (analyte) en présence d'autres composants potentiellement présents [2]. Cette capacité de distinguer l'analyte des autres composants a été désignée pendant plusieurs années par le nom de "sélectivité" ou, selon les guidelines par "spécificité" d'une méthode [1]. Seul le document ICH parle de la spécificité alors que les autres documents parlent bien de la sélectivité et Eurachem utilise les deux termes pour définir la même chose.

### **b. Plan d'expérience**

Aucun document ne propose un plan expérimental complet en terme de nombre exact d'échantillons. La FDA spécifie qu'il faut analyser des matrices biologiques (blanc) provenant de 6 sources au moins.

L'Eurachem exige pour le critère de spécificité :

- une analyse des échantillons et de la substance de référence en comparaison avec une méthode indépendante ;
- une analyse des échantillons contenant différents interférents potentiels en présence de la substance à analyser ;

sans aucune autre précision concernant le nombre des niveaux de concentration de mesures ou de séries.

Dans les autres documents, la seule indication mentionnée est l'analyse des matrices et des matrices enrichies par la substance à analyser (cf. Tableau 1)

**c. Tests statistiques & critères d'acceptation :**

- IUPAC exige le calcul d'un indice appelé Indice de sélectivité

$$\text{Indice de sélectivité} = b_{an}/b_{int}$$

Avec :

$b_{an}$  = pente de la courbe d'étalonnage

$b_{int}$  = pente de la réponse produite indépendamment par des substances susceptibles d'influencer les résultats

- AOAC fixe une limite pour la résolution du pic de réponse de tous les composants qui doit être :  $R_s \geq 2$
- La sélectivité doit être assurée à la LQI (limite de quantification inférieure) pour le document de la FDA.
- La réponse du blanc (ou de la matrice) doit être inférieure à 30% de LQ (limite de quantification) selon le document SANCO.

**d. Recommandations**

Une distinction claire doit être faite entre les deux termes "sélectivité" et "spécificité" :

"Une réaction spécifique est celle qui se produit seulement avec la substance à analyser, tandis qu'une réaction sélective est celle qui peut se produire avec d'autres substances, mais expose d'une manière préférentielle la substance à analyser. Peu de réactions sont spécifiques, mais plusieurs montrent une sélectivité". Ceci est compatible avec le concept de la sélectivité qui désigne quelque chose qui peut être graduée tandis que la spécificité est une caractéristique absolue. Pour plusieurs analystes, il est généralement admis que la spécificité est quelque chose d'exceptionnel. En effet, peu de méthodes répondent à un analyte unique [1]. Il paraît donc plus opportun de maintenir le terme «sélectivité» dans le cas des méthodes chromatographiques [2].

### 3. Linéarité & Fonction de réponse

#### a. Concept & Définition

Tous les documents étudiés donnent une définition de la linéarité (ou de fonction de réponse), à l'exception du document IUPAC qui ne donne aucune définition ni de la linéarité ni de la fonction de réponse (voir Tableau 2 & 2').

La première confusion rencontrée dans ces documents est celle existant entre le concept de la Fonction de réponse- ou courbe d'étalonnage - et celui de la Linéarité [1].

Comme défini dans [2] :

**La fonction de réponse** d'une procédure d'analyse traduit, à l'intérieur de l'intervalle de dosage, la relation existant entre la réponse (signal) et la concentration (quantité) en substance à examiner dans l'échantillon. La fonction de réponse monotone la plus simple qui exprime cette relation est appelée «courbe d'étalonnage »

**La linéarité** d'une procédure d'analyse est sa capacité à l'intérieur d'un certain intervalle de dosage d'obtenir des résultats directement proportionnels à la quantité (exemple : concentration) en analyte dans l'échantillon [2].

Le critère de linéarité se réfère à la relation entre la **quantité introduite** et la **quantité calculée** (résultat) d'analyte à partir de la courbe d'étalonnage tandis que la fonction de réponse se réfère à la relation entre la **réponse instrumentale** et la **concentration** [2].

Ceci est résumé dans la Figure 1 où les deux courbes de linéarité et de fonction de réponse sont présentées :



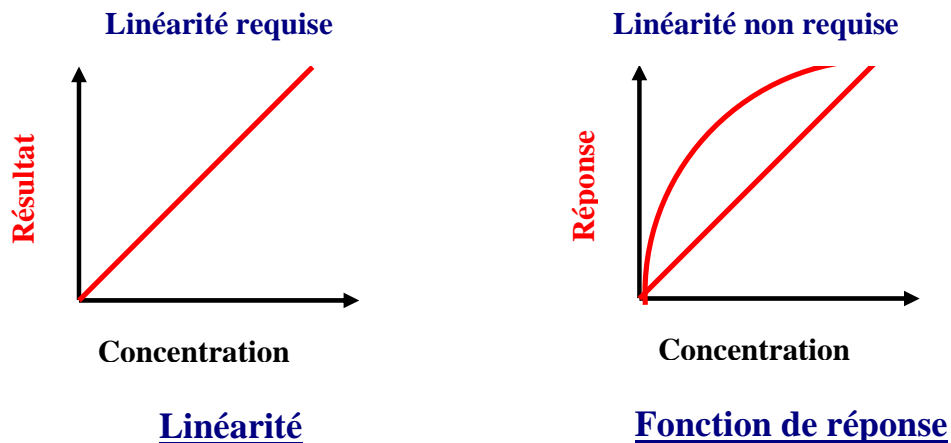


Figure III-1: Fonction de réponse et linéarité

La confusion est remarquée dans les textes de l'ICH et Eurachem. Dans ces documents, la linéarité est définie comme "... la capacité (dans un intervalle donné) d'obtenir des **résultats** qui sont directement proportionnels à la **concentration (la quantité)** de la substance analysée dans un échantillon"[5,8]. Mais plus tard, ces deux documents mentionnent que "la linéarité devrait être évaluée par l'examen visuel d'une courbe des valeurs du signal analytique tracées en fonction de la concentration de la substance à analyser."

Le texte indique clairement que c'est le signal et non plus le résultat qui importe dans la linéarité. Les documents confondent alors, d'une part, la linéarité et la courbe d'étalonnage et, d'autre part, les résultats et le signal. La suite du texte de l'ICH par exemple est auto-explicite [1]: "Si la relation est linéaire, les **résultats** doivent être évalués au moyen des méthodes statistiques appropriées, comme par exemple, la régression des moindres carrées."

Enfin, le fait qu'aucune linéarité ne soit nécessaire entre la quantité et le signal est - paradoxalement - contenu dans la dernière phrase de cette section consacrée à la linéarité [1] : "dans certains cas, pour mettre en évidence la linéarité entre les concentrations (trouvées) et les concentrations des échantillons (introduites), il faut soumettre les données à une transformation mathématique avant de faire l'analyse de régression." [5]

Nous remarquons que cette confusion n'existe pas au niveau **des textes de FDA & SANCO** qui ne contiennent désormais plus le mot "linéarité", mais seulement celui de "la courbe d'étalonnage», sans aucune restriction, en précisant qu' "on doit utiliser le modèle le plus simple qui décrit adéquatement la relation concentration – réponse".

#### **b. Plan d'expérience**

Pour le calcul de linéarité, le plan expérimental est de 5 concentrations dans le ICH, et au moins 6 concentrations pendant 2 à 3 jours dans les documents IUPAC & EURACHEM.

Pour le calcul de la fonction de réponse, le plan expérimental est de 6 à 8 concentrations pendant au moins 2 jours dans les documents FDA & AOAC

#### **c. Tests statistiques & critères d'acceptation :**

Les tests utilisés dans les différents documents pour évaluer la linéarité sont:

- **L'Examen visuel**
- **Le Coefficient de corrélation, ordonnée à l'origine, pente, courbe de régression, ainsi que la somme des carrés des résidus,**
- L'Examen Visuel de la courbe des résidus
- Le Test de manque d'ajustement

Seuls les tests des résidus et l'examen visuel sont décrits pour l'évaluation de la fonction de réponse.

Les critères d'acceptation sont explicitement recommandés dans 3 documents (FDA, AOAC & SANCO) :

**FDA :**

- Au niveau de la concentration équivalente à LQI , la limite est :  $\leq 20\%$  d'écart de la valeur réelle ;
- Les niveaux de concentrations supérieurs à la LQI, la limite est  $15\%$  d'écart ;
- Au moins 4 échantillons sur 6 (incluant le LQI et la concentration maximale) devraient être conformes à ces limites.

**AOAC**

- Un ajustement acceptable produit de façon aléatoire des résidus avec une moyenne 0. Tracer le signal de réponse en fonction de la concentration.
- Une réponse linéaire est souhaitable, mais il n'est pas nécessaire et n'est pas une caractéristique de performance requise.
- Le coefficient de corrélation élevé (par exemple  $> 0,99$ ) ne doit pas être utilisé comme une preuve de qualité de l'ajustement.

**SANCO**

- Les valeurs des résidus doivent être  $\leq \pm 20\%$  de la réponse du standard d'étalonnage si le niveau maximum de résidus (MRL) est dépassé et dans les limites de  $\pm 50\%$ , si le MRL n'est pas dépassé.

**d. Recommandations**

Les critères d'acceptation et les tests utilisés dans ces documents (tels que le R, le manque d'ajustement...) afin d'évaluation la linéarité et de démontrer la qualité d'ajustement d'un modèle sont fournis à titre informatif et ne sont pas, dans la plupart des cas, en accord avec l'objectif de la méthode analytique [1].

Le coefficient de détermination, par exemple, qui est largement utilisé dans la pratique exprime, en réalité, la part de la variabilité des réponses pouvant être expliquées par un modèle de régression. Il est souvent abusivement interprété comme une évaluation de la qualité d'ajustement d'un modèle. A cet effet, il ne sert à rien de s'imposer des

contraintes du type  $R^2 > 0.99$  qui n'est pas un gage de qualité des résultats qui seront obtenus [10].

Cette remarque est aussi confirmée dans les documents AOAC [7] et IUPAC [6]

Pour cette raison, plusieurs auteurs [1, 2, 3, 11] ont introduit l'utilisation du profil d'exactitude basée sur les intervalles de tolérance (ou intervalles de prédiction) pour décider si un modèle de calibration donnera des résultats de qualité ou non. Les modèles doivent être acceptés ou rejetés en fonction de l'exactitude des résultats calculés, indépendamment des propriétés statistiques [1].

Afin de mieux expliquer cette conclusion relative au choix du profil d'exactitude, nous illustrons notre étude par un exemple de dosage de la libération de principe actif (levonorgestrel).

Il s'agit d'un test de libération d'un principe actif, le levonorgestrel, à partir d'une matrice polymérique. La procédure de dosage est effectuée par chromatographie liquide à haute performance avec détection UV/Visible.

La Figure 2 illustre une présentation graphique de la courbe de la fonction de réponse (1) (surface des pics vs concentrations introduites), courbe des résidus (2) et ce en utilisant deux fonctions de réponse : linéaire simple (A) et linéaire simple après transformation racine carrée (B)

D'après la Figure 2, nous pouvons faire les remarques suivantes :

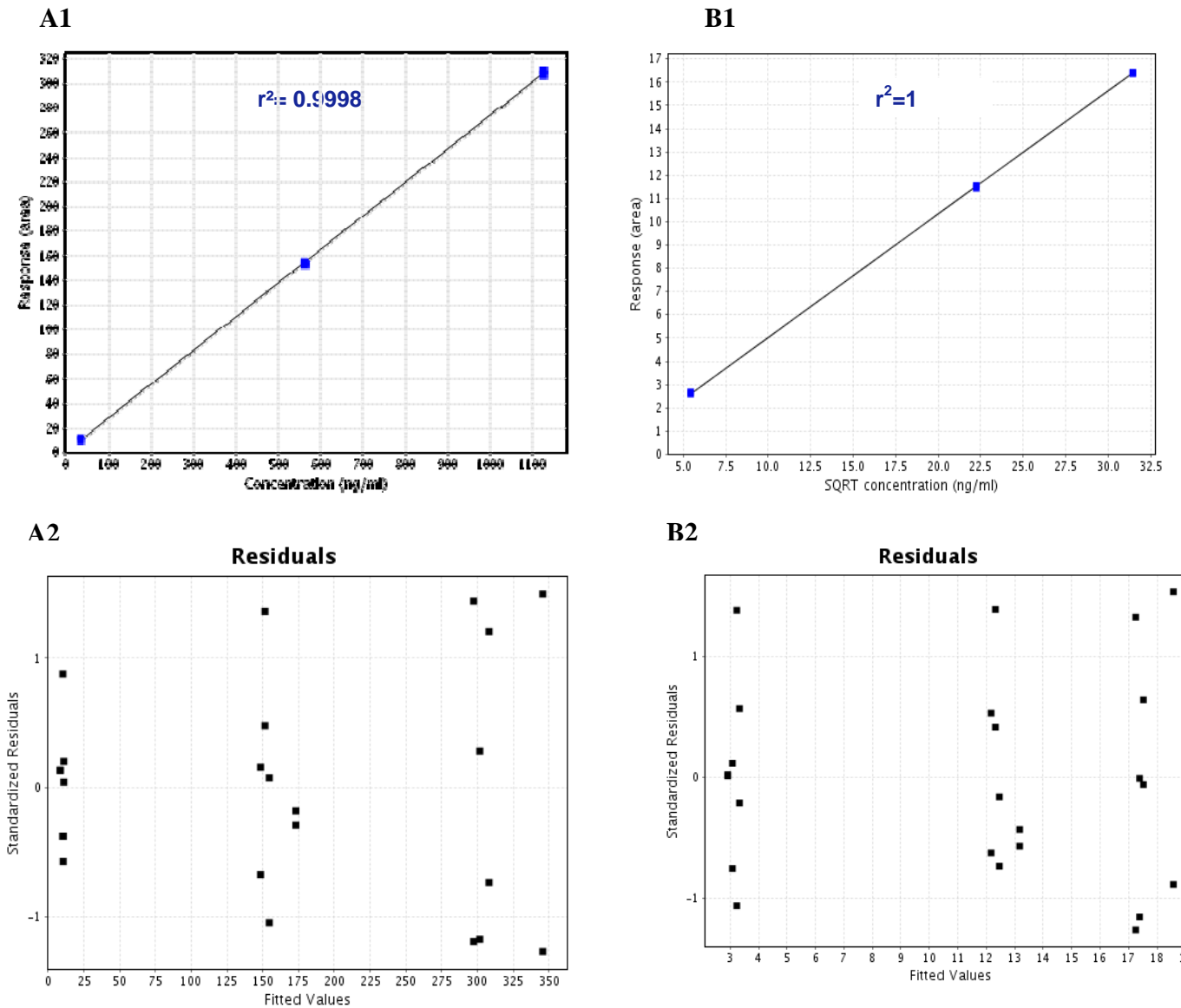
- L'évaluation visuelle des deux présentations graphiques de la courbe de la fonction de réponse ne montre aucune différence entre A1 & B1 puisque les deux courbes présentent des droites linéaires : Donc selon les différentes guidelines, la linéarité est prouvée pour les deux modèles.
- Les deux valeurs de  $R^2$  sont proches de la valeur 1 (égale 1 pour le modèle de racine carré) : encore une fois et selon les différents guidelines, la linéarité est prouvée pour les deux modèles.

- D'après la figure, nous remarquons que la courbe de résidus A2 présente un problème d'homogénéité des variances et donc un problème de linéarité et une nécessité de procéder à une transformation mathématique, ce qui est en contradiction avec les deux premières remarques.

Donc, selon cette remarque seule le modèle linéaire simple avec transformation racine carrée est linéaire.

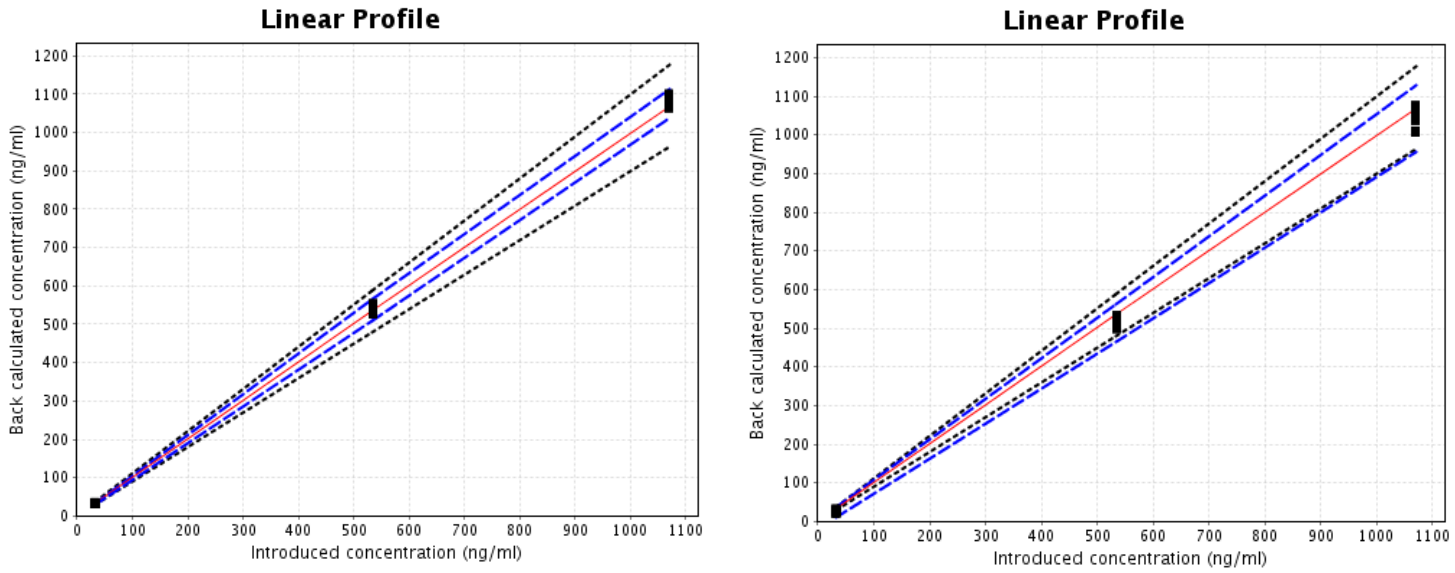
La Figure 3 montre les profils d'exactitude de la linéarité établis avec les mêmes données de la Figure 2 et basés sur les limites d'acceptation de la méthode et l'intervalle de tolérance (pour plus de détails concernant la construction du profil d'exactitude, se référer à la partie exactitude de ce même chapitre)

D'après cette figure, et contrairement à la Figure 2, seule la fonction linéaire simple est linéaire bien que le modèle racine carré présente des résultats conformes en termes d'évaluation visuelle de la courbe de la fonction de réponse, de la courbe des résidus et de la valeur de  $R^2$ , tests exigés par la plupart des documents étudiés afin de démontrer la linéarité.



**Figure II-2 : Présentation graphique de la fonction de réponse 1 et courbe des résidus 2 pour deux modèles de calibration : linéaire simple (A) et racine carrée (B)**

En conclusion, nous pouvons dire que seule l'approche basée sur l'intervalle de tolérance et les limites d'acceptation peut nous renseigner sur la linéarité de la méthode et la fonction de réponse à choisir pour mener une validation analytique. Cette dernière conclusion est très importante car le choix de la fonction de réponse est très critique pour une validation analytique et pour l'utilisation en routine de la méthode après validation.



**Figure III-3 : Ligne continue (rouge) représente la droite ( $Y=X$ ) sur laquelle les points doivent être localisés pour la validation de la méthode ; ligne discontinue (bleu): intervalle de tolérance  $\beta=95\%$ ; ligne en pointillés (noir) : limites d'acceptation à 5 % ; Points : concentrations calculées**

#### **4. Fidélité**

##### **a. Concept & Définition**

L'estimation de la fidélité est exigée par tous les documents (voir Tableau 3). Contrairement à la linéarité et à la fonction de réponse, nous pouvons trouver des définitions homogènes de la fidélité dans presque tous les documents [1].

Par exemple, la définition citée dans l'ICH est la suivante : la fidélité : "... correspond au degré d'accord (degré de dispersion) entre les résultats des mesures obtenues par l'analyse individuelle de plusieurs prélèvements d'un même échantillon homogène, prélevés dans des conditions prescrites" [5, 12]. Cette définition est en accord avec les définitions données par la FDA, AOAC, USP et IUPAC.

Néanmoins, trois types de fidélité sont généralement estimés, à savoir :

- 1) La répétabilité qui est une expression de la fidélité de la méthode lorsque celle-ci est reprise dans les mêmes conditions de réalisation, après un court intervalle de temps. La répétabilité est aussi désignée par fidélité intra-analyse (intra série ou intra run)
- (2) La fidélité intermédiaire qui correspond aux variations survenant dans un même laboratoire : analyses effectuées dans des jours différents, par des personnes différentes, au moyen d'appareils différents, etc.
- (3) La reproductibilité correspond à la concordance entre laboratoires (travaux de collaboration visant généralement la normalisation de la méthode).

Seuls les documents EURACHEM et SANCO recommandent l'évaluation de ces trois types de fidélité alors que la FDA & IUPAC exigent les types 1 et 2, alors que le AOAC exige les types 2 & 3.

Pour le document ICH (et USP), la reproductibilité est évaluée lorsque la méthode à valider doit être normalisée.



Le document de FDA, distingue aussi entre la fidélité intra-lot (appelé aussi répétabilité), qui évalue la fidélité pendant un seul jeu d'analyse et "la fidélité inter-lot (appelé aussi répétabilité), qui mesure la fidélité dans le temps et peut impliquer des différents analystes, équipement, réactifs et laboratoires" [1,4].

Comme nous pouvons le constater dans ce document, le même terme, à savoir « répétabilité » est utilisé deux fois pour les deux composants de variabilité, pouvant ainsi prêter à confusion pour l'analyste. De plus, ce document considère au même titre la variabilité dans un seul laboratoire ou dans différents laboratoires [1,12].

Les conditions de répétabilité impliquent la ré-exécution de l'ensemble de la procédure de la sélection à la préparation de la prise d'essai de l'échantillon dans le laboratoire et non seulement des répétitions instrumentales sur un seul échantillon préparé. Ce dernier s'appelle « fidélité instrumentale » qui ne comprend pas la répétition de toute la procédure d'analyse [1].

D'une manière générale, la validation d'une procédure analytique est réalisée dans un seul laboratoire puisqu'il doit être démontré que la procédure analytique est appropriée à son objectif attendu. L'évaluation de l'adéquation de la méthode d'un laboratoire à un autre laboratoire est d'habitude réalisée dans l'objectif de standardiser la procédure ou d'évaluer la performance de plusieurs laboratoires dans un " test de compétence", appelé aussi " test d'aptitude " et il est géré par des documents et des règles spécifiques [1].

Enfin, tous les documents exigent, explicitement ou implicitement, l'évaluation des erreurs produite par chaque type de fidélité. Cependant, aucun des documents étudiés (sauf l'IUPAC) n'exige l'évaluation de l'ensemble des erreurs aléatoires générées par le calcul de toutes les catégories de fidélité (répétabilité, fidélité intermédiaire).

**b. Plan d'expérience**

En règle générale, il est recommandé d'estimer la fidélité à 3 niveaux de concentrations différents (FDA, ICH). L'IUPAC exige deux niveaux de concentrations.

Seul le nombre total de répétitions est donné, sans aucune précision relative au plan expérimental en termes de nombre de jours [12] (3 à 10 répétitions)

Comme nous pouvons le constater dans les documents réglementaires qui font la différence entre répétabilité et fidélité intermédiaire, le concept de série s'impose. Ces séries sont composées au moins de dates différentes avec des opérateurs différents et/ou des équipements différents. Une série est une période pendant laquelle les analyses sont réalisées dans les conditions de répétabilité qui restent constantes. Il serait logique de choisir différents facteurs qui composeront les séries de telle façon à imiter les conditions qu'on rencontrera pendant l'utilisation en routine de la procédure analytique [1].

Il est évident que la procédure analytique ne sera pas utilisée qu'un seul jour. Ainsi, le fait d'inclure le facteur jour est obligatoire. Pendant son utilisation en routine, la procédure analytique ne sera pas utilisée par un seul opérateur, et/ou avec un seul équipement, d'où l'importance d'inclure ces facteurs de variabilité dans l'objectif et d'avoir une estimation représentative de la variabilité de la procédure analytique [1].

Quand la sélection des facteurs appropriés est faite, un plan expérimental peut être choisi pour optimiser le nombre de séries afin de représenter les effets principaux de ces facteurs avec un temps d'analyse effectif. Par exemple, si les facteurs choisis sont les jours, les opérateurs et les équipements, chaque facteur a deux niveaux, donc le plan factoriel fractionnaire permet de réaliser quatre séries dans seulement 2 jours. Le plan est montré dans le tableau 4 [1].

Série 1	Série 2	Série 3	Série 4
Jour 1	Jour 1	Jour 2	Jour 2
Opérateur 1	Opérateur 2	Opérateur 1	Opérateur 2
Equipement 2	Equipement 1	Equipement 1	Equipement 2

**Table III-4 : Plan d'expérience de 4 séries faites en prenant comme source de variabilité jour, opérateur et équipement**

**c. Tests statistiques & critères d'acceptation :**

Comme statué dans tous les documents, la fidélité est exprimée par l'écart-type ( $s$ ), la variance ( $s^2$ ) ou le coefficient de variation (CV). Elle mesure l'erreur aléatoire liée à une procédure analytique, c'est-à-dire la dispersion des résultats autour de leur valeur moyenne. L'estimation de la fidélité est indépendante de la valeur vraie ou de la moyenne de la justesse estimée.

Pour évaluer correctement les deux composantes de variabilité d'une procédure analytique pendant la phase de validation, il est recommandé [1] d'effectuer une analyse de variance (ANOVA) par niveau de concentration examiné. Tant que le plan expérimental est équilibré, c'est-à-dire le même nombre de répétitions par série pour un niveau de concentration donné, l'estimation de ces composantes par la méthode des moindres carrés peut être utilisée. Cependant, quand cette condition n'est pas respectée, les estimations de maximum de vraisemblance devraient être préférées [1,14].

De le tableau ANOVA, la répétabilité ou la fidélité intra-série et la fidélité inter-série sont obtenus comme suit :

$$MSM_j = \frac{1}{P-1} \sum_{i=1}^p n(X_{ij,calc} - \hat{\mu}_j)^2$$

MSM<sub>j</sub> : le carré moyen du modèle (mean square model) du niveau de concentration j

Avec  $X_{ij,calc}$  est la moyenne des résultats du niveau de concentration j de la série i; p est le nombre de séries; n est le nombre de répétitions par série;

$$\hat{\mu}_j = \frac{1}{pn} \sum_{i=1}^p \sum_{k=1}^n X_{ijk,calc}$$

$\hat{\mu}_j$  est la moyenne des concentrations calculées du niveau de concentration j

$$MSE_j = \frac{1}{pn - p} \sum_{i=1}^p \sum_{k=1}^n (x_{ijk,calc} - X_{ij,calc})^2$$

Avec  $MSE_j$  : Le carré moyen résiduel (mean square error) du niveau de concentration j

Si  $MSE_j < MSM_j$  alors:

$$\hat{\sigma}_{W.j}^2 = MSE_j$$

$$\hat{\sigma}_{B.j}^2 = \frac{MSM_j - MSE_j}{n}$$

Si non :

$$\hat{\sigma}_{W.j}^2 = \frac{1}{pn - p} \sum_{i=1}^p \sum_{k=1}^n (X_{ijk,calc} - X_{ijk,calc})^2$$

$$\hat{\sigma}_{B.j}^2 = 0$$

La fidélité intermédiaire est calculée comme suit:

$$\hat{\sigma}_{IP.j}^2 = \hat{\sigma}_{W.j}^2 + \hat{\sigma}_{B.j}^2$$

Avec  $\hat{\sigma}_{W.j}^2$  la variance intra série ou la répétabilité et  $\hat{\sigma}_{B.j}^2$  est la variance inter-série.

Après avoir évalué les composantes de la variance, un paramètre intéressant à observer est le rapport,  $R_j$  avec :

$$R_j = \frac{\hat{\sigma}_{B.j}^2}{\hat{\sigma}_{W.j}^2}$$

En effet, ce paramètre montre l'importance de la variance inter-série en comparaison avec la variance de répétabilité. Les valeurs élevées de  $R_j$ , par exemple plus grand que 4, pourraient conclure à un problème de variabilité de la procédure analytique dont les résultats peuvent varier d'une série à une autre, menant ainsi à revoir l'optimisation de la méthode, ou soit à interroger le nombre de série utilisées pendant le processus de validation pour obtenir une estimation fiable de la variance inter-série  $\hat{\sigma}_{B.j}^2$  [1].

La fidélité est généralement exprimée comme le pourcentage de coefficient de variation (CV). La formule classique est :

$$CV(\%) = 100 \times \frac{\sqrt{\hat{\sigma}^2}}{\hat{x}}$$

Avec  $\hat{\sigma}^2$  est l'estimation de variance et  $\hat{x}$  est la valeur de la moyenne estimée

Pour les critères d'acceptation, encore une fois, seuls les documents FDA & SANCO définissent des critères d'acceptation à savoir :

- Pour FDA :
  - A LQI:CV ≤20%
  - Autre que LQI ≤15%
- Pour Sanco
  - Critère d'acceptation ≤ 20%

#### **d. Recommandations**

- ↯ Notons que dans la pratique de plusieurs laboratoires et selon les recommandations du guide de validation SFSTP proposé en 1992 et qui fut le premier document proposant une méthodologie expérimentale pratique et complète ainsi qu'une procédure de traitement statistique des critères de validation analytique, l'évaluation de la fidélité n'est réalisée que sur le point 100%, chose qui est en contradiction avec tous les guidelines étudiés et avec l'objectif même de la validation analytique qui est de démontrer la fidélité et l'exactitude sur l'ensemble du domaine de concentration sur lequel nous voulons démontrer la validité de notre méthode.
- ↯ Afin de réaliser les répétitions suffisantes pour contrôler les erreurs aléatoires liés à la méthode étudiée, nous proposons d'utiliser les plans expérimentaux proposés par Hubert et Col [2, 3], qui proposent des plans adaptés aux différents types de méthodes analytiques utilisées dans l'industrie.
- ↯ Les critères d'acceptation de la fidélité doivent être fixés selon les objectifs de la méthode et doivent être associés aux erreurs systémiques liés à la justesse, afin de calculer l'erreur totale calculée par l'exactitude.

## 5. Justesse

### a. Concept & Définition

L'estimation de la justesse est requise par tous les guidelines (Cf. Tableau 4). Cependant, dans la majorité des documents, nous constatons une confusion entre le terme «exactitude» et le terme «justesse».

Comme nous pouvons le voir à partir de la définition suivante de la justesse trouvé dans les documents d'Eurachem « the Fitness for Purpose of Analytical Methods» [8] et IUPAC [6] (définition tirée de la norme ISO [14]), la justesse est: «Le degré de concordance entre la valeur moyenne obtenue à partir d'une large série de résultats d'essai et une valeur de référence acceptée. La mesure de la justesse est habituellement exprimée en termes de biais. La justesse a été appelée "exactitude de la moyenne". Cet usage n'est pas recommandé. "

Toutefois, lorsque nous recherchons la justesse dans les autres documents réglementaires, nous trouvons qu'il y a une incohérence dans la définition de la justesse à travers les différents guidelines.

Si nous lisons soigneusement les deux documents ICHQ2R1 [5] et FDA Bioanalytical Method validation [4], nous réalisons que ce concept n'est pas défini en soi. Dans l'ICH Q2R1 - partie 1, la justesse est citée comme suit dans la définition de l'exactitude: "L'exactitude correspond au degré de concordance entre la valeur de la méthode obtenue et la valeur de référence ou la valeur considérée comme vraie par convention. L'exactitude est aussi désignée par le mot « justesse »" Tandis que dans le document FDA nous trouvons une référence à ce concept au niveau du glossaire, "le degré de concordance entre la valeur obtenue et la vraie valeur connue dans des conditions prescrites est désignée aussi par « justesse »"[1].

Une confusion existe, donc, entre les deux concepts de « justesse » (ou exactitude de la moyenne) et « exactitude » (des résultats). Le document ISO précise également que

cette utilisation de la terminologie d'exactitude de la moyenne doit être évitée et remplacée par celle de justesse.

Lorsque nous comparons la définition de justesse dans les documents ICH & FDA et la définition de EURACHEM & IUPAC (et ISO), nous constatons que la principale différence est que dans les premiers documents, nous parlons d'écart entre la valeur réelle et la valeur trouvée ou la valeur calculée alors que dans les derniers documents, on se focalise sur la distance entre la valeur moyenne et la valeur réelle. Il est essentiel de distinguer la différence entre un résultat et une valeur moyenne [1].

Les résultats d'une procédure analytique sont les objectifs finaux. Lors de l'examen d'un échantillon de contrôle de qualité, le résultat impacte la décision de libérer un lot. Lorsque des échantillons inconnus sont analysés, les résultats donnent des informations sur l'effet thérapeutique d'un médicament ou de l'état pathologique ou physiologique d'un patient et ainsi de suite [1].

La valeur moyenne ne donne qu'une position centrale de la distribution des résultats de la même teneur réelle, pas la position de chaque résultat individuel. Par extension, le biais, le biais relatif ou le recouvrement va localiser la distribution des résultats produits par une méthode analytique par rapport à la vraie valeur [1].

Cette incohérence de définition n'est pas seulement trouvée d'un document à un autre, mais aussi dans des sections différentes du même document, en particulier dans le document ICH Q2R1.

Pour conclure, la justesse est liée à des erreurs systématiques des méthodes d'analyse. La justesse se réfère ainsi à une caractéristique ou une qualité de la procédure analytique et non à un résultat généré par cette procédure. Par contre l'exactitude, comme nous allons le voir par la suite, se réfère à la qualité des résultats.

**b. Plan d'expérience**

La majorité des guidelines (à l'exception de l'IUPAC & SANCO) définissent un plan expérimental pour évaluer l'exactitude :

- L'ICH recommande un minimum de 9 analyses à partir de 3 concentrations (n= 3) couvrant l'intervalle de linéarité càd (1 jour × 3 concentrations × 3 répétitions)
- Le FDA exige 3 niveaux de concentrations couvrant l'écart d'utilisation c-à-d la concentration la plus basse, intermédiaire et maximale de l'intervalle de dosage × 5 répétitions càd (1 j × 3 concentrations × 5 répétitions)
- AOAC recommande aussi 9 à 10 niveaux de concentration couvrant l'intervalle de dosage à réaliser pendant 3 jours

L'estimation de la justesse est généralement combinée avec les expériences de fidélité. Par conséquent, les remarques sur le plan expérimental, déjà faite sur la fidélité, s'appliquent également ici.

**c. Tests statistiques & critères d'acceptation :**

Trois méthodologies sont proposées par les différents documents afin d'évaluer la justesse des méthodes analytiques :

1. elle peut être déterminée en utilisant **une seule substance de référence certifiée** et en comparant la **valeur mesurée avec la vraie valeur** (cette approche est proposée par tous les documents) ;
2. Une deuxième approche compare **le résultat de la méthode proposée** avec le **résultat d'une méthode de référence** dont la justesse et la fidélité sont connues. (Cette approche est proposée par l'ICH, USP, AOAC et IUPAC)
3. Ces deux approches sont impraticables si la substance de référence ou la méthode de référence ne sont pas disponibles. Dans des telles situations, **une étude de recouvrement** est réalisée dans laquelle la matrice ou le placebo est dopé par



différentes quantités de l'analyte couvrant l'intervalle de la linéarité suivi par une quantification. Le résultat est alors exprimé en pourcentage de recouvrement [12]. (Cette approche est proposée par l'AOAC et IUPAC).

Notons que l'approche de comparaison des méthodes est recommandée par rapport à l'utilisation des matrices dopées par des substances de références certifiées. Ceci est en partie dû au nombre restreint de niveaux de concentration qu'elles couvrent (normalement seulement 2-3), mais surtout de la manière dont les échantillons sont préparés. [13]

Les paramètres statistiques nécessaires pour le calcul de la justesse sont cités dans tous les guidelines.

Les tests proposés sont :

- Le pourcentage de recouvrement
- Et la différence entre la valeur moyenne et la valeur considérée comme vraie avec l'intervalle de confiance correspondant (moyenne - valeur vraie)

En effet, la distance entre la valeur moyenne d'une série de mesures ( $\bar{x}_i$ ) et une valeur de référence  $\mu_T$  est appelé biais (biais, biais relatif ou recouvrement) :

$$\text{Biais} = \bar{x}_i - \mu_T$$

$$\text{Biais relatif (\%)} = 100 \times \left( \frac{\bar{x}_i - \mu_T}{\mu_T} \right)$$

$$\text{Recouvrement} = 100 \times \frac{\bar{x}_i}{\mu_T} = 100 - \text{Biais relatif (\%)}$$

Il est possible de calculer la valeur moyenne de ces résultats prédits ( $\bar{x}$ ) et, par conséquent estimer le biais, le biais relatif ou le recouvrement.

Lors de l'évaluation de l'acceptabilité du biais, du biais relatif ou du recouvrement, la méthodologie la plus utilisée est l'application du test de student en vérifiant les hypothèses suivantes [1]:

$$H_0 : \bar{x}_i - \mu_T = 0$$

$$H_1 : \bar{x}_i - \mu_T \neq 0$$

pour lequel un niveau de signification  $\alpha$  est fixé, généralement à 0,05. Cela signifie qu'il y a un risque, de 5 fois sur 100, de rejeter à tort l'hypothèse nulle  $H_0$ . Autrement dit, nous acceptons un risque, de 5 fois sur 100, de considérer à tort un biais différent de 0

En général, lorsque le quantile calculé de student est plus élevé que le quantile théorique correspondant, ou de manière équivalente lorsque le p-valeur est inférieur à  $\alpha$ , l'hypothèse nulle est rejetée. Par conséquent, il existe un degré élevé de confiance que le biais soit différent de 0 au niveau de signification  $\alpha$  fixé par l'analyste.

Une autre façon d'interpréter ce test est de regarder si le 0% de biais relatif ou 100% de recouvrement est incluse dans l'intervalle de confiance  $1-\alpha$ , respectivement, du biais relatif ou du recouvrement. Si ces valeurs sont en dehors de leur intervalle de confiance correspondant, alors l'hypothèse nulle est rejetée.

Toutefois, la seule conclusion qui peut être faite lorsque l'hypothèse nulle n'est pas rejetée c'est que ce n'est pas que le biais, le biais relatif ou le recouvrement qui est égal à 0, 0% ou 100%, mais que le test n'a pas pu démontrer que le biais, le biais relatif ou le recouvrement est différent de 0 ou 100.

Comme démontré clairement dans de nombreuses publications [1, 15-18], le risque  $\beta$ , qui est la probabilité d'accepter, à tort, l'hypothèse nulle, n'est pas fixé par l'utilisateur dans cette situation. En outre, cette approche peut conclure que le biais est significativement différent de 0, alors qu'il pourrait être analytiquement acceptable [1, 15-18].

Il est également toujours considéré que le biais n'est pas différent de 0 lorsque la variabilité de la procédure est relativement élevée. En fait, le test t de Student utilisé de cette façon

est un test de différence qui répond à la question: "Est-ce le biais de ma méthode analytique est différent de 0?". Toutefois, la question à laquelle l'analyste veut répondre au cours de l'étape de validation analytique est la suivante: «Est-ce que le biais de ma méthode d'analyse est acceptable?", le test nécessaire pour répondre à cette dernière question est un test d'équivalence ou un test d'intervalle d'hypothèse [1, 15-19]. Dans ces types de test, l'analyste doit choisir une limite d'acceptation de biais, biais relatif ou recouvrement, qui sont des limites dans lesquelles si le vrai biais, le biais relatif ou le recouvrement de la méthode d'analyse sont inclus, la justesse de cette procédure est acceptable. Différents auteurs ont recommandé l'utilisation de ce type de tests pour évaluer l'acceptabilité d'un biais [15,18].

En effet, une procédure tout à fait non biaisée est une utopie. En outre, le biais obtenu pendant l'expérience de validation n'est qu'une estimation du vrai biais inconnu de la procédure analytique. Néanmoins, ce dernier test de l'intervalle d'hypothèse ne répond pas à la vraie question analytique : le but même de la validation est de valider les résultats que la méthode va produire, et non pas la méthode elle-même... [1].

Encore une fois, pour les critères d'acceptation, seuls les documents de FDA & SANCO donnent des limites :

- Pour le FDA : la valeur moyenne devrait être dans un intervalle de 15 % de la valeur réelle sauf le niveau LQI, où il ne devrait pas dévier par plus de 20 %.
- Pour le document Sanco : La valeur moyenne de recouvrement doit être entre 70 et 120%

**d. Recommandations**

Nous recommandons de faire la distinction entre la justesse qui renseigne sur les erreurs systématiques et l'exactitude qui renseigne sur l'erreur totale.

A cet effet, nous recommandons que le critère d'acceptation de la justesse soit inclus dans le critère d'acceptation global qui englobe les 2 types d'erreurs systématique (justesse) et aléatoire (fidélité) comme proposé par l'approche du profil d'exactitude [2,3].

## 6. Exactitude

### a. Concept & Définition

L'estimation de l'erreur totale n'est pas définie d'une manière spécifique dans les documents étudiés à l'exception de l'EURACHEM et de la première partie du document ICH Q2R1 (Cf. Tableau 5).

Comme nous l'avons vu dans le document ICH Q2R1 partie 1 [5], l'exactitude est définie comme le : "...degré de concordance entre la valeur obtenue et la valeur de référence ou la valeur considérée comme vraie."

Cette définition correspond à celle du document EURACHEM [8], qui indique que l'exactitude est: «le degré de concordance entre le résultat d'analyse et la valeur acceptée comme référence." Par ailleurs, dans cette définition, une note est ajoutée en précisant que l'exactitude est **la combinaison des erreurs aléatoires et des erreurs systématiques (ou le biais).**

Ce degré de concordance observé est basé sur la somme des erreurs systématiques et aléatoires c. à. d. l'erreur totale liée à un résultat. Par conséquent, l'erreur de l'essai est l'expression de la somme de la justesse (ou biais) et de la fidélité (ou écart-type), c'est à dire l'erreur totale.

Cependant, quand nous regardons la section correspondante à l'exactitude dans la deuxième partie du document ICH Q2R1 : "Recommandations concernant les données relatives à l'exactitude" : «Pour exprimer l'exactitude, il est recommandé de déclarer le pourcentage de recouvrement de la substance dont une quantité connue a été ajoutée à l'échantillon, ou la différence entre la valeur moyenne et la valeur considérée comme vraie avec les intervalles de confiance correspondantes." Comme discuté auparavant, Il ne s'agit plus de l'exactitude, mais plutôt de la définition de la justesse qui est la valeur moyenne de

plusieurs résultats (par opposition à un seul résultat pour l'exactitude) qui est comparée à la vraie valeur.

Comme déjà constaté, cette section se réfère à des erreurs systématiques alors que l'exactitude telle que définie par l'ICH Q2R1 (première partie) et l'Eurachem correspond à l'évaluation de l'erreur totale de mesure.

Dans le document de la FDA Bioanalytical Method Validation [4], l'exactitude est définie comme «...le degré d'accord de la moyenne des résultats obtenus par la méthode et la vraie valeur (concentration) de l'analyte. (...). L'écart entre la moyenne de la valeur réelle sert à mesurer l'exactitude. »

Encore une fois, cette définition correspond à celle de la justesse de la méthode d'analyse.

**b. Tests statistiques et calculs :**

Nous rappelons que chaque mesure  $X$  a trois composantes: la vraie valeur de l'échantillon  $\mu_T$ , le biais de la méthode (estimé par la moyenne de plusieurs résultats) et la fidélité (estimée par l'écart-type ou, dans la plupart des cas, la fidélité intermédiaire). De manière équivalente, la différence entre une observation  $X$  et la valeur réelle est égale à la somme des erreurs systématiques et aléatoires, c'est à dire d'erreur totale ou erreur de mesure [1]:

$$X = \mu_T + \text{biais} + \text{fidélité}$$

$$\Leftrightarrow$$

$$X - \mu_T = \text{biais} + \text{fidélité}$$

$$\Leftrightarrow$$

$$X - \mu_T = \text{erreur total}$$

$$\Leftrightarrow$$

$$X - \mu_T = \text{erreur de mesure}$$

$$\Leftrightarrow$$

$$X - \mu_T = \text{Exactitude}$$

A cet effet, l'erreur totale (ou exactitude), se réfère à la distance entre un résultat et la vraie valeur de ce résultat.

**c. Recommandations**

Pour la plupart des utilisations, il n'est pas question de savoir si un écart par rapport à la valeur réelle est dû à une erreur aléatoire (manque de fidélité) ou à une erreur systématique (manque de justesse), tant que la quantité d'erreur totale reste acceptable. Ainsi, l'erreur totale d'analyse ou l'exactitude en fonction de l'erreur aléatoire et systématique est une notion indispensable. En outre, tous les analystes veulent faire en sorte que la quantité d'erreurs totales de la méthode n'aura aucune incidence sur l'interprétation du résultat de l'analyse comme sur la décision ultérieure [1-3].

Une décision fondée sur une évaluation du critère de justesse (erreur systématique) seul ou du critère de fidélité (erreur aléatoire) seul ne peut pas atteindre cet objectif, seule l'évaluation de l'exactitude des résultats qui prend en compte la notion d'erreur totale, donne des garanties à la fois aux laboratoires et aux autorités réglementaires sur la capacité de la méthode d'atteindre son objectif.

L'importance d'évaluer l'erreur totale est de nos jours très importante. Cependant et malheureusement, elle n'est pas encore incluse dans les guidelines de la validation analytique. Une des raisons en est que les concepts statistiques pour le calcul de l'erreur totale sont complexes et que, actuellement, aucun concept généralement admis pour le traitement statistique de l'erreur totale n'existe [1].

Récemment, une nouvelle stratégie de validation basée sur la notion de l'erreur totale et le profil d'exactitude a été introduite avec pour objectif, d'une part, harmoniser les démarches à suivre pour valider les procédures analytiques et d'autre part, offrir un outil pratique de décision basé sur le profil d'exactitude et la notion d'erreur totale [1, 20-26].

Cette stratégie de validation permet d'associer les deux éléments fondamentaux de la validation que sont la justesse et la fidélité au résultat final d'une mesure, et par

conséquent de tenir compte de l'erreur totale de mesure (erreur systématique + erreur aléatoire) [14].

Le principe de cette stratégie de validation peut être traduit par l'équation suivante, qui stipule que la différence entre une mesure ( $x$ ) et sa vraie valeur ( $\mu$ ) doit être inférieure à la limite d'acceptation ( $\lambda$ ) définie a priori.

$$-\lambda < x - \mu < \lambda \leftrightarrow |x - \mu| < \lambda$$

La notion de limite d'acceptation introduit donc un premier critère permettant à l'analyste de prendre des décisions basées sur l'objectif de la méthode analytique. Communément, la limite d'acceptation est de 1% ou 2 % pour le dosage de principes actifs dans une matière première, de 5 % pour les formes pharmaceutiques et de 15 % pour les analyses dans les matrices biologiques ou environnementales. Pour la détermination des impuretés, une limite d'acceptation minimale de 10 % est communément admise.

Le profil d'exactitude est construit à partir d'intervalle de tolérance d'espérance  $\beta$  de mesures attendues à chaque niveau de concentration.

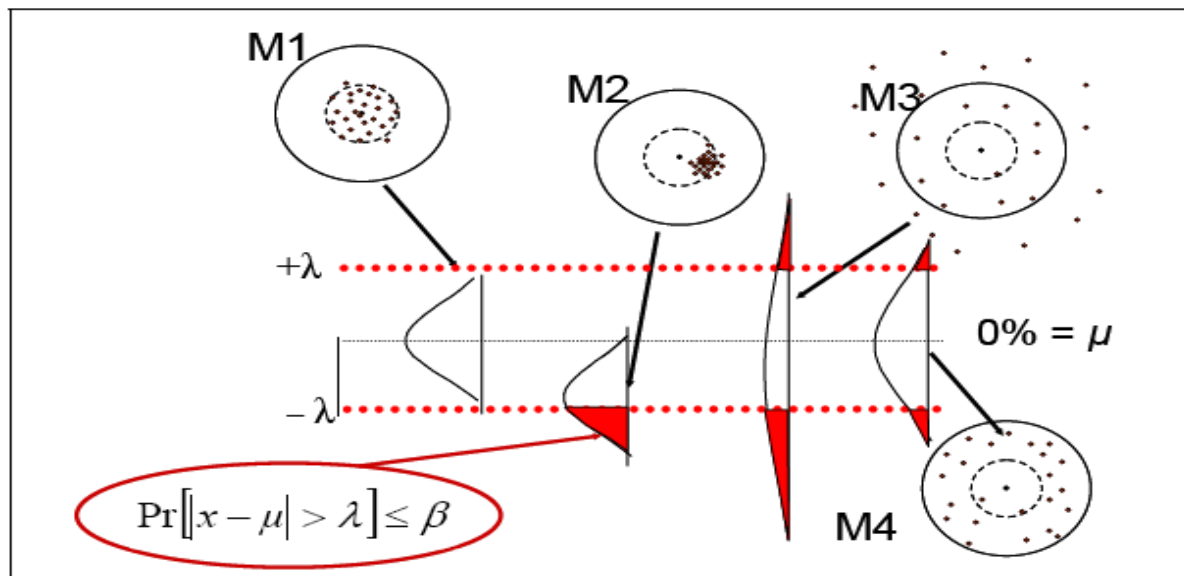
Une autre notion importante définie est celle de « bonne procédure analytique » avec un risque connu qui peut se traduire par la relation suivante [2,3] :

$$\Pr[|x - \mu| < \lambda] \geq \beta$$

avec  $\beta$  la proportion de mesures dans les limites d'acceptation, et  $\lambda$  la grandeur définissant les limites d'acceptation fixées *a priori* en fonction des contraintes du secteur d'activité.

Le risque associé d'une procédure s'évalue par la proportion attendue de mesures en dehors des limites d'acceptation. Le risque associé dépend des estimés du biais et de la fidélité de la procédure analytique obtenus en phase de validation comme le montre la Figure 4.





**Figure III-4 – Profil d'exactitude comme outil de décision pour quatre méthodes M1 (fidèle et juste), M2 (fidèle mais non juste), M3 (non fidèle et non juste) et M4 (non fidèle mais juste)**

Ainsi, le profil d'exactitude est un outil de décision fondé sur le risque associé à la méthode. La notion de risque est liée à la garantie concernant la future analyse des échantillons inconnus tout en appliquant la méthode validée. Aussi, le profil d'exactitude peut-il servir à accepter ou rejeter une méthode analytique suivant l'usage attendu [2,3]. Par ailleurs, le profil peut être utilisé comme outil de diagnostic. Il peut par exemple être utilisé pour sélectionner le modèle de régression le plus approprié à la calibration et pour déterminer les limites de quantification supérieure et inférieure et à sélectionner ainsi l'intervalle de dosage.

La stratégie couramment utilisée pour la validation d'une procédure analytique est basée sur les critères d'acceptation qui considèrent de manière dissociée le biais et la variance.

Cependant, la nouvelle stratégie est basée sur l'intervalle de tolérance d'espérance  $\beta$  de l'erreur totale d'une mesure (en anglais «  $\beta$ - expectation tolerance intervals for the total error measurement »), erreur qui prend en compte la justesse (biais) et la fidélité intermédiaire (écart-type) [2, 3, 20, 22].

Cette stratégie présente la possibilité de contrôler le risque d'acceptation d'une méthode non adéquate et ce, au moyen des profils d'exactitude. Un autre avantage présenté par cette stratégie est de fournir la garantie que les résultats des mesures qui seront obtenus lors de l'application future de la méthode validée seront inclus dans les limites d'acceptation fixées à priori selon les exigences définies.

D'un point de vue statistique, cette stratégie de validation répond aux besoins des analystes quant à la prise de décision en rapport avec les résultats fournis et en considérant les limites d'acceptation prédéfinies ainsi que le risque relatif à l'usage futur de la méthode. De ce fait, l'utilisation du profil d'exactitude se trouve être en accord avec l'objectif de la méthode qui peut être résumé en sa capacité à quantifier le plus exactement possible chacune des quantités inconnues devant être déterminées par le laboratoire.

D'un point de vue pratique [79-90], cette stratégie de validation peut se résumer en six étapes reprises ci-après :

- sélection des limites d'acceptation considérant l'usage attendu de la méthode,
- ajustement d'un modèle de régression à partir des standards de calibration,
- calcul des concentrations des standards de validation à partir du modèle sélectionné,
- détermination du biais moyen à chaque niveau de concentration,
- calcul des limites de tolérance bilatérales pour les résultats attendus au niveau  $\beta$  à chaque niveau de concentration en considérant le biais et l'écart-type de fidélité intermédiaire,
- établissement du profil d'exactitude, représentant comme fonction de la concentration, le biais moyen, l'intervalle de tolérance des valeurs attendues au niveau  $\beta$  et les limites d'acceptation.

L'intervalle de tolérance comprend les deux termes suivants: l'un étant la justesse et l'autre étant, à un facteur près, le coefficient de variation de fidélité intermédiaire. C'est la raison pour laquelle cet intervalle peut ainsi être considéré comme une expression de l'exactitude des résultats. Mais l'intervalle de tolérance intègre une dimension supplémentaire, celle de chance (ou risque) pour des résultats futurs, conditionnellement à des résultats passés [22].

Dès lors, la méthode peut être considérée comme exacte au niveau de chance  $\beta$  pour le niveau de concentration en question, si l'intervalle de tolérance est inclu dans les limites  $[-\lambda, +\lambda]$  définies a priori en fonction des objectifs de la méthode.

Les bornes de ces intervalles sont [22] :

$$L_j = \text{biais } (\%)_j - Q_t\left(v; \frac{1+\beta}{2}\right) \sqrt{1 + \frac{1}{pn B_j^2}} CV_{F.I.,j}$$

$$U_j = \text{biais } (\%)_j + Q_t\left(v; \frac{1+\beta}{2}\right) \sqrt{1 + \frac{1}{pn B_j^2}} CV_{F.I.,j}$$

Avec

$L_j$  borne inférieure de l'intervalle de tolérance du niveau de concentration  $j$

$U_j$  borne supérieure l'intervalle de tolérance du niveau de concentration  $j$

$Q_t\left(v; \frac{1+\beta}{2}\right)$  est le quantile  $\beta$  de la distribution t de Student à  $v$  degrés de liberté.  $p$  est le nombre de série et  $n$  le nombre de répétitions par série.

Avec

$$v = \frac{(R+1)^2}{\frac{\left(R + \frac{1}{n}\right)^2}{p-1} + \frac{1-\frac{1}{n}}{pn}}$$

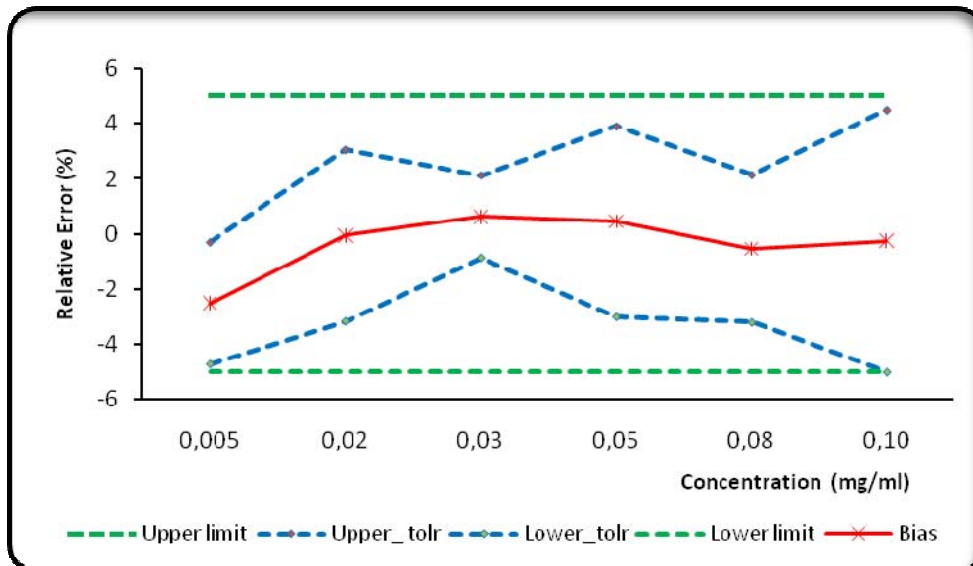
$CV_{F.I.,j}$  : Coefficient de variation de fidélité intermédiaire

$$B_j = \sqrt{\frac{R_j + 1}{nR_j + 1}}$$

$$R_j = \frac{\hat{\sigma}_{B,j}^2}{\hat{\sigma}_{W,j}^2}$$

$$\hat{\sigma}_{F.I.,j}^2 = \hat{\sigma}_{W,j}^2 + \hat{\sigma}_{B,j}^2$$

Le profil d'exactitude de la méthode s'obtient en reliant d'une part les bornes Lj entre elles (L1 → L2 →... → Lm) et d'autre part les bornes Uj entre elles (U1 → U2 → ... → Um) comme dans la Figure 5.



**Figure 5- Exemple de profil d'exactitude**

## **7. Limite de quantification: LQ**

### **a. Concept & Définition**

La limite de quantification LQ et la limite de détection LD sont des critères très souvent évoqués et utilisés en chimie analytique. Malheureusement, les définitions foisonnent, créant une confusion [28].

Nous pouvons trouver une définition de la limite de quantification dans presque tous les documents réglementaires (Cf. Tableau 6).

L'ICH considère que "La limite de quantification est un paramètre des méthodes de dosage des substances présentes en faibles quantités dans les matrices d'échantillon; elle est plus particulièrement utilisée dans le dosage des impuretés et (ou) des produits de dégradation ". Il définit la limite de quantification d'une procédure analytique individuelle comme "la plus faible quantité de la substance analysée que la méthode permet de doser avec un degré acceptable de fidélité et d'exactitude ". Le document ICH définit une seule limite de quantification. Mais l'intervalle de dosage d'une procédure analytique a deux limites : LQI (LQ inférieure) et LQS (LQ supérieure) [1,2,3]. La définition de la limite de quantification extraite d'Eurachem nous laisse comprendre qu'il y a plus qu'une limite de quantification : "les limites de dosage sont les caractéristiques de performance qui marquent la capacité d'un processus d'analyse chimique à quantifier adéquatement une substance". Mais, dans le document Eurachem, seule la "la limite de quantification" LQI est discutée [1].

Le document FDA Bioanalytical Method Validation distingue les deux limites et définit la limite inférieure de quantification (la plus faible quantité de la substance analysée que la méthode permet de doser avec un degré acceptable de fidélité et d'exactitude) et la limite supérieure de quantification (la plus haute quantité de la substance analysée que la méthode permet de doser avec un degré acceptable de fidélité et d'exactitude). Comme on peut le voir dans ce document, la seule différence est la substitution de "la plus faible" avec "la plus haute"[1].

**b. Plan d'expérience**

Le nombre de mesures pour évaluer la LQ n'est pas défini dans les documents ICH, AOAC, IUPAC et SANCO, le FDA recommande 5 mesures et l'EURACHEM en recommande 20.

Aucun document n'adresse un plan expérimental pour estimer la LQ, en terme de répétition par jour.

**c. Tests statistiques & critères d'acceptation :**

Rien n'est précisé sur la façon exacte de mesurer la LQ. Ainsi plusieurs méthodes d'estimation sont proposées par les différents guidelines:

1. Évaluation visuelle (ICH) : appliquée aux méthodes non instrumentales
2. Approche du rapport signal/bruit (ICH) : appliqué aux méthodes avec un bruit de fond
3. Approche de l'écart-type (s) de la réponse et de la pente de la courbe d'étalonnage (ICH)  
 $LQ = 10s/P$  (s= écart-type de la réponse & P= pente de la courbe d'étalonnage)
4.  $LQ = \text{Valeur du blanc} + k \text{ fois écart type (s) du blanc}$  (AOAC, EURACHEM)
5.  $LQ = 5 \times \text{réponse du blanc}$  (FDA)
6. La LQ est égale à la multiplication de la LD (généralement 2) (IUPAC)
7. Calcul de l'écart type à chaque niveau de concentration. Tracer la courbe **écart-type (s) vs concentration** et déterminer la valeur de la limite de quantification (EURACHEM)

↪ L'approche par l'évaluation visuelle (approche 1) n'est appliquée que pour les méthodes non instrumentales, d'où ses limites.

↪ Les principaux problèmes rencontrés avec **l'approche S/N** (Approche 2) sont les suivants :

- Cette méthode ne marche que pour les méthodes dont on peut observer une ligne de base [28], ce qui n'est pas toujours vrai. En outre, lorsque cela est possible, cette approche est tributaire de la manière avec laquelle ce bruit est mesuré et diffère d'un instrument à un autre. Cette approche est dès lors très subjective et dépend de l'équipement utilisé [1].
  - Le signal mesuré n'est pas toujours le signal utilisé pour quantifier l'analyte. Par exemple, dans les chromatographies avec une détection par spectrométrie, le signal mesuré représente les unités d'absorption, c'est-à-dire la hauteur de signal, mais pour la quantification ce sont les surfaces qui sont généralement utilisées. Donc, la LQ ne va pas exprimer le niveau le plus bas de l'analyte, mais la plus faible absorbance quantifiée. Le problème devient plus compliqué dans l'électrophorèse, où on considère d'habitude le signal comme la proportion entre la surface de pic et le temps de migration [1].
- ↪ Les critiques de l'approche basée sur la courbe d'étalonnage (approche 3) sont nombreuses :
- Cette méthode suppose qu'on a toujours une linéarité de la courbe d'étalonnage. Alors, comme nous l'avons vu dans la partie linéarité et fonction de réponse, cette supposition n'est pas toujours vraie (méthodes biologique par exemple).
  - Les mesures utilisées pour la courbe d'étalonnage sont issues des étalons purs alors que la LQ doit être estimée avec des mesures faites dans la matrice.
  - Les étalons doivent être utilisés avec prudence puisque l'estimation de l'écart-type dépend de la gamme de la courbe d'étalonnage: l'ordonnée à l'origine n'est bien estimée que si les concentrations utilisées sont assez petites [1].
- ↪ Les approches utilisant la réponse et l'écart-type du blanc (approche 4 & 5) ont ses limites, puisqu'elles sont applicable avec les médicaments, lorsqu'on maîtrise totalement la matrice (blanc) et qu'on est à même de la doper avec des concentrations très faibles, mais irréalisable dans bien d'autres cas, comme les analyses des matrices biologiques ou aussi les analyses environnementales [28].

Pour conclure, nous pouvons dire que toutes ces approches conduisent à des résultats variables de LQ [28]. C'est une grande problématique puisqu'elles ne permettent pas de comparer la LQ des différents laboratoires utilisant la même procédure analytique [1].

La FDA et l'Eurachem proposent d'autres approches (approches 5 pour le premier & approche 7 pour le deuxième document) pour évaluer la limite inférieure de quantification, basée sur un CV cible [1, 8]. Une courbe est tracée entre les concentrations proches de LQ attendu contre les CV.

Quand la courbe croise la CV cible, le niveau de concentration correspondant est la LQ. Cette approche résout la plupart des problèmes soulignés avec les approches précédentes, comme ça ne dépend désormais plus de l'opérateur et de l'équipement.

Cependant, aucune de ces approches ne répond à la définition de la Limite de Quantification. En effet, même avec cette dernière approche, seule la fidélité de la procédure analytique est évaluée sans une évaluation de la justesse et de l'exactitude (justesse + fidélité) comme exigé dans les définitions.

Les critères d'acceptation proposés dans les différents documents sont les suivants :

**FDA :**

- Réponse à LQI = 5× réponse du blanc
- Le pic de l'analyte doit être identifié avec une fidélité de 20% et justesse de 80–120%

**ICH**

- Un S/N de 10 :1
- $LQ = \frac{10\sigma}{P}$  avec  $\sigma$  = écart-type de réponse

P= pente de la courbe d'étalonnage



**d. Recommandations**

Comme on le voit, il n'existe pas de définition conventionnelle pour la LQ et le choix est large pour l'estimation de ce critère. En fait, la LQ est un critère empirique, censé estimer la limite à partir de laquelle une méthode analytique est capable de fournir un résultat quantitatif exact [28]. Cependant, quelle que soit la méthode employée, rares sont les situations qui répondent à ce besoin [28].

À notre avis, la meilleure façon de calculer les deux limites de quantification (LQS et LQI) est l'utilisation de l'approche du profil d'exactitude [1,2,21,27] qui répond à la définition du critère LQ, en démontrant que l'erreur totale du résultat est connue et acceptable à ces niveaux de concentration de LQ (sup et inf), c'est-à-dire avec les 2 erreurs systématique et aléatoire qui sont acceptables [1].

Comme on a l'vu sur la Figure 5, il existe des concentrations (sup et infé) pour lesquelles une des bornes de l'intervalle de tolérance coupe une des limites d'acceptabilité. Ces concentrations correspondent aux LQS et LQI

Comme montré par Feinberg [28], le choix du nombre de niveaux de concentrations et leurs positions relatives dans le domaine de validation influence le calcul de LQ.

La Figure 6 illustre un profil d'exactitude obtenu avec 3 niveaux de concentration. Dans ce cas-là, la LQ se situe aux environs de 0,5 mg/L. Par contre, avec la même méthode, si on utilise 4 niveaux de concentration (Fig 7), avec les niveaux les plus bas à la proximité de la valeur basse du domaine de validation. Nous pouvons remarquer que la LQ estimée se situe aux alentours de 0,2 mg/l.

Ces remarques montrent l'intérêt du choix des niveaux de concentration et l'importance de multiplier les niveaux dans la partie inférieure du domaine de validation [28].

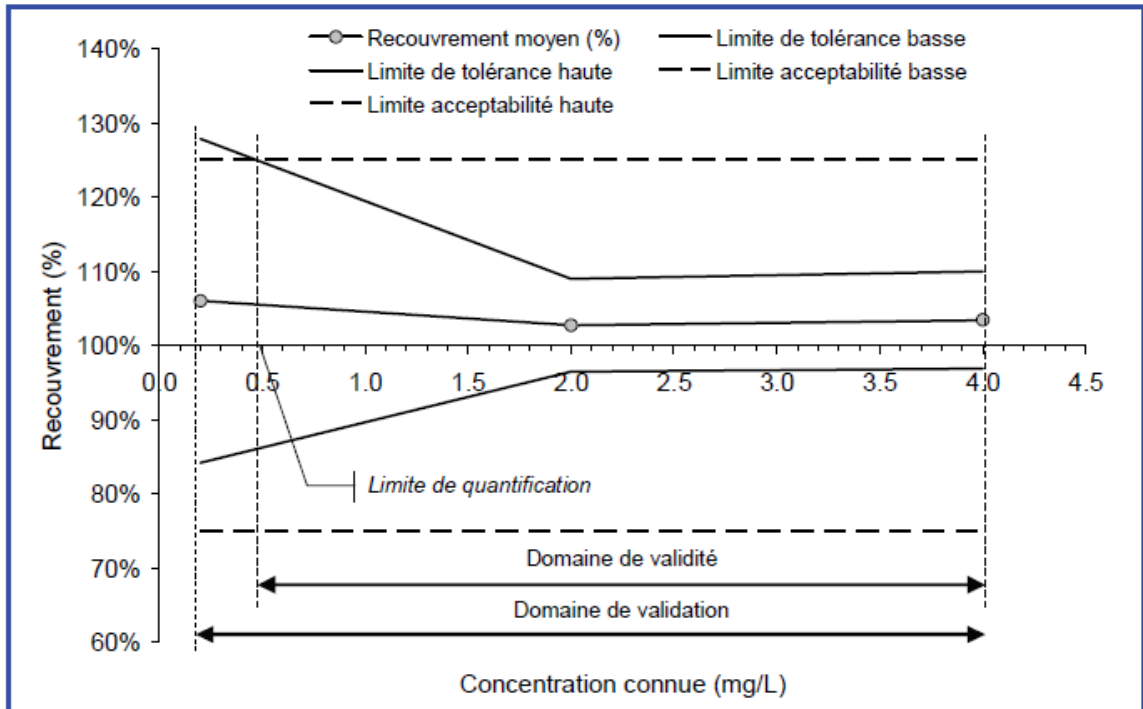


Figure III-6 : Profil d'exactitude (simulé) avec 3 niveaux de concentration [28]. Chaque cercle gris représente la concentration moyenne retrouvée pour le niveau et quantifie la justesse. Les lignes pointillées délimitent les limites d'acceptation et les lignes continues l'intervalle de tolérance. Les traits verticaux délimitent le domaine de validité

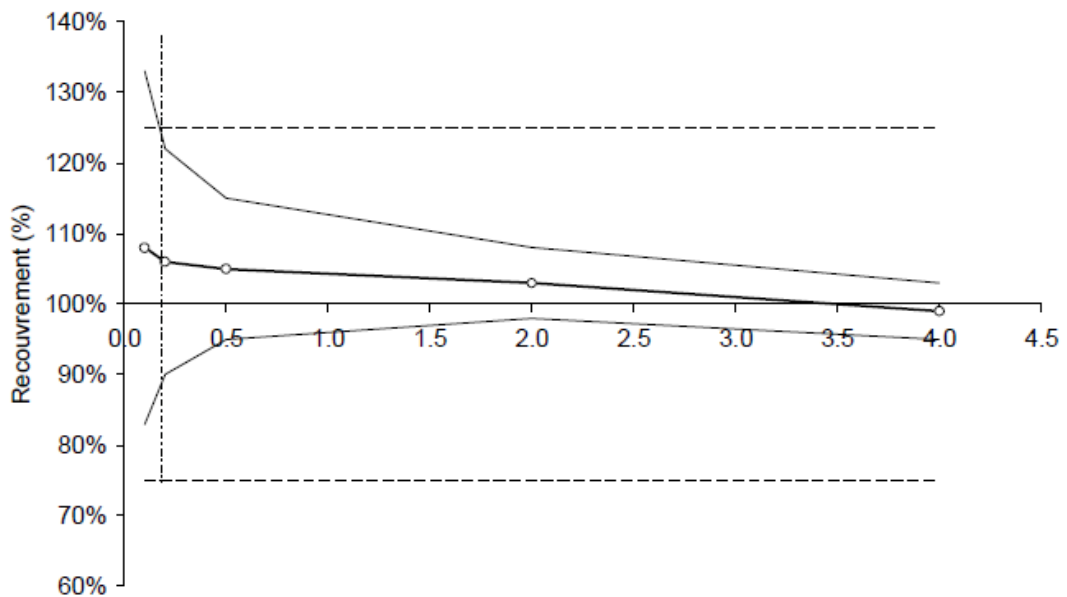


Figure III-7 : Profil d'exactitude (simulé) avec 4 niveaux choisis pour mieux calculer la LQ [28]

## **8. Limite de Détection (LD)**

### **a. Concept & Définition**

La limite de détection (LD) est définie dans tous les documents réglementaires (sauf le document SANCO), la définition générale donnée est comme suit : la limite de détection est la concentration la plus basse d'analyte qui peut être détectée en utilisant la méthode analytique, mais sans nécessairement être quantifiée (Cf. Tableau 6).

### **b. Plan d'expérience**

Le plan expérimental pour évaluer la LD, en termes de répétition par jour, n'est défini dans aucun document (sauf le AOAC qui exige 20 blancs à mesurer & 10 pour EURACHEM).

### **c. Tests statistiques & critères d'acceptation :**

Les mêmes approches proposées pour le calcul de la LQ sont utilisées pour le calcul de LD. Les mêmes critiques cités auparavant dans la partie LQ sont valables pour la LD.

**9. Bibliographie**

- [1] Rozet E., Ceccato A., Hubert C., Ziemons E., Oprean R., Rudaz S., Boulanger B., Hubert Ph., An Analysis of Recent Pharmaceutical Regulatory Documents on Analytical Method Validation, *Journal of Chromatography A*, 1158, 111-125 (2007)
- [2] Hubert Ph., Nguyen-Huu J.-J., Boulanger B., Chapuzet E., Chiap P., Cohen N., Compagnon P.-A., Dewe W., Feinberg M., Lallier M., Laurentie M., Mercier N., Muzard G., Nivet C., Valat L., Validation des procédures analytiques quantitatives Harmonisation des démarches Commission SFSTP, STP Pharma Pratiques - volume 13 – N°3 – (mai/juin 2003)
- [3] Hubert Ph., Nguyen-Huu J.-J., Boulanger B., Chapuzet E., Chiap P., Cohen N., Compagnon P.-A., Dewe W., Feinberg M., Lallier M., Laurentie M., Mercier N., Muzard G., Nivet C., Valat L., Harmonization of strategies for the validation of quantitative analytical procedures, A SFSTP proposal - Part I, *Journal of Pharmaceutical and Biomedical Analysis*, 36, 579-586 (2004)
- [4] US Federal Government, Guidance for industry: Bioanalytical Method Validation, US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Rockville, MD, USA, May 2001.
- [5] International Conference on Harmonization (ICH) of Technical Requirements for registration of Pharmaceuticals for Human Use, Topic Q2 (R1): Validation of Analytical Procedures: Text and Methodology, Geneva, Switzerland, 2005.
- [6] IUPAC Technical Report Harmonized guidelines for single laboratory validation of methods of analysis 2002
- [7] Guidelines for Single Laboratory Validation of Chemical Methods for Dietary Supplements and Botanicals. AOAC, Official Methods of Analysis, Vol. 1, Association of Official Analytical Chemists (AOAC), Arlington, VA, 15th ed., pp. 673-674, 1990,
- [8] The Fitness for Purpose of Analytical Methods A Laboratory guide to method validation and related Topics, Eurachem, Teddington, Middlesex, UK, 1998.
- [9] European Commission Directorate General Health and Consumer Protection SANCO/825/00 rev.7 Guidance document on residue analytical methods 2008 (available at the website: <http://www.crlpesticides.eu/library/docs/srm/AqcGuidance.pdf>).
- [10] Hubert Ph., Nguyen-Huu J.-J., Boulanger B., Chapuzet E., Cohen N., Compagnon P.- A., Dewé W., Feinberg M., Laurentie M., Mercier N., Muzard G., Valat L., Rozet E., Harmonization of strategies for the validation of quantitative analytical procedures A SFSTP proposal–Part III, *Journal of Pharmaceutical and Biomedical Analysis*, 45, 82–96 (2007)
- [11] Hoffman D., Kringle R., Two-Sided Tolerance Intervals for Balanced and Unbalanced Random Effects Models, *Journal of Biopharmaceutical Statistics*, 15, 283-293 (2005)
- [12] Chandran S., Singh R. S. P., Comparison of various international guidelines for analytical method validation *Pharmazie* 62: 4–14 (2007)
- [13] Stöckl D., D'Hondt H., Thienpont L, Method validation across the disciplines– Critical investigation of major validation criteria and associated experimental protocols, *J Chromatogr B Analyt Technol Biomed Life Sci.* 877(23):2180-90, 2009.
- [14] Application of the statistics-accuracy (trueness and precision) of the results and methods of measurement – Parts 1 to 6, International Organization of Standardization (ISO), ISO 5725, ISO, Geneva, Switzerland, 1994.

- [15] Hartmann C., Smeyers-Verbeke J., Massart D.L., McDowall R.D., Validation of bioanalytical chromatographic methods, *Journal of Pharmaceutical and Biomedical Analysis*, 17, 193-218 (1998)
- [16] Rosing H, Man WY, Doyle E, Bult A, Beijnen JH., Bioanalytical liquid chromatographic method validation, a review of current practices and procedures, *J Liq Chromatogr Relat Technol* 23:329-54, 2000
- [17] Ermer J, Validation in pharmaceutical analysis. Part I: an integrated approach, *J. Pharm. Biomed. Anal.* 24, 755-67 (2001)
- [18] Hartmann C, Massart DL, McDowall RD, An analysis of the Washington Conference Report on bioanalytical method validation, *J Pharm Biomed Anal.* 12(11) 1337-43 (1994)
- [19] Bouabidi A., Rozet E., Fillet M., Ziemons E., Chapuzet E., Mertens B., Klinkenberg R., Ceccato A., Talbi M., Streel B., Bouklouze A., Boulanger B., Hubert Ph., Critical analysis of several analytical method validation strategies in the framework of the fit for purpose concept, *Journal of Chromatography A*, 1217 3180–3192 (2010)
- [20] Hubert P., Nguyen-Huu J., Boulanger B., Chapuzet E., Chiap P., Cohen N., Compagnon P., Dewe W., Feinberg M., Laurentie M., Mercier N., Muzard G., Valat L., Validation des procédures analytiques quantitatives, Harmonisation des démarches, Partie II- Statistiques, *S.T.P Pharma Pratiques*, 16 (1), 28 – 58 (2006)
- [21] Hubert P Nguyen-Huu JJ, Boulanger B, Chapuzet E, Cohen N, Compagnon PA, Dewé W, Feinberg M, Laurentie M, Mercier N, Muzard G, Valat L, Validation des procédures analytiques quantitatives, Harmonisation des démarches, Partie III- Exemples d'application, *S.T.P Pharma Pratiques*, 16 (2), 87 – 121 (2006)
- [22] Hubert P., Nguyen-Huu J.-J., Boulanger B., Chapuzet E., Cohen N., Compagnon P.-A., Dewé W., Feinberg M., Laurentie M., Mercier N., Muzard G., Valat L., Rozet E., Harmonization of strategies for the validation of quantitative analytical procedures. A SFSTP proposal-Part II, *Journal of Pharmaceutical and Biomedical Analysis* 45(1), 70-81 (2007)
- [23] Hubert Ph., Nguyen-Huu J.-J., Boulanger B., Chapuzet E., Cohen N., Compagnon P.- A., Dewé W., Feinberg M., Laurentie M., Mercier N., Muzard G., Valat L., Rozet E., Harmonization of strategies for the validation of quantitative analytical procedures A SFSTP proposal-Part III, *Journal of Pharmaceutical and Biomedical Analysis* (45) 82–96 (2007)
- [24] Feinberg M., approche globale et harmonisée de la validation, *Spectra analyse* (249) 16-23 (2006)
- [25] Gustavo Gonzalez A., Angeles Herrador M., Accuracy profiles from uncertainty measurements. *Talanta* (70) 896–901(2006)
- [26] T. Singtoroj, J. Tarning, A. Annerberg, M. Ashton, Y. Bergqvist, N.J. White, N. Lindegardh, N.P.J. Day, *Journal of Pharmaceutical and Biomedical Analysis* (41) 219–227 (2006)
- [27] Streel B., Ceccato A., Klinkenberg R., Hubert Ph., Validation of a liquid chromatographic-tandem mass spectrometric method for the determination of loperamide in human plasma, *J. Chromatogr. B* 814 263–273 (2005)
- [28] Feinberg. M -LABO-STAT - Guide de validation des méthodes d'analyse; Lavoisier Edition Tec & Doc 2009
- [29] Caporal-Gautier J., Nivet J.M., Algranti P., Guilloteau M., Histe M., Lallier L., N'Guyen-Huu J.J., Russotto R., Guide de validation analytique. Rapport d'une commission SFSTP. I : Méthodologie, *STP Pharma Pratiques*, 2, 205-226 (1992)

## 10. Tableaux

Tableau III-1: Comparaison du critère de spécificité dans les différents guidelines

Guidelines	Concept & Définition	Plan expérimental	Tests & Critères d'acceptation
<b>ICH</b>	<b>(La spécificité)</b> La propriété qui fait qu'une méthode d'analyse rende compte sans ambiguïté de la substance analysée en présence d'autres composantes normalement présentes.	<ul style="list-style-type: none"> <li>▪ Impuretés disponibles : on enrichit la substance pure (substance médicamenteuse ou produit fini) en y ajoutant une quantité adéquate d'impuretés et (ou) d'excipients puis on démontre que les résultats obtenus ne sont pas influencés</li> <li>▪ Impuretés non disponibles : On compare les résultats obtenus avec des échantillons renfermant des impuretés ou des produits de dégradation aux résultats obtenus avec une autre méthode d'analyse bien connue. Ces comparaisons doivent inclure des échantillons exposés à des conditions de stress</li> </ul>	<ul style="list-style-type: none"> <li>▪ Démontrer une distinction entre la substance à analyser et les autres substances présentes</li> <li>▪ Evaluation de pureté des pics peut être utile pour démontrer que le pic chromatographique correspondant à la substance à analyser n'est pas attribuable à plus d'une substance</li> <li>▪ Pour la teneur, comparer les résultats obtenus par les deux méthodes et pour l'analyse des impuretés, comparer les profils des impuretés donnés avec les 2 méthodes.</li> </ul>
<b>US FDA</b>	<b>(Sélectivité)</b> La propriété d'une méthode d'analyse de différencier et de quantifier la substance à analyser en présence d'autres composantes dans l'échantillon	<ul style="list-style-type: none"> <li>▪ Analyse des échantillons blancs de matrice biologique appropriée (plasma, urine ou autre) obtenue d'au moins <b>six sources</b> et l'interférence doit être testée sur ces échantillons.</li> </ul>	Chaque échantillon doit être testé pour vérifier l'interférence La sélectivité doit être assurée au LQI limite inférieure de quantification
<b>AOAC</b>	<b>(Sélectivité)</b> est le degré avec lequel une méthode est capable de mesurer l'analyte-cible en présence d'autres analytes, matrices ou d'autres interférents	<ul style="list-style-type: none"> <li>▪ En présence des analytes d'accompagnement ou des matrices les plus susceptibles d'interférer.</li> <li>▪ d'un blanc de matrice (pour les faibles niveaux d'analyte)</li> <li>▪ Si le produit est mélangé avec d'autres substances, les substances ajoutées doivent être testés pour vérifier l'interférence</li> <li>▪ Si on est en présence d'un mélange de principe actif (2 ou plus), il faut procéder à une séparation.</li> </ul>	Critère pour la résolution du pic : $R_s \geq 2$
<b>IUPAC</b>	<b>(Sélectivité)</b> Le degré avec lequel une méthode peut évaluer quantitativement avec exactitude la substance analysée en présence d'interférents.		Indice de sélectivité doit être calculé Indice de sélectivité = $b_{an}/b_{int}$ $b_{an}$ = pente de la courbe d'étalonnage $b_{int}$ = pente de la réponse produite indépendamment par des substances susceptible d'influencer les résultats
<b>SANCO 10684</b>	<b>(Sa spécificité)</b> La capacité d'un détecteur à fournir des signaux pour identifier efficacement la substance à analyser	<ul style="list-style-type: none"> <li>▪ Réponse du blanc réactif et de l'échantillon</li> </ul>	<ul style="list-style-type: none"> <li>▪ La limite de quantification des méthodes doit être en relation avec le niveau minimum de risque (MRL)</li> </ul>
<b>EURACHEM</b>	<b>Sélectivité (ou Spécificité):</b> La capacité d'une méthode de déterminer d'une façon exacte et spécifique la substance à analyser en présence des autres composants dans la matrice de l'échantillon selon les conditions décrites de l'analyse	<ul style="list-style-type: none"> <li>▪ Analyser les échantillons et la substance de référence en comparaison avec des méthodes indépendantes (1 fois)</li> <li>▪ Analyser des échantillons contenant les différents interférents suspectés en présence de la substance à analyser (1 fois)</li> </ul>	<ul style="list-style-type: none"> <li>▪ Utiliser les résultats des techniques de confirmation pour évaluer la capacité de la méthode à confirmer l'identité de la substance à analyser et sa capacité à la mesurer séparément des autres interférences</li> <li>▪ examiner les effets des interférences – est ce que la présence interfère ou inhibe la détection ou la quantification des substances à analyser</li> </ul>

Tableau III-2: Comparaison du critère de linéarité dans les différentes directives

Guidelines	Concept & Définition	Plan expérimental	Tests & Critères d'acceptation
ICH	Capacité de donner des <b>résultats</b> qui sont directement (à l'intérieur de certaines limites) proportionnels à la <b>concentration (quantité)</b> de la substance analysée dans un échantillon.	5 concentrations au minimum sont recommandées Substance médicamenteuse (différentes dilutions) et (ou) en utilisant des portions pesées individuellement de mélanges synthétiques.	<b>Examen visuel</b> Les <b>résultats</b> doivent être évalués au moyen des méthodes statistiques appropriées, comme, par exemple, la régression des moindres carrés : <b>Coefficient de corrélation, ordonnée à l'origine, pente, courbe de régression, ainsi que la somme des carrés des résidus</b>
US FDA	NA		
AOAC	NA		
IUPAC	pas de définition	<ul style="list-style-type: none"> <li>• Six standards d'étalonnage <b>ou plus</b>, également espacés, couvrant l'intervalle de dosage</li> <li>• Intervalle : 0–150% ou 50–150% de la concentration cible (<b>en fonction de la performance recherchée</b>)</li> <li>• Les standards d'étalonnage doivent être analysés au moins en <b>double et de préférence en triple</b> ou plus</li> </ul>	<ul style="list-style-type: none"> <li>▪ Examen Visuel de la courbe des résidus</li> <li>▪ Application des tests significatifs statistiques</li> <li>▪ Test de manque d'ajustement peut être pondéré / régression simple</li> </ul>
SANCO	NA		
EURACHEM	Linéarité : Définit la capacité de la méthode à obtenir des résultats proportionnels à la concentration de l'analyte.	<ul style="list-style-type: none"> <li>• Blancs dopés aux différents niveaux de concentration (Il faut au moins <b>6 concentrations</b> plus blanc) (1 répétition)</li> <li>• substance de référence ou des échantillons de blanc enrichi à au moins <b>6 concentrations</b> différentes au sein de la gamme de linéarité (<b>3 répétitions</b>)</li> <li>• 3 au niveau LQ (limite de quantification)</li> </ul>	<ul style="list-style-type: none"> <li>▪ Tracer la mesure de réponse (axe Y) en fonction de la concentration de la substance (axe X)</li> <li>▪ Examiner visuellement les valeurs aberrantes qui peuvent ne pas être reflétées dans la régression</li> <li>▪ Calculer le coefficient de régression approprié.</li> <li>▪ Calculer et tracer les valeurs résiduelles (différence entre la valeur réelle y et la valeur y prédite par une ligne droite, pour chaque valeur de x). La distribution aléatoire autour de la ligne droite confirme la linéarité. Les tendances systématiques indiquent la non-linéarité</li> </ul>



Tableau III-2' Comparaison du critère de fonction de réponse dans les différentes directives

Guidelines	Concept & Définition	Plan expérimental	Tests & Critères d'acceptation
ICH	On parle de linéarité		
US FDA	Courbe d'étalonnage: Relation entre la réponse de l'équipement et la concentration connue d'analyte.	<p>La courbe d'étalonnage doit être préparée par dopage de la matrice par des concentrations connues de la substance à analyser :</p> <ul style="list-style-type: none"> <li>• Blanc : matrice sans standard interne SI</li> <li>• Un échantillon zéro (matrice + SI)</li> <li>• <b>6-8</b> échantillons non zéro couvriront l'intervalle de dosage incluant LQI (Limite de Quantification Inférieure)</li> </ul>	<p>A LQI: <math>\leq 20\%</math> d'écart de la valeur cible                      Plus que LQI: <math>15\%</math> d'écart                      Au moins 4 de 6 échantillons non zéro devrait être conformes aux critères incluant LQI et la concentration maximale</p>
AOAC	Les méthodes instrumentales modernes dépendent de la comparaison d'un signal d'une concentration inconnue d'un analyte à partir de celui d'une concentration connue de l'analyte identique ou similaire.	<ul style="list-style-type: none"> <li>○ Préparation d'une série de solution standard de <b>6 à 8</b> points, équidistants de la valeur nominale en <b>double répétition pendant 2 jours</b></li> </ul>	<ul style="list-style-type: none"> <li>○ Tracer la droite d'étalonnage en fonction de la concentration.</li> <li>○ Un ajustement acceptable produit de façon aléatoire des résidus avec une moyenne 0. Tracer le signal de réponse en fonction de la concentration.</li> <li>○ Une réponse linéaire est souhaitable, mais n'est pas nécessaire et n'est pas une caractéristique de performance requise.</li> <li>○ Le coefficient de corrélation élevé (par exemple, <math>&gt; 0,99</math>) ne doit être utilisé comme une preuve de qualité de l'ajustement.</li> <li>○ L'Examen visuel est suffisant pour vérifier la linéarité</li> <li>○ Test résiduel</li> </ul>
IUPAC	NA		
SANCO	<b>Etalonnage</b> : Détermination de la relation entre le signal observé (réponse produite par le système de détection) de la substance à analyser... et les quantités connues de l'analyte préparée comme solutions standard.		<ul style="list-style-type: none"> <li>○ A travers la courbe d'étalonnage</li> <li>○ Critères d'acceptation : Résidus <math>&lt; \pm 20\%</math></li> </ul>
EURACHEM	La relation de réponse de l'instrument à la concentration ne doit pas être parfaitement linéaire pour que la méthode soit efficace, mais cette courbe doit être répétable au jour le jour. Courbe de calibration : Représentation graphique de signal en fonction de la quantité d'analyte		

Tableau III-3: Comparaison du critère de fidélité selon les différentes directives

Guidelines	Concept & Définition	Plan expérimental	Tests & Critères d'acceptation
ICH	Correspond au degré d'accord (degré de dispersion) entre les résultats des mesures obtenues par l'analyse individuelle de plusieurs prélèvements d'un même échantillon homogène, prélevés dans des conditions prescrites.	<p><b>La répétabilité</b> (mêmes conditions de réalisation, après un court intervalle de temps): Au moins 9 mesures englobant l'écart d'utilisation de la méthode (3 concentrations avec 3 échantillons chacune); Ou Au moins 6 mesures d'une concentration à 100 % de la teneur escomptée.</p> <p><b>La précision intermédiaire</b> correspond aux variations survenant dans un même laboratoire : analyses effectuées dans des jours différents, par des personnes différentes, au moyen d'appareils différents, etc.</p> <p><b>La reproductibilité</b> correspond à la concordance entre laboratoires (travaux de collaboration visant généralement l'uniformisation de la méthodologie).</p>	Ecart type, Ecart type relatif et intervalle de confiance La méthode analytique est appliquée avec répétabilité sur différents prélèvements d'un même échantillon homogène de la matrice biologique
US FDA	Décrit le degré d'accord des mesures individuelles de l'analyte quand la méthode est appliquée avec répétabilité sur plusieurs prélèvements d'un même échantillon homogène de la matrice biologique	<p>5 répétitions pour chaque niveau de concentration (au moins 3 niveaux)</p> <ul style="list-style-type: none"> <li>▪ <b>concentration basse</b> de l'écart × 5 rép</li> <li>▪ <b>concentration intermédiaire</b> de l'écart × 5 répétitions</li> <li>▪ <b>concentration maximale</b> de l'écart × 5 rép</li> </ul> <p>2 types : Inter série/ Intra lot (ou répétabilité)</p>	<p>Coefficient de variation</p> <p>_ A LQI:CV ≤20%</p> <p>_ Autre que LQI ≤15%</p>
AOAC	Degré d'accord des mesures dans des conditions spécifiques	<p>5 séries</p> <p><b>Répétabilité:</b> sera établie dans le même laboratoire (jours différents/analystes différents/courbes d'étalonnages différents/Lots différents de réactifs/matrices différentes)</p> <p><b>Reproductibilité:</b> Inter laboratoires (8 laboratoires)</p>	<p>Détermination du CV et de l'écart-type du résultat de plusieurs prélèvements</p> <p>Ecart-type relatif</p>
IUPAC	Degré d'accord entre les résultats des mesures indépendants dans des conditions spécifiques	<p>Simple évaluation de la fidélité au niveau ou à proximité des extrémités de l'intervalle de validation</p>	<p>Détermination de la variance ou du coefficient de variation et test F du résultat de plusieurs prélèvements des erreurs normalement distribué est appliquée</p> <p>Deux conditions : <b>(a) Répétabilité</b> (variations au cours d'une même série), écart-type <math>\sigma_r</math>, et <b>(b) Précision</b> inter-série, biais <math>\delta</math>run écart-type <math>\sigma_{run}</math>.</p> <p>Les 2 sources d'erreur sont combinées</p> <p>Ecart type,  <math>\{(\sigma_{tot} = \sigma_r^2/n + \sigma_{run}^2)^{1/2}</math> Avec n = nombre de répétition</p> <p>Changement de l'opérateur, instrument, jour...</p>
SANCO	La valeur en dessous de laquelle la différence absolue entre deux résultats individuels pour un produit identique, obtenu dans des conditions de répétabilité ou de reproductibilité	<ul style="list-style-type: none"> <li>▪ Déterminée à partir des analyses répétées des standards d'étalonnage</li> <li>3 niveaux:</li> <li>▪ <b>Répétabilité r (RSD<sub>r</sub>)</b>: un seul laboratoire/ courte période/même analyte/ même équipement</li> <li>▪ <b>Reproductibilité R ((RSD<sub>R</sub>)</b>: plusieurs laboratoires/ différents analyte / Différents équipement</li> </ul>	<p>La mesure de la précision est généralement exprimée en terme d'écart-type. Critère d'acceptation ≤ 20%</p>

		<p>▪ <b>Within-laboratory reproducibility</b> : mêmes conditions de reproductibilité dans un seul laboratoire</p>	
<b>EURACHEM</b>	<p>.Étroitesse de l'accord entre les résultats indépendants d'essai conformément aux conditions stipulées</p>	<p>Les substances standards ou les échantillons blancs dopés à différentes concentrations dans l'intervalle de travail</p> <p>a) Même analyste, même équipement, même laboratoire, en courte période (10 fois)</p> <p>b) Différents : analytes, équipements, même laboratoire, différentes périodes (10 fois)</p> <p>c) Différents : analystes, équipements, laboratoires, différentes périodes (10 fois)</p>	<p>Déterminer l'écart type à chaque concentration</p>

Tableau III-4: Comparaison du critère de justesse dans les différentes directives

Guidelines	Concept & Définition	Plan expérimental	Tests & Critères d'acceptation
<b>ICH</b>	Correspond au degré de concordance entre la <b>valeur de la méthode obtenue</b> et la <b>valeur de référence</b> ou la valeur considérée comme véritable par convention. (L'exactitude est aussi désignée par la justesse).	Minimum 9 déterminations à partir de 3 concentrations (n = 3) couvrant l'intervalle de linéarité	<ul style="list-style-type: none"> <li>• Application de la méthode à des mélanges synthétiques avec une pureté connue</li> <li>• Comparaison des résultats avec ceux obtenus par une méthode bien caractérisée,</li> <li>• Déduction de l'exactitude après avoir établi la précision, la linéarité et la spécificité de la méthode en question.</li> <li>• Pourcentage de récupération de la substance dont une quantité connue a été ajoutée à l'échantillon</li> <li>• Différence entre la valeur moyenne et la valeur considérée comme véritable avec les intervalles de confiance correspondantes.</li> </ul>
<b>US FDA</b> Bioanalytical Method Validation	(exactitude) décrit le degré de concordance entre la moyenne des résultats obtenus par la méthode et la <b>valeur de référence</b> (concentration)  Parfois, Elle est dénommée justesse	Trois niveaux de concentrations couvrant l'écart d'utilisation : <ul style="list-style-type: none"> <li>▪ <b>concentration basse</b> de l'écart × 5 rép</li> <li>▪ <b>concentration intermédiaire</b> de l'écart × 5 répétitions</li> <li>▪ <b>concentration maximale</b> de l'écart × 5 répétitions des échantillons contenant des quantités connues de l'analyte</li> </ul>	Justesse = Moyenne - valeur vraie  A LQI: valeur moyenne doit être incluse dans ± 20% de la valeur vraie autre que LQI: valeur moyenne doit être incluse dans ± 15 % de la valeur vraie
<b>AOAC</b>	Correspond au degré de concordance entre la valeur trouvée et la <b>valeur de référence</b> Exactitude : degré de concordance entre la valeur trouvée et la <b>valeur vraie ou acceptée</b> Plusieurs sens selon qu'il s'agit d'une valeur individuelle ou d'une valeur moyenne Quand il s'agit de valeur moyenne, on utilise le terme «justesse».	Un compromis raisonnable consiste à obtenir 10 valeurs à partir d'une substance de référence, une matrice dopée ou par la méthode des ajouts dosés dispersés sur plusieurs jours ou dans des séries différentes comme base pour le contrôle de biais ou de recouvrement	<ul style="list-style-type: none"> <li>• utilisation des substances de référence</li> <li>• utilisation des méthodes de référence avec des incertitudes connues</li> <li>• utilisation de recouvrement (ou biais) à partir des échantillons chargés</li> </ul> % Recouvrement = (moyenne de concentration fortifiée — moyenne de concentration non fortifiée)/ajout connu dans la concentration Recouvrements de moins de 60-70% devraient être l'objet d'enquêtes menant à l'amélioration et le recouvrement
<b>IUPAC</b>	<b>La justesse</b> est le degré de concordance entre un résultat et une valeur de référence acceptée. Un biais plus petit signifie une bonne justesse.	NA	Prendre en compte les différents biais : Biais de la série/laboratoire/méthode Le biais de la méthode est déterminé en comparant le résultat de la méthode à une référence comme : <ul style="list-style-type: none"> <li>• Utilisation d'une substance de référence</li> <li>• Utilisation d'une méthode de référence</li> <li>• Utilisation d'un échantillon chargé/ recouvrement</li> </ul> Les tests de signification statistique sont recommandés
<b>SANCO 10684</b>	Le degré de la concordance entre la valeur moyenne obtenue à partir d'une série de résultats de mesures (c.-à-dire recouvrement moyen) et une référence acceptée ou une vraie valeur (ISO 5725-1). La mesure de la justesse est habituellement exprimée en biais	NA	Déterminer le recouvrement moyen de deux niveaux dopés Critère d'acceptation 70-120%
<b>EURACHEM</b>	La «justesse» (d'une méthode) est une expression de l'accord entre la moyenne d'une série de résultats (produits par la méthode) et de la vraie valeur. La justesse est généralement exprimée en termes de biais.	a) Blanc et substance de référence en utilisant la méthode étudiée (10 fois) b) Blanc et substance de référence et essai (échantillon) / en utilisant la méthode étudiée et une méthode indépendante (10 fois)	<ul style="list-style-type: none"> <li>▪ Valeur moyenne du blanc soustrait de la valeur moyenne des analytes de substance de référence. Comparer avec la valeur vraie ou acceptée comme vrai pour la substance de référence</li> <li>Donne une mesure du biais de la méthode</li> <li>▪ Valeur moyenne du blanc soustrait de la valeur moyenne de la substance de référence/essai</li> <li>Comparer avec des mesures analogues réalisées à l'aide d'une méthode indépendante / primaire. Donner une mesure du biais de la méthode par rapport à la méthode indépendante / primaire</li> </ul>

Tableau III-5: Comparaison du critère de la limite de quantification dans les différentes directives

Guidelines	Concept & Définition	Plan expérimental	Tests & Critères d'acceptation
<b>ICH</b>	La plus faible quantité de la substance analysée que la méthode permet de doser avec un degré acceptable de précision et d'exactitude.	Analyse d'un nombre suffisant d'échantillons dont la teneur en substance à analyser est proche de la limite de dosage ou coïncide avec cette limite.	<ol style="list-style-type: none"> <li>1. Par Évaluation visuelle</li> <li>2. Basée sur rapport signal/bruit (méthodes avec un bruit de fond) :</li> <li>3. Basée sur <ul style="list-style-type: none"> <li>_ l'écart-type de blancs</li> <li>_ l'écart-type résiduel d'une courbe de régression</li> <li>_ l'écart-type de l'ordonnée à l'origine de courbes de régression</li> </ul> </li> </ol>
<b>US FDA</b>	Limite de quantification inférieure & supérieure (LQI & LQS) : la plus faible/grande quantité de la substance analysée que la méthode permet de doser avec un degré acceptable de précision et d'exactitude.	Au minimum 5 échantillons indépendants	<p>LQI : doit satisfaire les conditions suivantes.</p> <ul style="list-style-type: none"> <li>▪ Réponse à LQI= 5× réponse du blanc</li> <li>▪ le pic de l'analyte doit être identifié avec une précision de 20% et exactitude de 80–120%</li> </ul>
<b>AOAC</b>	Limite de détermination : La plus faible quantité de la substance analysée que la méthode permet d'estimer avec une fiabilité acceptée		Valeur du blanc + 10 fois écart type de blanc: (xBI + 10sBI)
<b>IUPAC</b>	une concentration en dessous de laquelle la méthode analytique ne peut pas fonctionner avec une fidélité acceptable		<ul style="list-style-type: none"> <li>▪ Le CV à fidélité au niveau de LQ doit être = 10 %</li> <li>▪ La LQ est égale à la multiplication de la LD (généralement 2)</li> </ul>
<b>SANCO 10684</b>	Limite de quantification (aussi connue comme limite de détermination, LD), est la concentration minimale ou de la quantité de l'analyte qui peut être quantifiée avec une exactitude et fidélité acceptables		Devrait s'appliquer à la méthode complète. Diversement définie, mais doit être une valeur supérieure à la limite de détection. Avec la plupart des méthodes et des systèmes de détermination, la LQ n'a pas de valeur fixe. ≤ MRL (Maximum residue level)
<b>EURACHEM</b>	Egale ou supérieure au point de concentration la plus faible de la courbe d'étalonnage Elle est aussi connue sous le nom de la limite de déclaration (limit of reporting) : La plus faible concentration d'analyte qui peut être déterminé avec un niveau acceptable de fidélité (répétabilité) et exactitude justesse dans les '.	<ol style="list-style-type: none"> <li>a) 10 échantillons à blanc mesurés une seul fois</li> </ol> <p>Avoir des échantillons à blanc peut être difficile :</p> <ol style="list-style-type: none"> <li>b) Échantillons du blanc dopés avec différentes concentrations de l'analyte proche de la Limite de Détection (10 répétitions indépendantes pour chaque niveau)</li> </ol>	<ol style="list-style-type: none"> <li>a) La LQ est la concentration du blanc plus 5, 6 ou 10 écart-type (s) de la moyenne du blanc.</li> <li>b) Calculer l'écart-type s à chaque concentration. Tracer la courbe écart-type (s) vs concentration et déterminer la valeur de la limite de quantification</li> </ol>

Tableau III-6: Comparaison du critère de la limite de détection (LD) dans les différentes directives

Guidelines	Concept & Définition	Plan expérimental	Tests & Critères d'acceptation
<b>ICH</b>	La plus faible quantité de la substance analysée que la méthode permet de détecter, sans nécessairement fournir la valeur exacte.	Non spécifié	1. Par Évaluation visuelle 2. Basée sur le rapport signal/bruit ( Appliquée aux méthodes d'analyse avec un bruit de fond) 3. Basée sur l'écart-type : _ l'écart-type de blancs _ l'écart-type résiduel d'une courbe de régression _ l'écart-type de l'ordonnée à l'origine de courbes de régression.  Critères d'acceptation Rapport S/N > 2-3;
<b>US FDA</b>	La plus faible concentration d'un analyte que la méthode bioanalytique peut différencier, d'une façon fiable, du bruit de fond	Non spécifié	Non spécifié
<b>AOAC</b>	est utilisé pour le contrôle des impuretés indésirables qui sont spécifiées comme inférieures au niveau seuil de contamination	Non spécifié	Valeur du blanc + 3 fois écart type de blanc : $X_{BI} + 3\sigma_{BI}$ $X_{BI}$ est la valeur du blanc et $\sigma_{BI}$ est l'écart type du blanc
<b>IUPAC</b>	La plus faible quantité de la substance analysée qui peut être distincte de zéro (n'est pas une exigence)	Précisons qu'au moins 6 indépendante concentration de l'analyte (S0) dans une matrice de blanc ou à faible niveau de la substance à analyser	Limite de détection = 3S0.
<b>SANCO 10684</b>	NA	Non spécifié	Non spécifié
<b>EURACHEM</b>	AOAC (est utilisée pour le contrôle des impuretés indésirables qui sont spécifiées comme inférieures du niveau seuil de contamination)	a) 10 échantillons blanc indépendants mesurés une seule fois ou b) 10 échantillons blanc indépendants enrichis par la concentration acceptée comme la plus faible mesurée une seule fois	Ecart type de l'échantillon 'σ' de a) blanc ou b) échantillon dopé Exprimé la LD comme la concentration en analyte correspondent à a) la valeur moyenne du blanc + 3σ ou b) 0 + 3σ  Ecart type de l'échantillon 'σ' de Exprimé la LD comme la concentration en analyte correspondent à la valeur l'échantillon blanc + +4.65σ (dérive du test des hypothèses)

## **CHAPITRE IV : Analyse critique de différentes approches de validation analytique**

Ce chapitre correspond à l'article « **Critical analysis of several analytical method validation strategies in the framework of the fit for purpose concept** » publié dans *Journal of Chromatography A*, 1217 (2010) 3180–3192

- 1. Synthèse et conclusion**
- 2. Article**

### 1. Synthèse et conclusion

La Validation des méthodes analytiques est une étape obligatoire survenant à la fin du développement adoptée par tous les laboratoires d'analyse. C'est une étape hautement réglementée du cycle de vie d'une méthode d'analyse quantitative. Cependant, même si certains documents ont été publiés, il y a un manque de directives claires pour la méthodologie à suivre pour décider correctement quand une méthode peut être considérée comme valide. Cette situation a conduit à la disponibilité de plusieurs approches méthodologiques et il est donc de la responsabilité de l'analyste de choisir la meilleure. Dans ce travail, les processus de décision classiques rencontrés lors de l'évaluation de la validation des méthodes sont comparés, à savoir l'approche descriptive, l'approche de différence et l'approche d'équivalence. Par ailleurs, une approche de validation utilisant le profil d'exactitude calculé au moyen de l'intervalle de tolérance  $\beta$ -expectation et de la mesure de l'erreur totale est également disponible. Dans le présent travail, l'ensemble de ces différentes approches de validation a été appliqué à la validation des deux méthodes d'analyse à savoir :

- Le dosage par HPLC de la loratadine
- Le dosage par HPLC du paracétamol et de la codéine

L'évaluation des risques producteur et consommateur en utilisant des simulations Monte Carlo a également été réalisée afin de comparer la pertinence de ces différentes approches. Les méthodologies classiques donnent lieu à des conclusions insuffisantes et contradictoires qui ne leur permettent pas de répondre adéquatement aux objectifs de la validation de méthode. Cet objectif consiste à donner suffisamment de garanties que chacun des résultats futurs générés par la méthode lors de son utilisation en routine sera assez proche de la vraie valeur. Il est constaté que la méthodologie de validation qui donne le plus de garanties quant à la fiabilité ou la pertinence de la décision de considérer une méthode comme valide est celle qui est basée sur l'utilisation du profil d'exactitude.



2. Article

[Critical analysis of several analytical method validation strategies in the framework of the fit for purpose concept](#)

*Journal of Chromatography A, Volume 1217, Issue 19, 7 May 2010, Pages 3180-3192*

A. Bouabidi, E. Rozet, M. Fillet, E. Ziemons, E. Chapuzet, B. Mertens, R. Klinkenberg, A. Ceccato,  
M. Talbi, B. Streel, A. Bouklouze, B. Boulanger, Ph. Hubert



Volume 1217, Issue 19, 7 May 2010 ISSN 0021-9673

34TH INTERNATIONAL SYMPOSIUM ON HIGH PERFORMANCE LIQUID PHASE SEPARATIONS AND RELATED TECHNIQUES  
Dresden, Germany, 28 June–2 July 2009  
Guest Ed.: C.G. Huber

## JOURNAL OF CHROMATOGRAPHY A


INCLUDING ELECTROPHORESIS, MASS SPECTROMETRY AND OTHER SEPARATION AND DETECTION METHODS

**EDITORS**  
J.G. Dorsey (Tallahassee, FL)  
S. Fanali (Florence)  
R.W. Gliese (Boston, MA)  
P.R. Haddad (Hobart)  
C.F. Poole (Detroit, MI)  
M.-L. Riekkola (Helsinki)  
P.J. Schoenmakers (Amsterdam)  
N. Tanaka (Kyoto)

**EDITOR, SPECIAL VOLUMES**  
U.A.H. Brinkman (Amsterdam)

**HONORARY EDITORIAL BOARD**  
E. Heilmann (Water Creek, CA)  
E.L. Harger (Boston, MA)  
V. Stoyko (Chicago)  
L.R. Snyder (Orinda, CA)  
S. Terabe (Nagoya)

**EDITORIAL BOARD**  
A. Berthod (Vaulx-en-Velin)  
M. Caron (Paris)  
Y. Chen (Beijing)  
T. Chankar (Chattanooga, TN)  
A. Chiari (Modena)  
B. Dierckx (Ghent)  
F. Dujardin (Louvain-la-Neuve)  
A. Felinger (Paris)  
F. Fontijn (Louvain-la-Neuve)  
R. Frutiger (Louvain-la-Neuve)  
M. G. Garcia (Barcelona)  
M.C. Garcia-Abaza (Córdoba, Valencia)  
G.A. Guiochon (Cincinnati, OH)  
Y. Ishihama (Tokyo)  
F. Jandacek (Prague)  
H.-B. Jonsson (Münster)  
M. Junghans (Munich)  
M. Lammertink (Veenendaal)  
H.K. Lee (Singapore)  
C.A. Lucy (Edinburgh)  
D. McCool (Bristol)  
I. Némethy (Budapest)  
J. Némethy (Edinburgh)  
U. Niessner (Munich)  
H. Noth (Hofheim)  
S. Noth (Hofheim)  
Y. Picot (Nantes)  
P.S. Rieck (Munich)  
M. Roeske (Bonn)  
J. C. Sanchez (Cottbus, MD)  
F. Sauer (Hofheim)  
A. Sauerbrey (Magdeburg)  
Y. Sun (Tianjin)  
E. Szep (Berkeley, CA)  
R.E. Symcox (Seattle, WA)  
T.A. Van Beeck (Veghel)  
P.D. Williams (Raleigh, NC)  
S. Wolfgramm (San Antonio, TX)  
H. Zou (Tianjin)

Available online at  
  
www.sciencedirect.com

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

## Journal of Chromatography A

journal homepage: [www.elsevier.com/locate/chroma](http://www.elsevier.com/locate/chroma)

## Critical analysis of several analytical method validation strategies in the framework of the fit for purpose concept

A. Bouabidi<sup>a,1</sup>, E. Rozet<sup>b,1</sup>, M. Fillet<sup>c</sup>, E. Ziemons<sup>b</sup>, E. Chapuzet<sup>d</sup>, B. Mertens<sup>e</sup>, R. Klinkenberg<sup>e</sup>, A. Ceccato<sup>f</sup>, M. Talbi<sup>a</sup>, B. Streel<sup>e</sup>, A. Bouklouze<sup>g</sup>, B. Boulanger<sup>h</sup>, Ph. Hubert<sup>b,\*</sup>

<sup>a</sup> Analytical Chemistry Laboratory, UFR: Sciences et Procédés d'Analyse physico-chimiques Université Hassan II - Mohammedia - Faculté des Sciences Ben M'Sik, Casablanca, Morocco

<sup>b</sup> Analytical Chemistry Laboratory, CIRP, Institute of Pharmacy, University of Liège, Liège, Belgium

<sup>c</sup> Analytical Pharmaceutical Chemistry Laboratory, CIRP, Institute of Pharmacy, University of Liège, Liège, Belgium

<sup>d</sup> Qualilab, 1121 rue de la Bergeresse, 45160 Olivet, France

<sup>e</sup> Galephar M/F, 39 Rue du Parc Industriel, 6900 Marche en Famenne, Belgium

<sup>f</sup> Mithra Pharmaceuticals SA, 1 Rue Sur le Foulons, 4000 Liège, Belgium

<sup>g</sup> Laboratoire de Pharmacologie-Toxicologie, Faculté de Médecine et de Pharmacie, Université Med V Soussi - Rabat, Morocco

<sup>h</sup> UCB Pharma SA, Chemin du Foriest, B-1420 Braine-L'alleud, Belgium

## ARTICLE INFO

## Article history:

Available online 25 August 2009

## Keywords:

Validation methodology  
Accuracy profile  
Tolerance intervals  
Hypothesis tests  
Total error  
Fit for purpose

## ABSTRACT

Analytical method validation is a mandatory step at the end of the development in all analytical laboratories. It is a highly regulated step of the life cycle of a quantitative analytical method. However, even if some documents have been published there is a lack of clear guidance for the methodology to follow to adequately decide when a method can be considered as valid. This situation has led to the availability of several methodological approaches and it is therefore the responsibility of the analyst to choose the best one. The classical decision processes encountered during method validation evaluation are compared, namely the descriptive, difference and equivalence approaches. Furthermore a validation approach using accuracy profile computed by means of  $\beta$ -expectation tolerance interval and total measurement error is also available. In the present paper all of these different validation approaches were applied to the validation of two analytical methods. The evaluation of the producer and consumer risks by Monte Carlo simulations were also made in order to compare the appropriateness of these various approaches. The classical methodologies give rise to inadequate and contradictory conclusions which do not allow them to answer adequately the objective of method validation, i.e. to give enough guarantees that each of the future results that will be generated by the method during routine use will be close enough to the true value. It is found that the validation methodology which gives the most guarantees with regards to the reliability or adequacy of the decision to consider a method as valid is the one based on the use of the accuracy profile.

© 2009 Elsevier B.V. All rights reserved.

### 1. Introduction

Before routine analysis, validation of each analytical method in all analytical laboratories of the chemical, pharmaceutical, bio-pharmaceutical or agro-food industries is a mandatory step [1–5]. However, there is a lack of clear methodological guidance in order to help the analyst to adequately decide that his method can be considered as valid, even in the regulatory documents [6,7]. Different industrial or scientist groups have proposed practical guidelines

for analytical method validation which propose decision methodologies [8–15]. Room for interpretation is left leading to diverse interpretations of these documents and hence to various decision processes. Furthermore adequacy of these methodologies with the objective of method validation is not always checked. Before going further on the discussion about the weaknesses or interests of the different validation methodologies available, it seems important to us to remind what are the objective of a quantitative analytical method and of its validation. Every analyst and regulatory agents together with the client or consumer concerned by the result of an analytical method will obviously agree that the objective of a quantitative analytical method is to be able to quantify as accurately as possible each of the unknown quantities that the laboratory will have to determine [12–15]. This means that what is expected from an analytical procedure is that the difference between the result ( $x_i$ ) and the unknown “true value” ( $\mu_T$ ) of the test sample be small

\* Corresponding author at: Laboratory of Analytical Chemistry, Institute of Pharmacy, Université de Liège, CHU, B 36, B-4000 Liège, Belgium. Tel.: +32 4 3664316; fax: +32 4 3664317.

E-mail address: [Ph.Hubert@ulg.ac.be](mailto:Ph.Hubert@ulg.ac.be) (Ph. Hubert).

<sup>1</sup> Both authors contributed equally to this work.

or inferior to a predefined acceptance limit  $\lambda$ . This latter limit can be set according to regulatory or end user's requirements. If we agree with this definition of the objective of an analytical method, it becomes thus evident that the one of its validation is to give guarantees as soon as possible that each of the future results that will be generated by the method during routine use will remain close enough to the true value. Therefore the appropriate validation methodology should give this guarantee in order to reduce simultaneously the producer and consumer risks.

The classical decision processes encountered for method validation either in the pharmaceutical industry or in the scientific literature can be classified into three groups:

- the descriptive approach which uses only point estimates of statistical parameters [6,8],
- the difference approach which uses two sided hypothesis statistical tests [8,9],
- the equivalence approach which compares parameters confidence intervals to acceptance limits [8,16–19].

Beside these classical approaches, an original one based on accuracy profile using tolerance intervals and total measurement error as single statistical decision tool has been introduced [12–15,20–22]. Facing the availability of these different approaches, it is the duty of the analyst to choose the most appropriate approach.

The aim of this paper is therefore to compare those approaches in order to stress their pros and cons, compare the risk encountered when declaring valid or not an analytical method and to check their adequacy with the aim of method validation. To achieve this objective all these approaches were applied to the validation of two quantitative analytical LC–UV methods. These quality controls methods are aimed at evaluating the conformity of three active ingredients, loratadine, paracetamol and codeine in two pharmaceutical formulations. In addition, the producer risk (i.e. the risk not to declare valid a method when it should be) and the consumer risk (i.e. the risk to declare erroneously a method as valid) of each of the validation methodologies were evaluated by Monte Carlo simulations.

## 2. Experimental

### 2.1. Classical validation methodologies

All the classical validation methodologies start by the evaluation of the response function (e.g. [8–11]). The steps commonly performed are depicted in the flowchart of Fig. 1. By means of the responses or signals obtained for each calibration standards a regression model is fitted, in general a simple linear curve using the least square regression model ( $y = ax + b$ ) for methods with UV detection as it is the case in this work. Then four main steps are most often carried out, the first one is the evaluation of the homoscedasticity of the variances through all the concentration levels of the calibration standards. If variance heterogeneity is observed a solution is to fit a weighted linear model. The following step is the lack of fit evaluation, if this test is significant, the linear model is considered not adequate and another type of regression model should be tested (e.g. a model with mathematical transformations or a quadratic model). Then a test of significance of the slope (or alternatively a test of model significance if a more complex model than the linear regression is used such as a quadratic model) and of the intercept are performed. If the slope is found not significant then another regression model should be evaluated. Finally if the intercept is significantly different from 0, the adequate regression model is the simple linear regression model, whereas if not, a single point

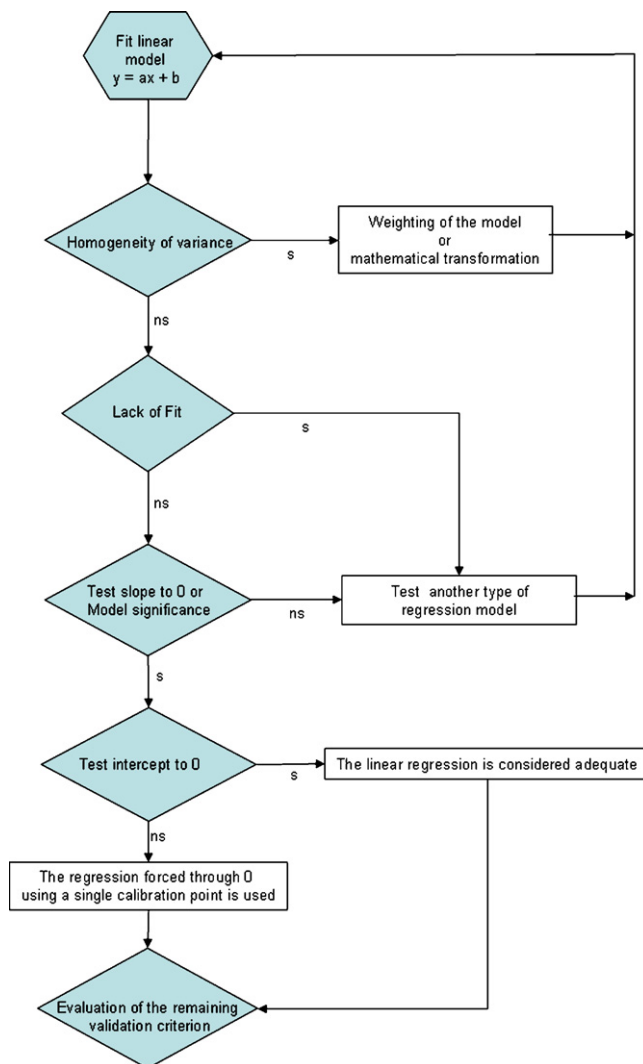


Fig. 1. Flowchart of the steps used to evaluate the response function using the classical validation methodology (s: statistically significant; ns: not significant).

calibration curve going through 0 can be considered as the adequate response function ( $y = ax$ ).

The next steps are the evaluation of the remaining validation criterion where three subcategories of validation methodologies can be made based on the decision process used. All of them can be characterized by the fact that they evaluate and compare separately trueness and precision, i.e. the systematic and random errors of the method, respectively.

#### 2.1.1. Descriptive approach

For this methodology, the point estimates of the relative bias and intermediate precision relative standard deviation are computed at each concentration level of the validation standards and are compared to a priori settled acceptance limits. The most common acceptance limits encountered in the pharmaceutical industries are:

- $\pm 2\%$  for the relative bias
- $3\%$  for the  $RSD_{ip}$ .

This means that if the estimated bias is included in the interval  $[-2\%, +2\%]$ , the trueness of the method for the concentration level studied can be accepted. For the precision criterion if the

RSD<sub>IP</sub> is smaller than 3% at the concentration level studied then the precision of the method can be accepted.

#### 2.1.2. Difference approach

The difference methodology is performed by using statistical hypothesis tests which are made of two mutually excluded hypotheses namely the null hypothesis  $H_0$  and the alternative hypothesis  $H_1$ . The trueness criterion (evaluation of the method bias) is evaluated by comparison of a 95% confidence interval of the overall average bias (or average recovery) to the 0% relative bias value (or equivalently the 100% recovery value). If this interval contains the 0% bias (or the 100% recovery) the trueness of the method under investigation is accepted, else it should be rejected.

For the precision criterion, one has to observe whether or not the maximum acceptable intermediate precision relative standard deviation (RSD<sub>IP</sub>) value is included inside the 95% confidence interval of the RSD<sub>IP</sub> [23,24] and this for each concentration level of the validation standards. For instance, in the present paper the maximum acceptable value for RSD<sub>IP</sub> is set at 3% for each analytical method. Therefore, if the 3% value is included in this interval the precision criterion is accepted; if not, it is rejected.

#### 2.1.3. Equivalence approach

The equivalence approach differs from the previous one in that, rather than looking if the estimated relative bias or estimated RSD<sub>IP</sub> are included in the acceptance limits, it looks at whether the true value of these parameters are included in their respective acceptance limits or not, this for each concentration level of the validation standards studied. To achieve this for the trueness criterion, the 90% two sided confidence interval of the relative bias is computed for each concentration level and compared to the  $\pm 2\%$  acceptance limits. For the precision criteria, it is the upper limit of the 95% confidence interval of the RSD<sub>IP</sub> [23,24] that is compared to the 3% limit which must be smaller than this maximum value.

### 2.2. Accuracy profile methodology

The accuracy profile methodology is based on the concept of total measurement error, i.e. the simultaneous combination of the systematic error (measured by biases, i.e. method trueness) and of the random error (measured by RSDs, i.e. method precision) [12–15]. The criterion related to total error is the result accuracy. What is needed is to give at the end of the validation step guarantees that each of the future results that the laboratory will obtain during routine analyses will be accurate enough. Therefore to reach this objective, instead of a whole set of statistical tests, the accuracy profile methodology uses only one statistical decision methodology, namely a  $\beta$ -expectation tolerance interval computed at each concentration level of the validation standards [13–15]. This interval represents a region where, on average, it is expected to find a defined proportion of the population of the results, i.e.  $\beta$ . When applied at different concentration levels an accuracy profile can be drawn [13–15,20]. This profile is then compared to “a priori” settled acceptance limits, which are set to  $\pm 5\%$  in the following of this paper. Indeed, this value is a conventional limit used in the framework of the evaluation of the conformity of active ingredients in pharmaceutical formulations.

### 2.3. Chemicals and solvents

#### 2.3.1. LC determination of loratadine

All chemicals and solvents used were of analytical or HPLC grade. Loratadine was supplied by the European Pharmacopoeia (Strasbourg, France). Potassium dihydrogen phosphate was purchased from Acros Organics (Geel, Belgium). Methanol, acetonitrile and phosphoric acid (85%) were obtained from Merck (Darmstadt,

Germany). Deionized water was generated from Milli-Q water purifying system (Millipore, Watford, UK). A phosphate buffer at pH 3.0 was prepared by dissolving 2.7 g of potassium dihydrogen phosphate in 1.0 L of deionized water. The pH was adjusted to 3.0 with phosphoric acid.

#### 2.3.2. LC determination of paracetamol and codeine

All chemicals and solvents used were of analytical or HPLC grade. Paracetamol and codeine phosphate were supplied by the European Pharmacopoeia (Strasbourg, France). Potassium dihydrogen phosphate was purchased from Acros Organics. Sodium octanesulfonate was purchased from Sigma–Aldrich (Steinheim, Germany). Methanol and phosphoric acid (85%) were obtained from Merck. Deionized water was generated from Milli-Q water purifying system (Millipore). An aqueous phosphate solution at pH 3.5 was prepared by dissolving 1.4 g of potassium dihydrogen phosphate in 1.0 L of deionized water. The pH was adjusted to 3.5 with phosphoric acid.

### 2.4. Apparatus

#### 2.4.1. LC determination of loratadine

The HPLC system was an HP 1100 (Agilent, Palo-Alto, CA, USA) composed of a quaternary pump G1311A, an autosampler G1313A, an oven G1316A an online degasser G1322A and a DAD detector G1315A.

#### 2.4.2. LC determination of paracetamol and codeine

The HPLC system consisted in a LaChrom (Merck–Hitachi, Darmstadt, Allemagne) composed of a quaternary pump L-7100, an autosampler L-7200, an oven L-7360 and a DAD detector L-7455.

### 2.5. Chromatographic conditions

#### 2.5.1. LC determination of loratadine

The mobile phase consisted in a mixture of acetonitrile and aqueous phosphate buffer at pH 3.0 (58:42, v/v). The separation was made in isocratic mode with a LiChroCart PuroSpher RP-18e column (125 × 4 mm I.D.—5  $\mu$ m; Merck) thermostatised at 30 °C. The flow rate was 1 mL/min, the injection volume was 25  $\mu$ L and the UV detection was performed at 245 nm.

#### 2.5.2. LC determination of paracetamol and codeine

The mobile phase consisted in a mixture of methanol and aqueous phosphate buffer at pH 3.5 with 1.73 g/L of octanesulfonate (27:73, v/v). The separation was made in isocratic mode with a PuroSpher RP-18e column (55 × 4 mm I.D.—3  $\mu$ m; Merck) thermostatised at 35 °C. The flow rate was 1 mL/min, the injection volume was 10  $\mu$ L and the UV detection was performed at 296 nm and 214 nm for paracetamol and codeine, respectively.

### 2.6. Standard solutions

#### 2.6.1. LC determination of loratadine

A stock solution of loratadine was prepared by accurately weighting 25.0 mg of loratadine and diluting this in 100.0 mL of methanol. The calibration standards for loratadine were prepared by diluting the stock solution with its mobile phase to reach five concentration levels ranging from 12.5 to 37.5  $\mu$ g/mL. The calibration standards were repeated once for three series of analyses. The validation standards were prepared in reconstituted solutions to reach five concentration levels: 12.5, 20, 25, 30 and 37.5  $\mu$ g/mL. The validation standards at 12.5, 25 and 37.5  $\mu$ g/mL were repeated six times whereas the others were analyzed in triplicates. The validation standards were analyzed during three series.

**Table 1**

Results of the evaluation of the response function used by the classical validation methodologies for the LC–UV methods dedicated to the quantification of loratadine and paracetamol together with codeine. Significant values ( $\alpha = 0.05$ ) are in bold and italic.  $C_{\text{obs}}$  and  $C_{\text{theo}}$  are the observed and critical values of the Cochran tests for homogeneity of variance, respectively.

Sample types	Statistical tests	Loratadine	Analyte paracetamol	Codeine
Calibration standards	Model significance	<b><i>p &lt; 0.0001</i></b>	<b><i>p &lt; 0.0001</i></b>	<b><i>p &lt; 0.0001</i></b>
	Lack of fit	<i>p = 0.436</i>	<i>p = 0.151</i>	<i>p = 0.147</i>
	Homogeneity of variances	$C_{\text{obs}} = 0.46 < C_{\text{theo}} = 0.60$	$C_{\text{obs}} = 0.56 < C_{\text{theo}} = 0.87$	$C_{\text{obs}} = 0.86 < C_{\text{theo}} = 0.87$
	Comparison of intercepts to 0	<i>p = 0.371</i>	<b><i>p = 0.0013</i></b>	<i>p = 0.817</i>
	Comparison of slopes to 0	<b><i>p &lt; 0.0001</i></b>	<b><i>p &lt; 0.0001</i></b>	<b><i>p &lt; 0.0001</i></b>

### 2.6.2. LC determination of paracetamol and codeine

A stock solution of paracetamol and codeine was prepared by accurately weighting 500.0 mg of paracetamol and 30.0 mg of codeine phosphate and diluting them in 100.0 mL of methanol. The calibration standards for paracetamol and codeine were prepared by diluting the stock solution with their pH 3.5 aqueous phosphate solution to reach three concentration levels for both analytes: 200, 400 and 600  $\mu\text{g/mL}$  for paracetamol and 20, 25 and 30  $\mu\text{g/mL}$  for codeine. The calibration standards were repeated once for three series. The validation standards were prepared in reconstituted solutions to reach the same three concentration levels for both active substances. They were all analyzed in triplicates during three series.

### 2.7. Simulations

Simulations were performed in order to estimate the performances of the different decision approaches for method validation in terms of producer and consumer risks.

To achieve this, independent validation results were generated from the ANOVA I model described below:

$$X_{ij} = \delta + \phi_B + \varepsilon_W$$

where  $X_{ij}$  is the result of the  $j$ th measurement in series  $i$ ,  $\delta = \mu_{\text{Lab}} - \mu_T$  is the bias between the true value of the measurand ( $\mu_T$ ) and the average value of the results of the laboratory ( $\mu_{\text{Lab}}$ ),  $\phi_B$  is the between series random error effect supposed to be independent and  $N(0, \sigma_B^2)$  and  $\varepsilon_W$  is the within-series (or repeatability) random error effect supposed to be independent and  $N(0, \sigma_W^2)$ .

$\mu_T$ , the true value was fixed at 100. Simulations were performed for analytical methods with biases ranging from  $-5$  to  $+5\%$  and for intermediate precision relative standard deviation ranging from 0% to 5%. The ratios  $R$  corresponding to the series-to-series variance and the within-series (or repeatability) variance was settled at  $R = 1$ . The number of replicates was fixed at  $J = 3$  and the number of series  $i$  used were 3 and 10. For each situation 2000 simulations were performed.

The various acceptance limits were fixed as in Sections 2.1 and 2.2 of this work: a maximum acceptable relative bias of  $\pm 2\%$ , a maximum acceptable  $\text{RSD}_{\text{IP}}$  of 3%, and a maximum total error of  $\pm 5\%$ . Furthermore the minimum proportion of results to be included in this last acceptance limits was set at  $\beta = 95\%$ , i.e. we want to have guarantees that, on average, 95% of the future results obtained by the laboratory will be within  $\pm 5\%$  of the true value. The region of truly valid methods is also computed and defined as the region where at least 95% of the results are included within the  $\pm 5\%$  limits.

### 2.8. Computations

Data treatments were realized using the AVA software (Qualilab, Olivet, France) for the classical validation methodologies and e-noval V2.0 (Arlenda, Liege, Belgium) for the accuracy profile approach. All simulations were performed with R v2.2.1 (CRAN, <http://cran.r-project.org>).

## 3. Results

### 3.1. LC determination of loratadine

#### 3.1.1. Classical validation methodologies

When following the classical validation methodology for the assay dedicated to the LC determination of loratadine in tablets, it can be seen in Table 1 that the adequate response function is a linear regression forced through zero with the highest concentration level as single calibration point. Indeed, the lack of fit test and the test of the significance of the intercept are non-significant. Furthermore, the test of homogeneity of variance is also non-significant. Considering this strategy, the response function that should provide accurate results would be the linear regression through zero. Nevertheless, the trueness and precision criteria were also estimated using the full linear regression for the sake of the critical comparison of the validation strategies and reported in Table 2.

When assessing the trueness of the analytical method using the difference validation approach (Section 2.1.2) and based on the previously selected response function, the confidence interval of the average relative bias does not include the 0% bias value as shown in Table 2. Therefore, the conclusion is that the method cannot be considered as valid despite the fact that the first part of the methodology ensured that the selected response function was the adequate one. When using the first classical validation approach (Section 2.1.1) the estimated relative bias at each validation standards concentration level of the analytical method is compared to the  $\pm 2\%$  acceptance limits. As can be seen in Table 2, the method can be accepted as valid as each estimated relative bias is included in the interval  $[-2\%, +2\%]$  irrespective of the concentration level. For the third classical validation approach (Section 2.1.3) it is the 90% confidence interval of the relative bias at each concentration level that is compared to the  $\pm 2\%$  acceptance limits. Here also as seen in Table 2 the method can be considered as valid for this criterion as all the confidence intervals are included in the interval  $[-2\%, +2\%]$ .

When evaluating the precision criterion, the difference approach for precision rejects the validity of the analytical method for three concentration levels: 20, 25 and 37.5  $\mu\text{g/mL}$  while, the use of the descriptive validation approach shows that the method can be considered acceptable for the precision criterion. Indeed for this last methodology, the estimated  $\text{RSD}_{\text{IP}}$  are smaller than the 3% acceptance limit for every concentration level studied. However the third approach (equivalence) will reject the validity of the analytical method as the upper confidence limit of the  $\text{RSD}_{\text{IP}}$  exceeds this acceptance limits for the lowest concentration level (12.5  $\mu\text{g/mL}$ ) and is equal to 3.0% for the fourth concentration level (30  $\mu\text{g/mL}$ ) as shown in Table 2.

#### 3.1.2. Accuracy profile validation methodology

By comparison, Fig. 2 gives accuracy profiles obtained with the same data but when applying the accuracy profile approach for the validation of the analytical method. Two calibration functions were tested, namely the linear regression through zero fitted using the highest calibration level and the simple linear regression models. As shown by this figure, both calibration models tested are acceptable

**Table 2**  
Results of the validation of the LC–UV method for the quantification of loratadine in tablets using the classical and accuracy profile methodologies with either the linear regression model through 0 using the highest calibration standard or the simple linear model as calibration curves.

	Loratadine					
	Linear regression through 0 using the highest level			Simple linear regression		
	Slope	Intercept	r <sup>2</sup>	Slope	Intercept	r <sup>2</sup>
<b>Response functions</b>						
Day 1	1.9751E+04	na	na	1.9617E+04	5263	0.999
Day 2	1.9826E+04	na	na	1.9746E+04	2159	0.999
Day 3	1.9966E+04	na	na	2.0002E+04	–1441	0.999
<b>Difference approach</b>						
Trueness (95% confidence of the mean bias (%))	<b>[–0.49, –0.13]</b>			<b>[–0.64, –0.34]</b>		
Precision (95% upper limit confidence interval of RSD <sub>IP</sub> (%))	12.5 µg/mL		3.7	12.5 (g/mL)		<b>1.6</b>
	20 (g/mL)		<b>2.8</b>	20 (g/mL)		<b>1.9</b>
	25 (g/mL)		<b>1.7</b>	25 (g/mL)		<b>0.9</b>
	30 (g/mL)		3.0	30 (g/mL)		3.2
	37.5 (g/mL)		<b>1.6</b>	37.5 (g/mL)		<b>1.6</b>
<b>Descriptive approach</b>						
Trueness (relative bias (%))	12.5 (g/mL)		0	12.5 (g/mL)		–0.6
	20 (g/mL)		0	20 (g/mL)		–0.2
	25 (g/mL)		–0.5	25 (g/mL)		–0.6
	30 (g/mL)		–0.4	30 (g/mL)		–0.4
	37.5 (g/mL)		–0.5	37.5 (g/mL)		–0.5
Precision (RSD <sub>Rep</sub> /RSD <sub>IP</sub> (%))	12.5 (g/mL)		0.6/1	12.5 (g/mL)		0.6/0.6
	20 (g/mL)		0.3/0.7	20 (g/mL)		0.3/0.5
	25 (g/mL)		0.7/0.7	25 (g/mL)		0.7/0.7
	30 (g/mL)		0.5/0.8	30 (g/mL)		0.5/0.8
	37.5 (g/mL)		0.6/0.7	37.5 (g/mL)		0.6/0.7
<b>Equivalence approach</b>						
Trueness (90% confidence interval of bias (%))	12.5 (g/mL)		[–0.8, 0.7]	12.5 (g/mL)		[–0.9, 0.2]
	20 (g/mL)		[–0.6, 0.7]	20 (g/mL)		[–0.5, 0.2]
	25 (g/mL)		[–1.1, 0.1]	25 (g/mL)		[–1.0, –0.3]
	30 (g/mL)		[–12, 0.5]	30 (g/mL)		[–10, 0.2]
	37.5 (g/mL)		[–1.0, 0.0]	37.5 (g/mL)		[–0.8, –0.2]
Precision (95% upper limit confidence interval of RSD <sub>IP</sub> (%))	12.5 (g/mL)		<b>3.7</b>	12.5 (g/mL)		1.6
	20 (g/mL)		2.8	20 (g/mL)		1.9
	25 (g/mL)		1.7	25 (g/mL)		0.9
	30 (g/mL)		<b>3.0</b>	30 (g/mL)		<b>3.2</b>
	37.5 (g/mL)		1.6	37.5 (g/mL)		1.6
<b>Accuracy profile approach</b>						
Accuracy (95% β-expectation tolerance interval (%))	12.5 (g/mL)		[–3.1, 3.0]	12.5 (g/mL)		[–2.0, 0.9]
	20 (g/mL)		[–2.7, 2.7]	20 (g/mL)		[–1.8, 1.5]
	25 (g/mL)		[–2.2, 1.1]	25 (g/mL)		[–2.1, 0.8]
	30 (g/mL)		[–2.9, 2.2]	30 (g/mL)		[–3.1, 2.3]
	37.5 (g/mL)		[–2.0, 0.9]	37.5 (g/mL)		[–2.0, 1.0]
<b>Results linearity</b>						
Slope	0.9921			0.9951		
Intercept	0.1013			–2.03E – 05		
Range	[12.5, 37.5]			[12.5, 37.5]		
r <sup>2</sup>	0.9996			0.9996		
LOQ ((g/mL)	12.5			12.5		
LOD ((g/mL)	4.2			4.2		

since each accuracy profile is fully included into the acceptance limits settled at ±5% for all the validation standards concentration levels. As these profiles were built with tolerance intervals at β=95%, it is expected that 95 out of 100 future results obtained with this method during its routine use will be situated on average in this interval for the whole concentration range studied.

Therefore, the method can be considered as valid either by using the simple linear regression or the forced through zero one. It is this last one that will be used for routine analysis, as it is easier and more cost effective. Table 2 gives all the validation results obtained when using the two tested calibration functions.

### 3.2. LC determination of paracetamol and codeine

#### 3.2.1. Paracetamol

3.2.1.1. *Classical validation methodologies.* For the validation of the method used for the determination of paracetamol, the lack of fit

test is not significant indicating that the linear regression is an adequate model. Also, the slope and the intercept are significant (Table 1). Consequently, the response function that is indicated as suitable by the classical validation approach is the simple linear regression model. In this case also, the estimation of trueness and precision were nonetheless made using the linear regression forced through 0 and the highest concentration level as well as with a weighted (1/X<sup>2</sup>) linear regression model. The results for method trueness using this formerly selected calibration curve show however that the method cannot be considered as valid by the difference approach. Indeed the confidence interval of the average relative bias does not include the 0% value as shown in Table 3. This is again in contradiction with the conclusion of the evaluation of the response function. The first classical validation approach (Section 2.1.1) which compares the estimated relative bias at each validation standards concentration levels of the analytical method to the ±2% acceptance limits shows that the method is valid for all the con-

**Table 3**  
Results of the validation of the LC–UV method for the quantification of paracetamol in effervescent tablets using the classical and accuracy profile methodologies with either the linear regression model through 0 using the highest calibration standard, the simple linear model or the weighted ( $1/X^2$ ) linear model as calibration curves.

		Paracetamol					
		Linear regression through 0 using the highest level			Weighted ( $1/X^2$ ) linear regression		
		Slope	Intercept	$r^2$	Slope	Intercept	$r^2$
<b>Response functions</b>							
Day 1		1060	na	na	1030	1.97E+04	0.999
Day 2		1061	na	na	1030	2.07E+04	0.999
Day 3		1066	na	na	1052	8474	0.999
<b>Difference approach</b>							
Trueness (95% confidence of the mean bias (%))		<b>[1.05, 2.58]</b>					
Precision (95% upper limit confidence interval of RSD <sub>IP</sub> (%))		200 (g/mL)		3.7	<b>[-0.98, -0.09]</b>		6.9
		400 (g/mL)		2.7	200 (g/mL)		3.4
		600 (g/mL)		<b>1.8</b>	400 (g/mL)		<b>1.1</b>
					600 (g/mL)		
<b>Descriptive approaches</b>							
Trueness (relative bias (%))		200 (g/mL)		<b>4.1</b>	200 (g/mL)		-1.2
		400 (g/mL)		1.3	400 (g/mL)		-0.1
		600 (g/mL)		0	600 (g/mL)		-0.2
Precision (RSD <sub>rep</sub> /RSD <sub>IP</sub> (%))		200 (g/mL)		0.9/1.1	200 (g/mL)		0.9/1.8
		400 (g/mL)		0.7/0.8	400 (g/mL)		0.7/1.0
		600 (g/mL)		0.3/0.5	600 (g/mL)		0.3/0.3
<b>Equivalence approaches</b>							
Trueness (90% confidence interval of bias (%))		200 (g/mL)		<b>[3.3, 4.9]</b>	200 (g/mL)		<b>[-2.4, -0.0]</b>
		400 (g/mL)		[0.7, 1.9]	400 (g/mL)		[-0.8, 0.6]
		600 (g/mL)		[-0.4, 0.3]	600 (g/mL)		[-0.5, 0.0]
Precision (95% upper limit confidence interval of RSD <sub>IP</sub> (%))		200 (g/mL)		<b>3.7</b>	200 (g/mL)		<b>6.9</b>
		400 (g/mL)		2.7	400 (g/mL)		<b>3.4</b>
		600 (g/mL)		1.8	600 (g/mL)		1.1
<b>Accuracy profile approach</b>							
Accuracy (relative 95% $\beta$ -expectation tolerance interval (%))		200 (g/mL)		[1.1, 7.2]	200 (g/mL)		[-7.0, 5.0]
		400 (g/mL)		[-0.9, 3.6]	400 (g/mL)		[-3.0, 2.7]
		600 (g/mL)		[-1.6, 1.5]	600 (g/mL)		[-1.2, 0.7]
<b>Results linearity</b>							
Slope			0.9787			1.002	0.9969
Intercept			12.98			-2.446	-0.3892
Range			[200, 600]			[200, 600]	[200, 600]
$r^2$			0.9997			0.9997	0.9997
LOQ ((g/mL)			319.8			309.3	200
LOD ((g/mL)			106.6			103.1	66.7



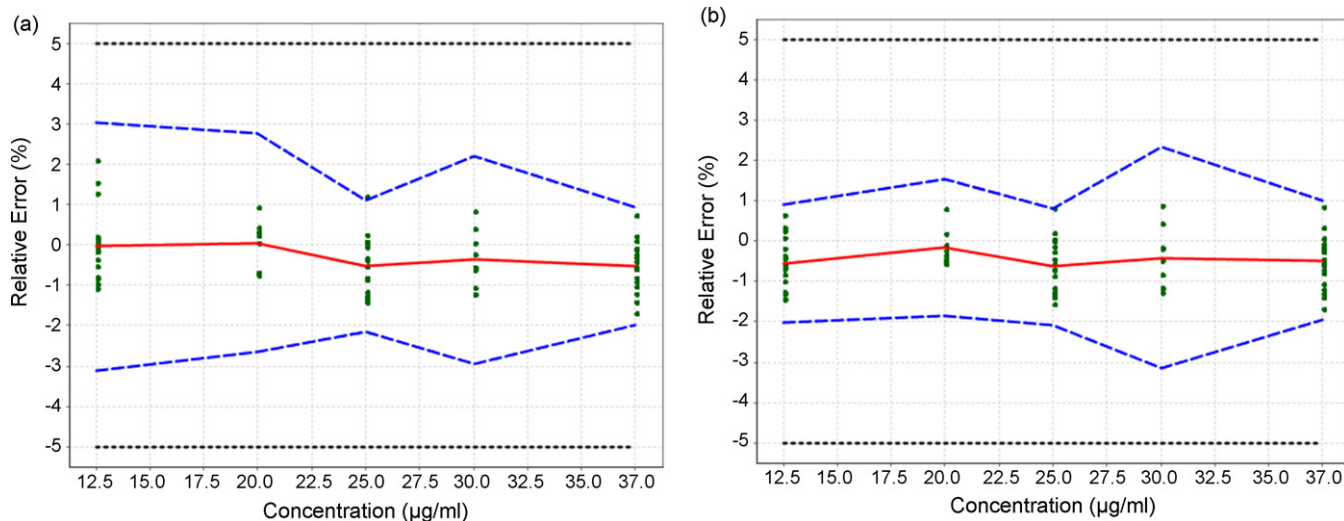


Fig. 2. Loratadine. Accuracy profile obtained for the validation of the LC–UV analytical method for the quantification of loratadine by considering (a) the linear regression through 0 using the highest calibration level and (b) the simple linear regression as calibration curves; plain line: relative bias, dashed lines:  $\beta$ -expectation tolerance limits, dotted curves: acceptance limit (%) and dots: relative back-calculated concentrations of the validation standards.

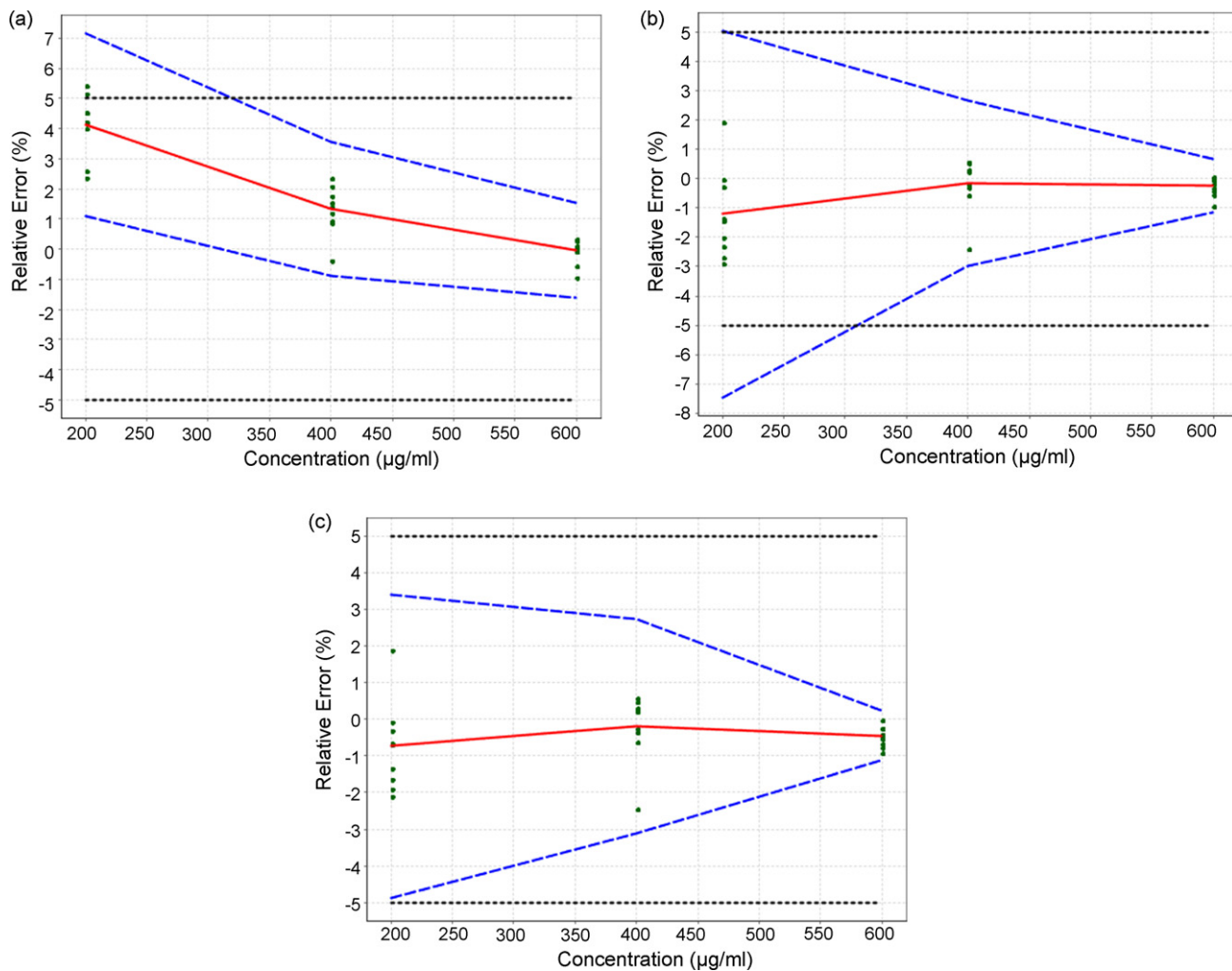


Fig. 3. Paracetamol. Accuracy profile obtained for the validation of the LC–UV analytical method for the quantification of paracetamol by considering (a) the linear regression through 0 using the highest calibration level, (b) the simple linear regression and (c) the weighted ( $1/X^2$ ) linear regression as calibration curves; plain line: relative bias, dashed lines:  $\beta$ -expectation tolerance limits, dotted curves: acceptance limit (%) and dots: relative back-calculated concentrations of the validation standards.

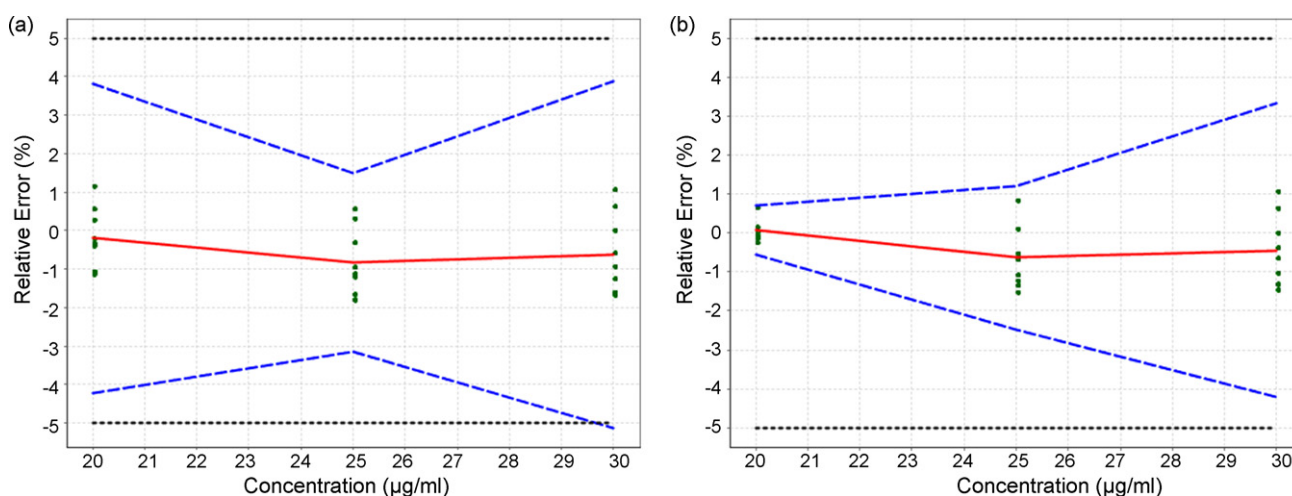
**Table 4**

Results of the validation of the LC–UV method for the quantification of codeine in effervescent tablets using the classical and accuracy profile methodologies with either the linear regression model through 0 using the highest calibration standard or the simple linear model as calibration curves.

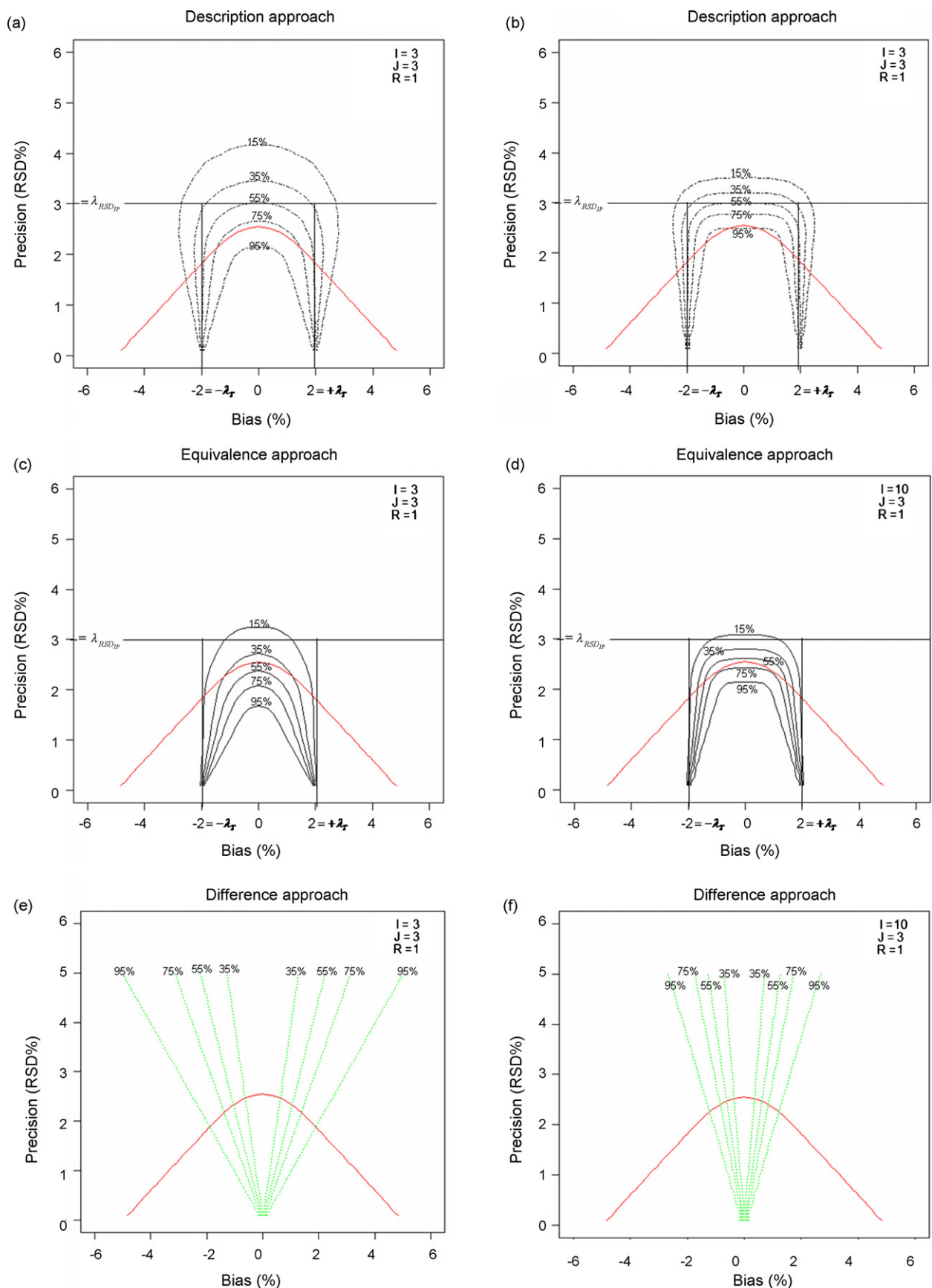
	Codeine					
	Linear regression through 0 using the highest level			Simple linear regression		
	Slope	Intercept	$r^2$	Slope	Intercept	$r^2$
Response functions						
Day 1	1.99E+04	na	na	1.99E+00	–1636	0.999
Day 2	1.98E+04	na	na	1.96E+04	6185	0.999
Day 3	2.00E+04	na	na	2.03E+04	–8643	0.999
Difference approach						
Trueness (95% confidence of the mean bias (%))	[–0.91, –0.19]			[–0.63, –0.04]		
Precision (95% upper limit confidence interval of RSD <sub>IP</sub> (%))						
20 (g/mL)			3.9	20 (g/mL)		<b>0.4</b>
25 (g/mL)			<b>2.8</b>	25 (g/mL)		<b>1.3</b>
30 (g/mL)			4.7	30 (g/mL)		4.1
Descriptive approaches						
Trueness (relative bias (%))						
20 (g/mL)			–0.2	20 (g/mL)		0.1
25 (g/mL)			–0.8	25 (g/mL)		–0.6
30 (g/mL)			–0.6	30 (g/mL)		–0.4
Precision (RSD <sub>Rep</sub> /RSD <sub>IP</sub> (%))						
20 (g/mL)			0.3/0.9	20 (g/mL)		0.3/0.3
25 (g/mL)			0.8/0.9	25 (g/mL)		0.8/0.8
30 (g/mL)			0.5/1.1	30 (g/mL)		0.5/1.0
Equivalence approaches						
Trueness (90% confidence interval of bias (%))						
20 (g/mL)			[–0.8, 0.4]	20 (g/mL)		[–0.1, 0.3]
25 (g/mL)			[–1.5, –0.2]	25 (g/mL)		[–1.2, –0.1]
30 (g/mL)			[–1.4, 0.2]	30 (g/mL)		[–12, 0.3]
Precision (95% Upper limit confidence interval of RSD <sub>IP</sub> (%))						
20 (g/mL)			<b>3.9</b>	20 (g/mL)		0.4
25 (g/mL)			2.8	25 (g/mL)		1.3
30 (g/mL)			<b>4.7</b>	30 (g/mL)		<b>4.1</b>
Accuracy profile approach						
Accuracy (95% $\beta$ -expectation tolerance interval (%))						
20 (g/mL)			[–4.2, 3.8]	20 (g/mL)		[–0.6, 0.7]
25 (g/mL)			[–3.1, 1.5]	25 (g/mL)		[–2.5, 1.2]
30 (g/mL)			[–5.1, 3.9]	30 (g/mL)		[–4.2, 3.3]
Results linearity						
Slope	0.9855			0.9850		
Intercept	0.2179			0.2813		
Range	[20, 30]			[20, 30]		
$r^2$	0.9968			0.9978		
LOQ ((g/mL))	20			20		
LOD ((g/mL))	6.7			6.7		

centration level tested (Table 3). As shown in Table 3 when using the third classical validation approach (Section 2.1.3) the method cannot be considered as valid for this criterion as the confidence interval of the lowest concentration level studied (200  $\mu\text{g/mL}$ ) is not included in the interval [–2%, +2%].

The difference approach for precision will reject the validity of the method for the highest concentration level (600  $\mu\text{g/mL}$ ) since the maximum tolerated 3% RSD<sub>IP</sub> value is not smaller than the 95% upper confidence limit of the RSD<sub>IP</sub>. For the precision criterion the descriptive validation approach shows that the method seems valid



**Fig. 4.** Codeine. Accuracy profile obtained for the validation of the LC–UV analytical method for the quantification of codeine by considering (a) the linear regression through 0 using the highest calibration level and (b) the simple linear regression as calibration curves; plain line: relative bias, dashed lines:  $\beta$ -expectation tolerance limits, dotted curves: acceptance limit (%) and dots: relative back-calculated concentrations of the validation standards.



**Fig. 5.** Results of the simulations using the descriptive (a, b), equivalence (c, d), difference (e, f) and accuracy profile (g, h) approaches. Iso-probability curves for each approach are depicted, i.e. the methods on these curves have the same probability to be declared as valid (with respect to their respective acceptance limits).  $i$  is the number of series,  $J$  the number of replicates and  $R$  the ratio of the between run variance over the within run variance.  $[-\lambda_T; +\lambda_T]$  is the acceptance limits for trueness and  $\lambda_{RSD_{IP}}$  is the acceptance limit for precision. The continuous bell-shaped curve present in all figures delimits the region of methods truly providing 95% of their results in the  $\pm 5\%$  acceptance limits for total error (or accuracy).

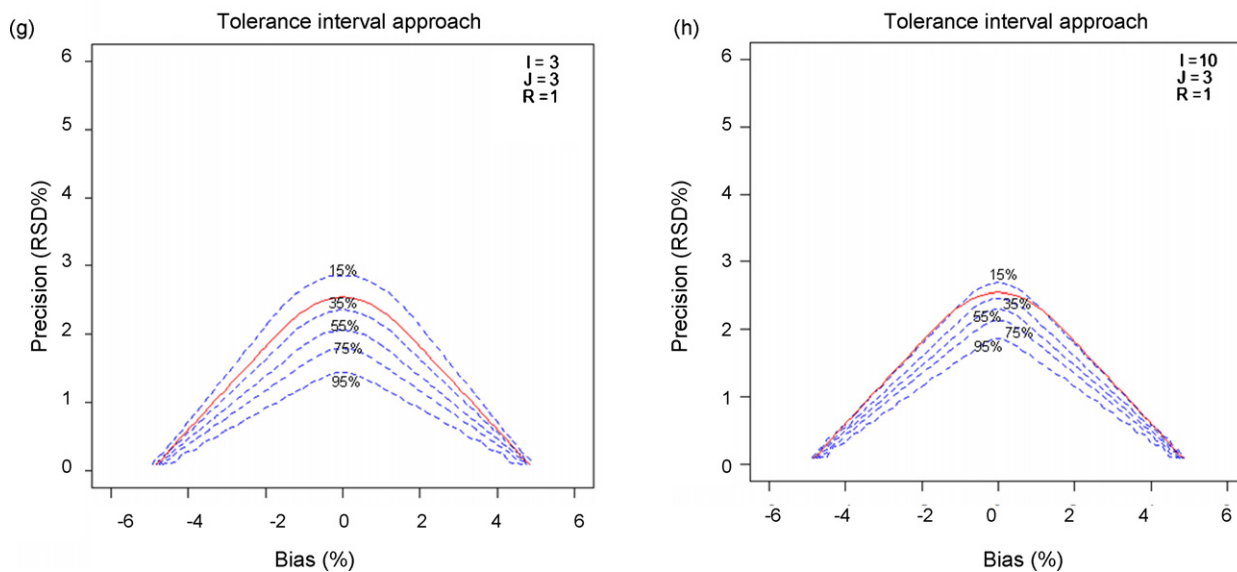


Fig. 5. (Continued).

as the estimated  $RSD_{IP}$  are smaller than the 3% acceptance limit for all the concentration levels studied. However the third approach (equivalence) will reject the validity of the analytical method as the upper confidence limit of the  $RSD_{IP}$  exceed this acceptance limits for the lowest (200  $\mu\text{g/mL}$ ) and the middle concentration levels (400  $\mu\text{g/mL}$ ) as shown in Table 3.

**3.2.1.2. Accuracy profile validation methodology.** Application of the accuracy profile methodology for the validation of the paracetamol assay is shown in Fig. 3, where three accuracy profiles are shown build with three different calibration functions: the linear through 0 using only the highest level of the calibration standards, the simple linear and the weighted  $1/X^2$  linear regression. As can be seen in Fig. 3 and by opposite to the classical validation methodology, the linear regression is not an adequate calibration function as the corresponding profile is not totally included in the acceptance limits. Similarly, the linear regression through 0 is also not fit for the objective of the analytical method as the profile is stepping outside the acceptance limits. The only acceptable calibration function is the weighted  $1/X^2$  linear regression. Indeed with this model the  $\beta$ -expectation tolerance interval is fully included in the acceptance limits irrespective to the validation standard concentration level. The results obtained with the accuracy profile methodology for the validation of the assay dedicated to the determination of paracetamol in effervescent tablets are shown in Table 3 for the three types of calibration function tested.

### 3.2.2. Codeine

**3.2.2.1. Classical validation methodologies.** The classical validation methodology results considering the evaluation of the response function for the validation of the quantitative determination of codeine in effervescent tablets by HPLC–UV are given in Table 1. As can be seen from this table, the linear regression is adequate as the Lack of Fit test is not significant and the test for the slope significance is significant. Furthermore, the test of significance of the intercept is not significant. Therefore, it can be concluded that the response function that should be adequate using this methodology is the linear regression through 0 fitted with the highest calibration level. Again for the purpose of comparing the different validation approaches, another response function was nevertheless used, namely the simple linear regression. All the corresponding results are shown in Table 4.

For the evaluation of the trueness of the method using the difference validation methodology based on the calibration curve selected earlier, the confidence interval of the average relative bias does not include the 0% bias value as shown in Table 4. This therefore means that the method is not valid using the proposed calibration function which is in contradiction with the first part of the validation process. When using the descriptive classical validation approach, the estimated relative bias at each validation standards concentration level of the analytical method are included in the interval  $[-2\%, +2\%]$  for all concentration level as shown in Table 4. Thus the method can be accepted as valid for the trueness criterion using the descriptive validation methodology. For the third classical validation approach (equivalence) it is the confidence interval of the relative bias at each concentration level that is compared to the  $\pm 2\%$  acceptance limits. Here also Table 4 shows that the method can be considered as valid for this criterion as all the confidence intervals are included in the interval  $[-2\%, +2\%]$ .

The difference approach for precision rejects the validity of the method for the middle concentration level (25  $\mu\text{g/mL}$ ) since the maximum tolerated 3%  $RSD_{IP}$  value is not smaller than the 95% upper confidence limit of the  $RSD_{IP}$ . Using the descriptive validation approach the method is considered as valid for the precision criteria as all the estimated  $RSD_{IP}$  are smaller than the 3% acceptance limit (Table 4). However the third approach (equivalence) rejects the validity of the analytical method as the upper confidence limit of the  $RSD_{IP}$  exceeds this acceptance limits for the lowest concentration level (20  $\mu\text{g/mL}$ ) and the highest one (30  $\mu\text{g/mL}$ ) as shown in Table 4.

**3.2.2.2. Accuracy profile validation methodology.** However, the application of the accuracy profile approach does give other conclusions. Indeed as depicted by the accuracy profiles of Fig. 4, build either with a linear regression forced through 0 and using the highest calibration level or a simple linear regression as calibration functions, only this last model allows to compute accurately over the whole range studied. It can also be seen that only for the two first concentration level, the forced through 0 calibration curve gives accurate results. As the objective of this method is to quantify codeine over the 3 concentration levels of the validation standards tested, the only adequate response function is the simple linear one. Table 4 summarizes the validation results for the two response functions tested.

### 3.3. Simulations

Fig. 5a–h shows iso-probability curves for the various decision methodologies in the 2D plan of true bias and precision, using either 3 replicates and 3 runs or 3 replicates and 10 runs. Iso-probability curves show where the probability to accept analytical methods as valid is the same all over the curve. For example in Fig. 5a, the iso-probability curve labeled 95% show all the methods that have 95% chance to be declared valid using the descriptive approach. In all these figures are also included the region of methods that can generate 95% of their results within the  $\pm 5\%$  acceptance limits (the continuous hyperbolic curve).

For the descriptive approach, it can be seen from Fig. 5a that this approach will accept as valid methods with biases and RSD values greater than their respective acceptance limits with a probability around 35–55%. Indeed the last iso-probability curves included inside the acceptance limits are the 35 and 55% ones. Thus the risk to accept a method as valid when it should not is between 35% and 55% using such an approach, i.e. the consumer risk. When increasing the sample size, the descriptive approach does not behave better. The consumer risk estimation is rather improved and is about 55% (Fig. 5b). Indeed the last iso-probability curve included inside the acceptance limits is the one labeled 55%. Furthermore this approach does not allow to guarantee that at least 95% of the results generated by the method declared valid will be within the  $\pm 5\%$  acceptance limits for total error as most iso-probability curves lay outside the total error  $\pm 5\%$  acceptance limits.

When performing 3 runs of 3 replicates with the equivalence approach (Fig. 5c) the consumer risk is much more controlled than with the descriptive approach: it lies between 15% and 35%. Furthermore, when the sample size increases as shown in Fig. 5d, this risk is reduced to values smaller than 15%. However, this approach does not answer the objective of validation: giving guarantees that the validated method will provide accurate results. Indeed there is a non-negligible amount of methods that will be declared valid using this approach but that will provide less than 95% of their results in the  $\pm 5\%$  acceptance limits as shown by the iso-probabilities curve stepping outside the truly acceptable methods region in Fig. 5c and d. Furthermore, a great number of perfectly acceptable methods with respect to that objective are abusively discarded using this approach.

Fig. 5e and f for the difference approach is showing the peculiar behavior of this method validation approach. Method with small bias and small enough variance to allow detection of significant biases will never be accepted as valid, whereas method with high bias and variability sufficiently high to prevent the detection of biases as significant will be considered as valid. The producer risk is controlled: methods with 0% bias are accepted, whatever their precision. The consumer risk is not at all controlled: methods with high bias can have an extremely high probability to be accepted. For example methods with biases higher than 2% and with precision higher than 2% are accepted about 95% of the times as shown in Fig. 5e. Finally, it can be seen that this approach is not at all fit to answer the objective of methods validation: extremely few methods declared valid with this approach will effectively provide 95% of their results within the  $\pm 5\%$  total error acceptance limits.

The application of the accuracy profile approach using tolerance intervals shows that the consumer risk is well controlled, indeed the risk to consider a method as valid when it should not is between 35% and 15% as shown in Fig. 5g. Furthermore this risk decreases as the sample size increases: it is about 15% in Fig. 5h. The producer risk is however only partially controlled when the number of series is equal to 3. However this last risk also decreases when increasing the size of the experimental design. Finally, as depicted by Fig. 5g and h, the only approach investigated which can give enough guarantees that the method declared valid will provide

results inside the acceptance limits or specifications with a high probability is the accuracy profile using  $\beta$ -expectation tolerance intervals as statistical methodology.

### 4. Discussion

As can be seen from Fig. 1, the classical validation approach relies highly on the use of many statistical tests. Statistics should be used to help making the correct decisions; however, the successive tests used here do not allow this as shown by the examples illustrated above and confirmed by the simulations. Indeed, it is the erroneous application and/or interpretation of those tests that lead the analyst either to consider erroneously as valid or not an analytical procedure, leading in all cases at higher risks for the consumer and increasing method development costs.

The first problem is the research of a linear calibration or standard curve (i.e. a straight line) by using the Lack of Fit test. Indeed in this test if the variability of the method is small relative to the variability of the fitting error, then the  $F$  value computed will exceed the theoretic  $F$  value that cannot be exceeded, and therefore the conclusion will be that the linear fit is not adequate. In this case, this conclusion is not necessarily analytically sound:

1. The aim of the calibration curve is to provide by back-calculation accurate results, therefore adequacy of a calibration curve should be made by evaluating the quality of the results provided, as it is proposed using the accuracy profile approach [7].
2. The second problem is a practical one: usually, the replicates used to build the calibration curve are not independent, hence the variability of the method will be small and the previously stated problem will arise. In such situation of dependency of replicates, the mean value should be used to fit the functions tested for calibration curves.
3. Finally the approach proposed in the 92' SFSTP guide for the validation of analytical methods [9], where the calibration curve of 3 distinct days of analysis are pooled together to fit the response functions is not acceptable as it does not reflect the real use that will be made of the method during routine analysis. Indeed, for each day or run of analysis a new calibration curve is made, and unknown samples results are computed on the basis of its corresponding day or run calibration curve, not on the one of the previous day or run. Therefore this approach cannot ensure that the analytical method that will be used in routine is fit for its intended purpose.

The following problem is the interpretation of all the difference tests (Student  $t$ -tests,  $F$  tests,  $\chi^2$  tests) that are used by the classical validation methodologies for the evaluation of the response function and by the difference approach. All these statistical tests are of the following form:

$$\begin{aligned} H_0 &: \theta_1 = \theta_2 \\ H_1 &: \theta_1 \neq \theta_2 \end{aligned}$$

for which a confidence level  $\alpha$  is set, generally at 0.05 in the pharmaceutical field, and where  $\theta_i$  are the parameter of interest (e.g. method bias). This means that it is accepted that the null hypothesis  $H_0$  will be rejected wrongly 5 times out of 100. When the computed quantile is higher than the corresponding theoretical quantile, or equivalently when the  $p$ -value is smaller than  $\alpha$ , the null hypothesis is rejected. Therefore there is high confidence that the difference is real as the confidence level  $\alpha$  is fixed by the analyst. Another way to interpret this test is to look at the  $1-\alpha$  confidence interval of the parameter. If the tested value is outside its corresponding confidence interval then the null hypothesis is rejected.

However the only conclusion that can be made when the null hypothesis is not rejected is not that the two parameters are equal but rather that the test could not demonstrate that they were different from each other. As already discussed in numerous publications [8,19,25,26], the  $\beta$  risk, which is the probability to wrongly accept the null hypothesis, is not fixed by the user in this situation. It has to be noted that this risk does not corresponds to the probability  $\beta$  used in  $\beta$ -expectation tolerance intervals for the accuracy profile methodology (Section 2.2). Nonetheless, this approach can conclude that the parameters are significantly different from each other, whereas they could be analytically acceptable as shown by the simulations performed [8,19,25,26]. Furthermore, these types of tests will also always consider that for example the bias is not different from 0 when the variability of the procedure is relatively high. In fact, the Student *t*-test used this way is a difference test and does not allow the analyst to demonstrate that the analytical method is fit for its intended purpose.

Another issue is the confusion made between estimated parameters such as the mean bias or the RSD<sub>IP</sub> (Descriptive validation methodology—Section 2.1.1) and their true values (Equivalence validation methodology—Section 2.1.3). All the parameters computed during the validation process, such as the mean bias per concentration level or the repeatability and intermediate precision RSDs, are not the true ones. If the analysts restart from scratch the validation process, with everything remaining equivalent in the validation protocol, he will obtain different values of those parameters. They are random values. Indeed, as shown in Section 2.1.1 the descriptive validation approach compares the point estimates of bias or RSD<sub>IP</sub> to a priori fixed acceptance limits. However, even if those estimated parameters are found acceptable there is no guarantee that their corresponding true values are included in the former acceptance limits.

Other authors have introduced the use of equivalence tests to avoid these last limitations [8,16–19]. These approaches are based on the use of confidence intervals for the studied parameter, such as the confidence interval for the mean bias. These intervals represent a region where it is expected to find with a defined probability the true value of the parameter. It is not anymore the estimated parameter that is compared to acceptance limits, but their confidence interval. With these approaches there are good guarantees that the conclusions made are in accordance with the acceptance limits.

A final confusion remains, which is directly observable in regulatory documents relative to the validation of analytical methods: the difference between good methods and good results [7]. All the approaches described above, except the accuracy profile one, look at the performance of the method in terms of bias and precision rather than at the reliability of the results. The analyst has to remind himself what is the final objective of its analytical method as stated in the introduction. Consequently, the objective of an analytical method validation is evident: to give guarantee that the analytical method will provide, when used in routine, accurate (or acceptable) results. This can be achieved by the accuracy profile approach. This tool gives to the analyst a sense of what will be the credibility of the results generated by its analytical method. Furthermore it is a predictive approach [27]. This approach is therefore in good agreement with what will be the daily use of the analytical method, i.e. it allows answering “is the analytical method fit for its intended purpose?”. The accuracy profile methodology allows to control the risk of accepting an unsuitable assay while providing guarantee that the results of the measurements that will be obtained during the future use of the validated method will be included within acceptance limits fixed according to the end user's requirements. This last point is not only essential for the analyst as his conclusions or diagnostics are made on behalf of the results he obtained. It is also an increasing request of new regulatory requirements to manage

the risk associated to the use of these methods in routine analysis [28,29].

## 5. Conclusion

The lack of clear guidance for the methodology to follow to adequately decide when a method can be considered as valid has led to the availability of several methodological approaches and it is therefore the responsibility of the analyst to choose the best one. In this paper, the different conclusions that can be made depending on the decision methodology chosen were illustrated by mean of two examples. Furthermore, simulations have been performed to highlight the risk to use one methodology or the other. This will certainly help analysts to select the decision methodology knowing the risk inherent to its application. Indeed, it has to be reminded that when released to routine applications, the analytical method declared as valid will be used to make highly crucial decisions, going from the marketing of batches of drug products to the evaluation and monitoring of patients health.

## Acknowledgements

The authors are very grateful to the anonymous reviewers for providing important comments that led to significant improvements of this article. A research grant from the Belgium National Fund for Scientific Research (FRS-FNRS) to E. Rozet is gratefully acknowledged.

## References

- [1] International Conference on Harmonization (ICH) of Technical Requirements for registration of Pharmaceuticals for Human Use Topic Q2 (R1): Validation of Analytical Procedures: Text and Methodology, Geneva, 2005.
- [2] ISO 5725 parts 1–6, Accuracy (Trueness and Precision) of Measurement Methods and Results, International Organization for Standardization (ISO), Geneva, 1994.
- [3] Guidance for Industry: Bioanalytical Method Validation, US Department of Health and Human Services, US Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Rockville, 2001.
- [4] The Fitness for Purpose of Analytical Methods, Eurachem, Teddington, 1998.
- [5] ISO/CEI 17025: General Requirements for the Competence of Testing and calibration Laboratories, International Organization for Standardization (ISO), Geneva, 2005.
- [6] B. Boulanger, P. Chiap, W. Dewe, J. Crommen, Ph. Hubert, J. Pharm. Biomed. Anal. 32 (2003) 753.
- [7] E. Rozet, A. Ceccato, C. Hubert, E. Ziemons, R. Oprean, S. Rudaz, B. Boulanger, Ph. Hubert, J. Chromatogr. A 1158 (2007) 111.
- [8] C. Hartmann, J. Smeyers-Verbeke, D.L. Massart, R.D. McDowall, J. Pharm. Biomed. Anal. 17 (1998) 193.
- [9] J. Caporal-Gautier, J.M. Nivet, P. Algranti, M. Guilloteau, M. Histe, L. Lallier, J.J. N'Guyen-Huu, R. Russotto, STP Pharma Pratiques 2 (1992) 205.
- [10] E. Chapuzet, N. Mercier, S. Bervoas-Martin, B. Boulanger, P. Chevalier, P. Chiap, D. Grandjean, Ph. Hubert, P. Lagorce, M. Lallier, M.C. Laparra, M. Laurentie, J.C. Nivet, STP Pharma Pratiques 7 (1997) 169.
- [11] Ph. Hubert, P. Chiap, J. Crommen, B. Boulanger, E. Chapuzet, N. Mercier, S. Bervoas-Martin, P. Chevalier, D. Grandjean, P. Lagorce, M. Lallier, M.C. Laparra, M. Laurentie, J.C. Nivet, Anal. Chim. Acta 391 (1999) 135.
- [12] Ph. Hubert, J.-J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.-A. Compagnon, W. Dewe, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, STP Pharma Pratiques 13 (2003) 101.
- [13] Ph. Hubert, J.-J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.-A. Compagnon, W. Dewe, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, J. Pharm. Biomed. Anal. 36 (2004) 579.
- [14] Ph. Hubert, J.J. N'guyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewé, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, E. Rozet, J. Pharm. Biomed. Anal. 45 (2007) 70.
- [15] Ph. Hubert, J.J. N'guyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewé, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, E. Rozet, J. Pharm. Biomed. Anal. 45 (2007) 82.
- [16] D.J. Schuurmann, J. Pharmacol. Biopharm. 15 (1987) 657.
- [17] C. Hartmann, J. Smeyers-Verbeke, W. Penninckx, Y. Vander Heyden, P. Vankeerberghen, D.L. Massart, Anal. Chem. 67 (1995) 4491.
- [18] R. Kringle, R. Khan-Malek, F. Snikeris, P. Munden, C. Agut, M. Bauer, Drug Inf. J. 35 (2001) 1271.
- [19] C. Hartmann, D.L. Massart, R.D. McDowall, J. Pharm. Biomed. Anal. 12 (1994) 1337.

- [20] E. Rozet, V. Wascotte, N. Lecouturier, V. Pr eat, W. Dew e, B. Boulanger, Ph. Hubert, *Anal. Chim. Acta* 591 (2007) 239.
- [21] D. Hoffman, R. Kringle, *J. Biopharm. Stat.* 15 (2005) 283.
- [22] D. Hoffman, R. Kringle, *Pharm. Res.* 24 (2007) 1157.
- [23] R. Burdik, F. Graybill, *Confidence Interval on Variance Components*, Marcel Dekker, Inc., New York, NY, 1992.
- [24] E. Rozet, W. Dew e, E. Ziemons, A. Bouklouze, B. Boulanger, Ph. Hubert, *J. Chromatogr. B* 877 (2009) 2214.
- [25] H. Rosing, W.Y. Man, E. Doyle, A. Bult, J.H. Beijnen, *J. Liq. Chrom. Rel. Technol.* 23 (2000) 329.
- [26] J. Ermer, *J. Pharm. Biomed. Anal.* 24 (2001) 755.
- [27] E. Rozet, C. Hubert, A. Ceccato, W. Dew e, E. Ziemons, F. Moonen, K. Michail, R. Wintersteiger, B. Streel, B. Boulanger, Ph. Hubert, *J. Chromatogr. A* 1158 (2007) 126.
- [28] International Conference on Harmonization (ICH): Quality Risk Management Q9, Geneva, 2005 <http://www.emea.eu.int/Inspections/docs/ICHQ9Step4QRM.pdf>.
- [29] Process Analytical Technology (PAT) Initiative, US Food and Drugs Administration, Rockville, MD, 2004, <http://www.fda.gov/cder/OPS/PAT.htm>.

**CHAPITRE V : Flexibilité et applicabilité de l'approche de l'intervalle de tolérance  $\beta$ -expectation**

Ce chapitre correspond à l'article « **Flexibility and Applicability of  $\beta$ -expectation tolerance interval approach to assess the fitness of purpose of pharmaceutical analytical methods**»

Publié dans le journal Drug Testing and Analysis le 21 Mai 2012

**1. Synthèse et conclusion****2. Article**



## 1. Synthèse et conclusion

Dans l'industrie pharmaceutique, l'utilisation efficace de n'importe quelle procédure analytique requiert d'avoir la certitude que chaque mesure future de l'analyse de routine, dans les études de stabilité ou les études de validation de processus, sera assez proche de la vraie valeur inconnue de la quantité ou de la concentration de l'analyte contenu dans l'échantillon.

Comme vu précédemment, les approches classiques de la validation vérifient seulement les performances quantitatives par rapport aux valeurs de référence, mais ceci ne reflète pas les besoins réels de l'utilisateur final des résultats produits par la méthode analytique [15]. Une stratégie universelle et innovatrice utilisant le profil d'exactitude calculé à partir de l'intervalle de tolérance d'espérance  $\beta$  et d'erreur totale de mesure a été proposée pour décider de la validité de la méthode. Cette approche contrôle le risque d'accepter une méthode non valide et en même temps la capacité de prévoir la fiabilité des résultats futurs.

Le but de ce travail est d'illustrer l'applicabilité de cette méthodologie de validation dans le cadre de l'industrie pharmaceutique et d'appliquer sa flexibilité d'adaptation à n'importe quel type de méthode analytique rencontrée dans ce secteur, en couvrant différentes matrices pharmaceutiques.

Pour atteindre cet objectif, cette approche harmonisée de validation a été appliquée à diverses techniques analytiques (chromatographie liquide, spectrophotométrie UV), pour différents types de matrices (comprimé, sirop) et pour différentes catégories de test (dosage, dissolution, recherches des impuretés). Chaque exemple a été choisi parce qu'il illustre une situation particulière, classiquement rencontrée par les analystes de l'industrie pharmaceutique.

A travers les différents exemples présentés dans ce travail, l'utilisation du profil d'exactitude comme approche de validation pour les différentes matrices, différentes techniques et pour différentes catégories de tests couramment rencontrés dans l'industrie pharmaceutique a été démontrée. Son applicabilité universelle comme méthodologie de validation efficace dans l'industrie pharmaceutique est démontrée. L'approche du profil d'exactitude évalue bien l'aptitude de ces méthodes à répondre à leur objectif d'utilisation en routine.

Avec ces exemples, nous avons également pu démontrer les éléments suivants:

- L'aptitude de cette approche de validation à prendre en compte l'effet de matrice et la possibilité d'utiliser toute méthode analytique en routine malgré son effet de matrice.
- le concept de série doit incorporer les facteurs appropriés influençant chaque méthode analytique et peut prendre plusieurs dimensions, en plus des facteurs classiques tels que les jours / l'équipement / l'opérateur.

Par ailleurs, nous avons aussi discuté le problème de l'absence des normes relatives aux critères d'acceptation de plusieurs tests: recherche des impuretés, test de dissolution, dosage de conservateur... et nous avons proposé des critères d'acceptation pour certains de ces tests.

## 2. Article

*Article « Flexibility and Applicability of  $\beta$ -expectation tolerance interval approach to assess the fitness of purpose of pharmaceutical analytical methods”*

*Journal of Drug Testing and Analysis, (February 2012), DOI 10.1002/dta.1345*

[Bouabidi A](#), [Talbi M](#), [Bourichi H](#), [Bouklouze A](#), [El Karbane M](#), [Boulanger B](#), [Brik Y](#), [Hubert P](#), [Rozet E](#).

# Flexibility and applicability of $\beta$ -expectation tolerance interval approach to assess the fitness of purpose of pharmaceutical analytical methods

A. Bouabidi,<sup>a,b</sup> M. Talbi,<sup>b</sup> H. Bourichi,<sup>b</sup> A. Bouklouze,<sup>c</sup> M. El Karbane,<sup>c</sup>  
B. Boulanger,<sup>d</sup> Y. BRIK,<sup>e</sup> Ph. Hubert<sup>a§</sup> and E. Rozet<sup>a\*,#,\$</sup>

An innovative versatile strategy using Total Error has been proposed to decide about the method's validity that controls the risk of accepting an unsuitable assay together with the ability to predict the reliability of future results. This strategy is based on the simultaneous combination of systematic (bias) and random (imprecision) error of analytical methods. Using validation standards, both types of error are combined through the use of a prediction interval or  $\beta$ -expectation tolerance interval. Finally, an accuracy profile is built by connecting, on one hand all the upper tolerance limits, and on the other hand all the lower tolerance limits. This profile combined with pre-specified acceptance limits allows the evaluation of the validity of any quantitative analytical method and thus their fitness for their intended purpose.

In this work, the approach of accuracy profile was evaluated on several types of analytical methods encountered in the pharmaceutical industrial field and also covering different pharmaceutical matrices. The four studied examples depicted the flexibility and applicability of this approach for different matrices ranging from tablets to syrups, different techniques such as liquid chromatography, or UV spectrophotometry, and for different categories of assays commonly encountered in the pharmaceutical industry i.e. content assays, dissolution assays, and quantitative impurity assays. The accuracy profile approach assesses the fitness of purpose of these methods for their future routine application. It also allows the selection of the most suitable calibration curve, the adequate evaluation of a potential matrix effect and propose efficient solution and the correct definition of the limits of quantification of the studied analytical procedures. Copyright © 2012 John Wiley & Sons, Ltd.

**Keywords:** accuracy profile; tolerance intervals; total error; analytical method validation

## Introduction

In the pharmaceutical industry, efficient use of any analytical procedures requires to certainty that every future measurement in routine analysis, stability studies or process validation studies, and so on, will be close enough to the unknown true value of the amount or concentration of the analyte contained in the sample.

Classical approaches to validation only check quantitative performances against reference values, but this does not reflect the needs of the end users of the results generated by the analytical method.<sup>[1]</sup> An innovative versatile strategy using accuracy profile computed by means of  $\beta$ -expectation tolerance interval and total measurement error has been proposed to decide about the method's validity that controls the risk of accepting an unsuitable method together with the ability to predict the reliability of future results.

The concept of accuracy profile was first introduced in the papers of Hubert *et al.*<sup>[1]</sup> and Boulanger *et al.*<sup>[2]</sup> After that, in 2003, the commission on the validation of analytical procedures of the Société Française des Sciences et Techniques Pharmaceutiques (SFSTP) used the accuracy profiles for improving the assessment of results' accuracy in method validation studies.<sup>[3]</sup>

Since then, several publications (over 70) have used this new approach in analytical validation,<sup>[1–78]</sup> 18 publications are devoted to discussions on the concept of this approach,<sup>[1–18]</sup> 31 have applied the approach in the validation of bioanalytical methods,<sup>[19–49]</sup> 10 for

pharmaceutical methods,<sup>[50–59]</sup> 7 for the analysis of plant materials,<sup>[60–66]</sup> 6 to transfer of analytical methods from one laboratory to another,<sup>[67–72]</sup> and 6 publications for other applications (food processing, environmental control, etc.).<sup>[73–78]</sup>

The aim of this publication is to demonstrate the high applicability of this validation methodology in the pharmaceutical

\* Correspondence to: Dr E. Rozet, Laboratory of Analytical Chemistry, Institute of Pharmacy, Université de Liège, CHU, B 36, B-4000 Liège, Belgium.  
E-mail: Eric.Rozet@ulg.ac.be

§ These authors contributed equally to this work

# F.R.S.-FNRS Postdoctoral Researcher (Belgium)

a Analytical Chemistry Laboratory, CIRP, Institute of Pharmacy, University of Liège, Belgium

b Analytical Chemistry Laboratory, University Hassan II - Mohammedia - Faculty of Sciences Ben M'Sik, Casablanca, Morocco

c Pharmacology-Toxicology Laboratory, Faculty of Medicine and Pharmacy, University Med V Soussi, Rabat, Morocco

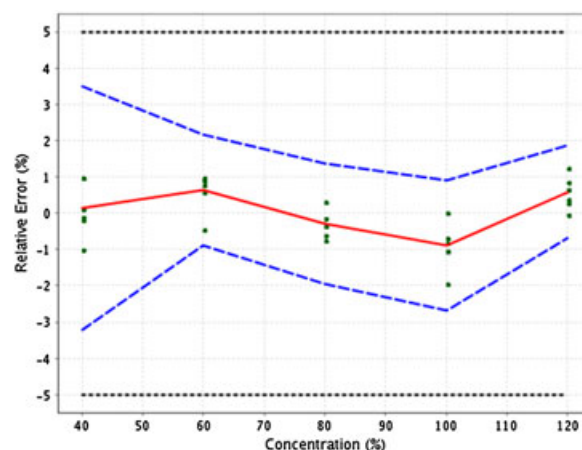
d Arlenda SA, Liège, Belgium

e Drug Control National Laboratory Ministry of Health, Rabat, Morocco

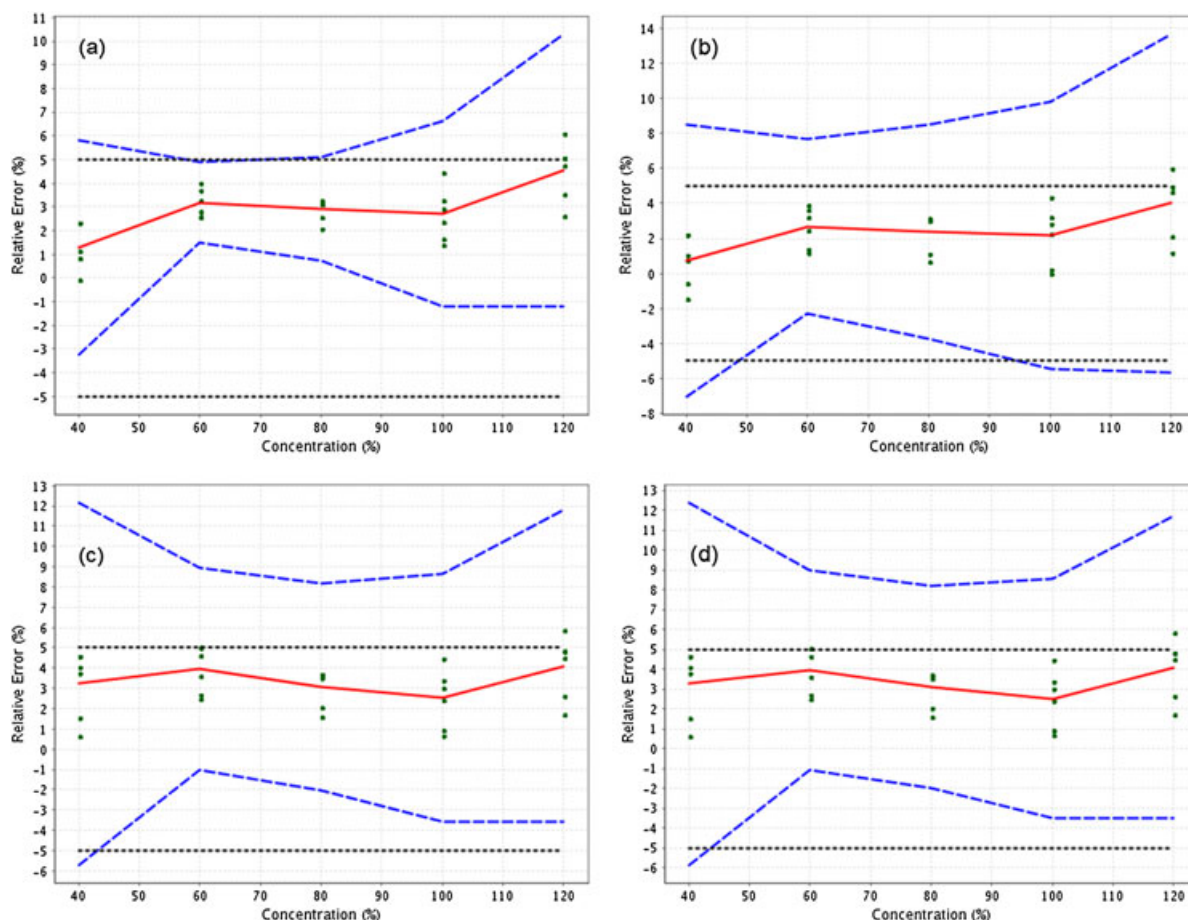
industry field and to illustrate its flexibility with regard to several types of pharmaceutical analytical methods.

To achieve this aim, this harmonized validation approach was applied to various analytical techniques (liquid chromatography (LC), UV spectrophotometry), for different types of matrices (tablet, syrup) and for different category of test (assay, dissolution, impurities). Each example was chosen because it illustrates a specific situation classically faced by analysts in the pharmaceutical industry.

All examples are presented following the same structure. A brief reminder of the procedure provides the type of analytical technique used, as well as the goals to be achieved. Then, the applied experimental design is presented with the description of the calibration and validation standards used. After that, the accuracy profiles obtained with the adequate response function model (the relationship between the response (signal) and the concentration (quantity) of the analyte in the sample, also sometime called calibration or standard curve) is presented and allows interpreting and deciding about the validity of the method. Finally, the corresponding trueness, precision, accuracy data, limit of quantification (LOQ), valid concentration range, as well as the higher and lower tolerance interval limits are then summarized in tables.



**Figure 2.** Accuracy profile of metformin determination in tablets by UV spectrophotometry obtained after application of linear regression using calibration standards prepared with the matrix. The plain line is the relative bias, the dashed lines are the 95%  $\beta$ -expectation tolerance limits and the dotted curves represent the acceptance limits ( $\pm 5\%$ ). The dots represent the relative back-calculated concentrations of the validation standards.



**Figure 1.** Accuracy profiles of metformin determination in tablets by UV spectrophotometry obtained using calibration standards prepared without the matrix after application of (a) Linear regression forced through the origin 0 fitted only with 100% calibration level; (b) Linear regression forced through the origin 0 fitted only with the highest calibration level; (c) Simple linear regression; and (d) Weighted ( $1/X$ ) linear regression. The plain line is the relative bias, the dashed lines are the 95%  $\beta$ -expectation tolerance limits and the dotted curves represent the acceptance limits ( $\pm 5\%$ ). The dots represent the relative back-calculated concentrations of the validation standards.

<b>Table 1.</b> Performance criteria of metformine content validation obtained with the model linear regression (a) within matrix and (b) without matrix				
<b>(a)</b>		<u>Linear regression within matrix</u>		
<b>Response function</b>				
Y = a + bX		<u>Run 1</u>	<u>Run 2</u>	<u>Run 3</u>
	Intercept	-0.01693	-0.01619	0.008182
	Slope b	0.008663	0.008619	0.008337
<b>Trueness</b>				
		<u>Absolute bias, % w/w (Relative, %)</u>		
40 %		0.05705 (0.1)		
60 %		0.3824 (0.6)		
80 %		-0.2381 (-0.3)		
100 %		-0.8983 (-0.9)		
120 %		0.6969 (0.6)		
<b>Precision</b>				
		<u>Repeatability (RSD, %)</u>	<u>Intermediate precision (RSD, %)</u>	
40 %		0.4	0.8	
60 %		0.5	0.5	
80 %		0.2	0.4	
100 %		0.6	0.6	
120 %		0.4	0.4	
<b>Accuracy</b>				
		<u><math>\beta</math>-expectation tolerance limit (% w/w) ; relative <math>\beta</math>-expectation tolerance limit (%)</u>		
40 %		[38.7, 41.4] ; [-3.2, 3.5]		
60 %		[59.5, <b>61.3</b> ] ; [ -0.9, <b>2.2</b> ]		
80 %		<b>[78.4, 81.1]</b> ; [-1.9, 1.4]		
100 %		<b>[97.3, 100.9]</b> ; [-2.7, 0.9]		
120 %		[119.2, 122.3] ; [-0.7, 1.8]		
<b>Linearity</b>				
	Interval (% w/w)	40 – 120		
	Slope	1.000		
	Intercept	-0.00008064		
	r <sup>2</sup>	0.9994		
		40% (w/w)		
<b>LOQ (lower)</b>				
<b>(b)</b>		<u>Linear regression without matrix</u>		
<b>Response function</b>				
Y = a + bX		<u>Run 1</u>	<u>Run 2</u>	<u>Run 3</u>
	Intercept	-0.01212	-0.01530	0.007683
	Slope b	0.008245	0.008269	0.008187
<b>Trueness</b>				
		<u>Absolute bias, % w/w (Relative, %)</u>		
40 %		1.295 (3.2)		
60 %		2.370 (4.0)		
80 %		2.466 (3.1)		
100 %		2.527 (2.5)		
120 %		4.926 (4.1)		
<b>Precision</b>				
		<u>Repeatability (RSD, %)</u>	<u>Intermediate precision (RSD, %)</u>	
40 %		0.4	1.9	
60 %		0.6	1.2	
80 %		0.2	1.1	
100 %		0.9	1.6	
120 %		0.6	1.7	
<b>Accuracy</b>				
		<u><math>\beta</math>-expectation tolerance limit (% w/w) ; relative <math>\beta</math>-expectation tolerance limit (%)</u>		
40 %		[37.7, 44.9] ; [-5.7, 12.2]		
60 %		[59.4, 65.4] ; [ -1.0, 8.9]		
80 %		[78.4, 86.6] ; [-2.0, 8.2]		

(Continues)

Table 1. (Continued)

(a)		Linear regression within matrix		
		Run 1	Run 2	Run 3
<b>Response function</b>				
Y = a + bX				
	Intercept	-0.01693	-0.01619	0.008182
	Slope b	0.008663	0.008619	0.008337
100 %		[96.4 , 108.6] ; [-3.6 , 8.6]		
120 %		[115.7 , 134.2] ; [-3.6 , 11.8]		
<b>Linearity</b>				
	Interval (% w/w)	40 – 120		
	Slope	1.037		
	Intercept	-0.2505		
	r <sup>2</sup>	0.9982		
<b>LOQ (lower)</b>				
		No LOQ		

## Materials and methods

### Determination of metformin in tablets by UV spectrophotometry

#### Chemicals and reagents

Metformin hydrochloride standard reference was supplied by the European Pharmacopoeia (Strasbourg, France).

#### Instrumentation

Determination of metformin was performed with an Agilent UV-VIS 8453E double beam spectrophotometer (Agilent, Palo Alto, CA, USA) at a wavelength of 232 nm using purified water as blank.

#### Calibration and validation standards

The following samples were prepared:

- **Calibration standards:** Prepared out of the matrix, 3 concentration levels: 80, 100, 120%, 2 repetitions per level during 3 days. The same calibration standards are prepared with the matrix.
- **Validation standards:** Prepared in a reconstituted matrix (placebo) at 3 concentration levels: 80, 100, 120% of the target value, analyzed during 3 days. The levels 80 and 120% were replicated twice per day of analysis, while for the level 100% it was 6 times.

### Determination of fluconazole in dissolution media by high performance liquid chromatography (HPLC)

#### Chemicals and reagents

All chemicals and solvents used were of analytical or HPLC grade. Fluconazole was supplied by the European Pharmacopoeia. Potassium dihydrogen phosphate was purchased from Acros Organics (Geel, Belgium).

Methanol and phosphoric acid (85%) were obtained from Merck (Darmstadt, Germany). Deionized water was generated from the Milli-Q water purifying system (Millipore, Watford, UK).

#### Instrumentation and chromatographic conditions

Analyses were performed on an Agilent technologies HPLC 1100 series system (Hewlett-Packard, Palo Alto, CA, USA) equipped

with a solvent delivery quaternary pump G1311A, an on-line degasser G1322A, an autosampler G1313A, a column oven G1316A and a diode-array detector G1315A.

Chromatographic analysis was performed on a C18 Thermo Hypersil BDS column (150 x 4.6 mm i.d., 5 µm particle size) and kept at 25 °C. The mobile phase was prepared by mixing separately measured methanol and phosphate aqueous solution (prepared by dissolving 1.4 g of potassium dihydrogen phosphate in 1000 ml of water and adjusted to pH 3.6 with phosphoric acid) in a ratio 40:60 (v/v) and was degassed before use. The HPLC system was operated isocratically at a flow rate of 1.0 ml/min and the injection volume was 20 µl. UV detection was performed at 261 nm.

#### Calibration and validation standards

A stock solution of fluconazole was prepared by accurately weighting 100.0 mg of fluconazole and diluting this in 100.0 ml of methanol. The calibration standards for fluconazole were prepared by diluting an appropriate volume of stock solution with 0.1 M HCl aqueous solution to reach the concentration levels corresponding to the range 25% to 125% of fluconazole i.e. 12.5 (25%), 42.5 (85%), 50 (100%), 57.5 (115%) and 62.5 mg (125%) of fluconazole. The standards were analyzed in duplicate.

The validation standards were prepared similarly with addition of the excipients of the formulation of the drug product to reach four concentration levels ranging from 25% to 125% of fluconazole i.e. 12.5 (25%), 25.52 (50%), 50 (100%), and 62.5 mg (125%) of fluconazole. Each solution of the validation standard is introduced in two breakers and each validation standard (from each breaker) was analyzed two times. So the validation study was performed on six different series leading to a total of 96 analyses.

### Determination of methyl parahydroxybenzoate content (preservative) in syrup by HPLC

#### Chemicals and reagents

Methyl parahydroxybenzoate standard reference was supplied by the European Pharmacopoeia. Acetonitrile of HPLC grade was purchased from Merck. Sodium acetate of analytical grade was supplied from Sigma-Aldrich (Steinheim, Germany).

Deionized water was generated from a Milli-Q water purifying system.

#### Instrumentation and chromatographic conditions

The HPLC system consisted in a LaChrom (Merck-Hitachi, Darmstadt, Germany) composed of a quaternary pump L-7100, an autosampler L-7200, an oven L-7360 and a DAD detector L-7455. The mobile phase consisted in a 33/67 (v/v) mixture of acetonitrile and an aqueous buffer (1.36 g of sodium acetate in 1000 ml water adjusted to pH 5.0 with glacial acetic acid). The chromatographic isocratic separation was made with a Nucleosil C18 column (150 x 4.6 mm ID; particle size: 5  $\mu$ m; Macherey-Nagel, Hoerd, France) and the UV detection was performed at 245 nm. The flow rate was set at 1.5 ml/min and 20  $\mu$ l of each sample was injected onto the column. The analysis run time was of 15 min.

#### Calibration and validation standards

Calibration standards and validation standards were prepared at five concentration levels: 60% (9  $\mu$ g/ml), 80% (12  $\mu$ g/ml), 100% (15  $\mu$ g/ml), 120% (18  $\mu$ g/ml) and 140% (21  $\mu$ g/ml) of the target nominal concentration of methyl parahydroxybenzoate in the syrup formulation. The calibration standards prepared in methanol were analyzed once during 3 days. Each validation standard prepared in a reconstituted matrix of the formulation (or placebo) was analyzed in triplicates during the same three days, except the 100% level that was replicated 6 times each day.

#### Quantification of detection of known impurities in amoxicilline tablets

##### Chemicals and reagents

Amoxicilline standard reference was supplied by the European Pharmacopoeia. Acetonitrile of HPLC Grade was purchased from Merck.  $\text{KH}_2\text{PO}_4$  of analytical grade was supplied from Sigma-Aldrich. Deionized water was generated from a Milli-Q water purifying system.

#### Instrumentation and chromatographic conditions

The HPLC system consisted is the same LaChrom system previously described. The mobile phase consisted in a 96/4 (v/v) mixture of acetonitrile and an aqueous buffer (6.8 g of  $\text{KH}_2\text{PO}_4$  in 1000 ml of water, adjusted to pH 5.0 with KOH 45% (m/v)). The chromatographic isocratic separation was made with a Lichrosphere C18 column (250 x 4.6 mm ID; particle size: 5  $\mu$ m) and the UV detection was performed at 230 nm. The flow rate was set at 1.5 ml/min and 20  $\mu$ l of each sample was injected onto the column. The analysis run time was of 15 min.

#### Calibration and validation standards

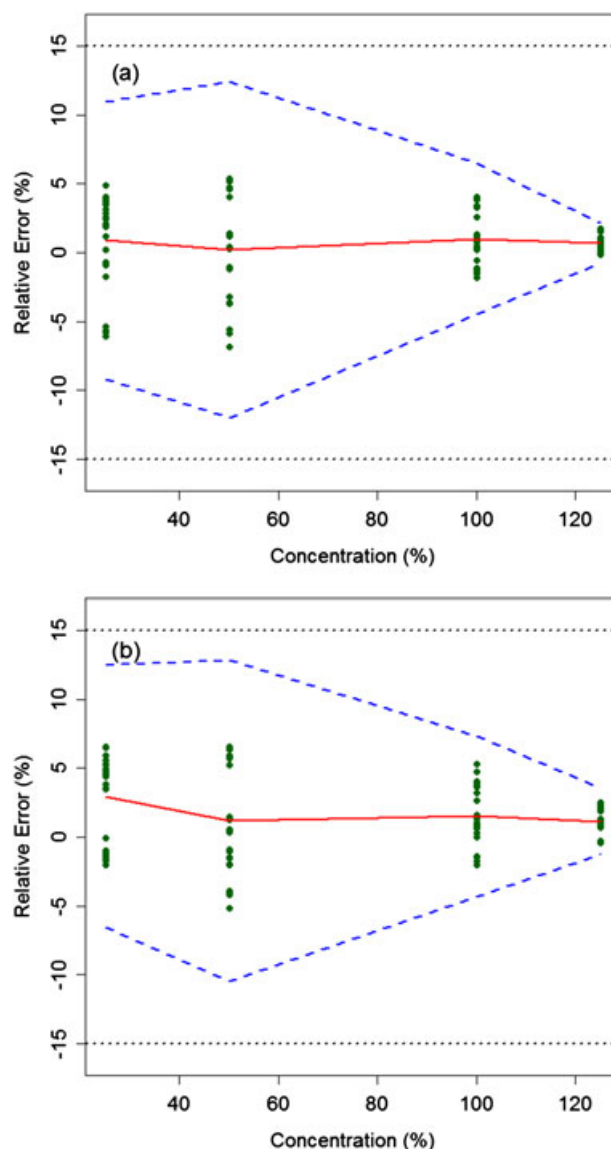
Calibration standards were prepared at twelve concentration levels: 20% (5 mg), 40% (10 mg), 50% (12.5 mg), 60% (15 mg), 80% (20 mg), 90% (22.5 mg), 100% (25 mg), 110% (27.5 mg), 120% (30 mg), 140% (35 mg), 150% (37.5 mg), and 160% (40 mg) of the specification of three impurities (impurity A, B & C). The calibration standards prepared in the mobile phase were analyzed twice during 3 days. Each validation standard was prepared in a reconstituted matrix of the formulation (or placebo) at five concentration levels (80%, 90%, 100%, 110%, and 120%) and was analyzed in duplicates during the same three days.

## Results and discussion

### Determination of metformin in tablets by UV spectrophotometry

By using the approach proposed by Hubert *et al.*,<sup>[6,11,12]</sup> the method is considered as not valid within the studied range of concentration when the accuracy profile crosses the accuracy acceptance limits set at  $\pm 5\%$ . This validation approach gives the guarantee that each future result generated by the method will be included within  $\beta$ -expectation tolerance limits with a user defined guarantee. Here this guarantee is set at 95.0%.

Four calibration models were investigated, namely the simple linear regression, the linear regression forced through the origin (0) fitted only with the 100% calibration level, the linear regression



**Figure 3.** Accuracy Profiles of fluconazole dissolution obtained after application of linear regression through 0 fitted using the highest level only (a) and simple linear regression (b). The plain line is the relative bias, the dashed lines are the 95%  $\beta$ -expectation tolerance limits and the dotted curves represent the acceptance limits ( $\pm 15\%$ ). The dots represent the relative back-calculated concentrations of the validation standards.



**Table 2.** Performance criteria of fluconazole dissolution obtained with the model; linear Regression through 0 fitted using the highest level only (a) and simple linear regression (b)

a) linear regression through 0 fitted using the highest level only							
<b>Response function</b>							
Y = b x		<u>Run 1</u>	<u>Run 2</u>	<u>Run 3</u>	<u>Run 4</u>	<u>Run 5</u>	<u>Run 6</u>
Slope b		1.2276E + 04	1.2276E + 04	1.2349E + 04	1.2349E + 04	1.2262E + 04	1.2262E + 04
<b>Trueness</b>							
<u>Absolute bias, % w/w (Relative, %)</u>							
25 %		0.2240 (0.9)					
50.0%		0.0870 (0.2)					
100.0%		1.034 (0.9)					
125.0%		0.8782 (0.7)					
<b>Precision</b>							
		<u>Repeatability (RSD, %)</u>			<u>Intermediate precision (RSD, %)</u>		
25 %		0.5			3.7		
50.0%		0.4			4.4		
100.0%		0.6			2.0		
125.0%		0.2			0.5		
<b>Accuracy</b>							
<u>β-expectation tolerance limit (% w/w) ; relative β-expectation tolerance limit (%)</u>							
25 %		[ 23.0 , 28.2] ; [-9.2 , 10.9]					
50.0%		[ 44.8 , 57.4] ; [-12.1 , 12.4]					
100.0%		[ 96.4 , 107.6] ; [-4.5 , 6.5]					
125.0%		[ 122.8 , 127.2] ; [-0.7 , 2.1]					
<b>m</b>							
		Interval (% w/w)			25 – 125		
		Slope			1.009		
		Intercept			-0.06430		
		r <sup>2</sup>			0.9985		
<b>LOD</b>							
1.45 (% w/w)							
<b>LOQ</b>							
Lower LOQ (% w/w) = 25							
Upper LOQ (% w/w) = 125							
b) Linear regression							
<b>Response function</b>							
Y = a + b x		<u>Run 1</u>	<u>Run 2</u>	<u>Run 3</u>	<u>Run 4</u>	<u>Run 5</u>	<u>Run 6</u>
Slope b		1.2472E + 04	1.2472E + 04	1.2290E + 04	1.2290E + 04	1.2130E + 04	1.2130E + 04
Intercept		-8803	-8800	1187	1190	-2028	-2030
<b>Trueness</b>							
<u>Absolute bias, % w/w (Relative, %)</u>							
25 %		0.7446 (2.9)					
50.0%		0.5956 (1.2)					
100.0%		1.5146 (1.5)					
125.0%		1.3972 (1.1)					
<b>Precision</b>							
		<u>Repeatability (RSD, %)</u>			<u>Intermediate precision (RSD, %)</u>		
25 %		0.5			3.5		
50.0%		0.4			4.2		
100.0%		0.6			2.2		
125.0%		0.2			0.9		
<b>Accuracy</b>							
<u>β-expectation tolerance limit (% w/w) ; relative β-expectation tolerance limit (%)</u>							
25 %		[23.6 , 28.4] ; [-6.6 , 12.5]					
50.0%		[45.6 , 57.6] ; [-10.4 , 12.8]					
100.0%		[96.6 , 108.4] ; [-4.3 , 7.3]					
125.0%		[123.0 , 129.0] ; [-1.3 , 3.5]					
<b>Linearity</b>							
		Interval (% w/w)			25 – 125		
		Slope			1.009		

(Continues)

Table 2. (Continued)						
a) linear regression through 0 fitted using the highest level only						
Response function						
Y = b x	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6
Slope b	1.2276E + 04	1.2276E + 04	1.2349E + 04	1.2349E + 04	1.2262E + 04	1.2262E + 04
LOD	Intercept $r^2$			0.1933		
LOQ				0.9984		
				5.27 %, w/w		
				Lower LOQ (% w/w) = 25		
				Upper LOQ (% w/w) = 125		

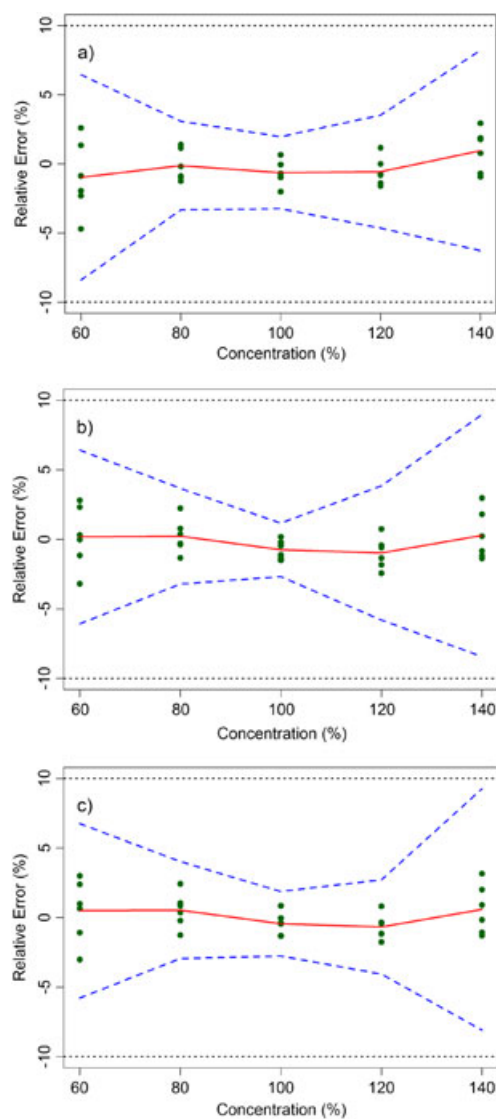
forced through the origin (0) fitted only with the 120% calibration level and a weighted (1/X) linear regression. Accuracy profiles obtained with these calibration curves are shown in Figure 1.

As illustrated in Figure 1, the method is thus considered not valid over the concentration range investigated whatever the calibration model used. It can be observed that the tolerance intervals are outside the  $\pm 5\%$  acceptance limits.

In order to understand the problem encountered with this validation, we suspected the presence of matrix effect problem. This was observed by using calibration standards prepared with all the excipients to construct different response functions. Indeed, after using these calibration curves prepared within the matrix, the method can now be considered as valid. This is shown on Figure 2 that depicts the accuracy profile obtained after application of linear regression with calibration standards prepared within the matrix.

Indeed, with the classical validation methodologies,<sup>[15,2,10,78–84]</sup> when a matrix effect is detected, the usually decision made is to go back to the development or optimization phase of the method or simply its rejection. So this example illustrates that an analytical method can still be useful and valid with this new validation approach, even if a matrix effect has been detected. A solution to circumvent this matrix effect problem is to prepare calibration standards with all the matrix components.

Table 1 presents the validation criteria such as trueness, precision, accuracy of the results, linearity of the results, and the limit of quantification obtained with the linear calibration curve prepared with all the matrix components included into the calibration standards (Table 1a) as well as the validation criteria obtained with the linear calibration curve prepared without the matrix component (Table 1b) for comparison. As can be seen in Table 1b, the main problem arising from using a calibration curve prepared without the matrix component is that it provides a high relative bias (ranging from 2.5 to 4.1%) while intermediate precision RSD never exceeds 2.0%. As it seems that the good precision of the method could compensate the weakness of the trueness of the method, it could be tempting to declare this method as valid. However the accuracy profile depicted in Figure 1c and the values of the relative 95%  $\beta$ -expectation tolerance intervals show that there is high risk that the method will provide inaccurate and unreliable results when performed with calibration standards prepared without the matrix. Indeed, from the various methodologies available to decide about the validity of an analytical method, the accuracy profile approach using statistical tolerance is almost the only one providing that the results will be of enough reliability.<sup>[15]</sup> This example of



**Figure 4.** Accuracy profiles of the HPLC-UV method for the determination of methyl parahydroxybenzoate obtained for the validation standards with the simple linear regression model (a) and with the linear model forced through the origin and fitted only with the 100.0% calibration level (b) or fitted only with the 140.0% calibration level (c). The plain line is the relative bias, the dashed lines are the 95%  $\beta$ -expectation tolerance limits and the dotted curves represent the acceptance limits ( $\pm 10\%$ ). The dots represent the relative back-calculated concentrations of the validation standards.

analytical method validation aiming at controlling the concentration of an active ingredient in tablets illustrates the accuracy profile approach's flexibility in regards to matrix effects. It allows demonstrating that a method can be validated even if a matrix effect is present. It further guarantees the quality of the future results that will be generated by the validated analytical method.

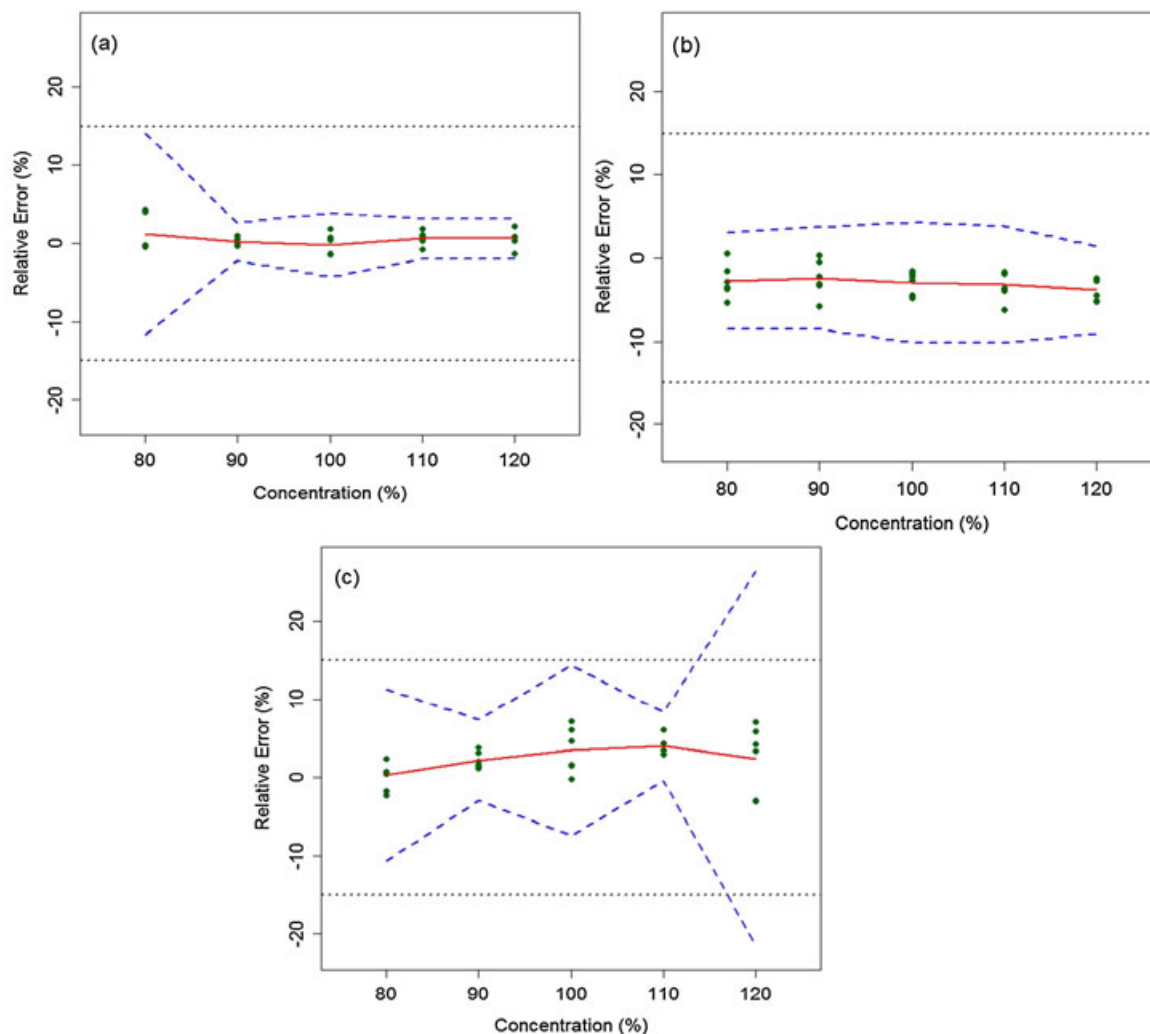
### Determination of fluconazole in dissolution media by HPLC

In order to evaluate the validation of the analytical method aiming at quantifying fluconazole in dissolution baths with the total error approach, we have to set the acceptance limits. Such limits are missing in the guidelines. We will compare the dissolution test to the test of uniformity of content of single-dose preparations.

In the European Pharmacopoeia<sup>[85]</sup> definition of the uniformity content test we found that: *the test . . . is based on the assay of the individual contents of active substance(s) of a number of single-dose units to determine whether the individual contents are within limits set with reference to the average content of the sample.*

This definition is similar to the definition of dissolution tests as mentioned in the European Pharmacopoeias since the objective of a dissolution test is the assay of the individual contents of active substance(s) of each breaker in order to determine whether the individual contents are within limits set with reference to the average content of the sample. In another paragraph of the same document,<sup>[85]</sup> we found that these limits are set at 85% and 115% of the nominal concentration. Therefore, the acceptance limits are set at  $\pm 15\%$  for the analytical method validation and the risk is set at 5%.

Several calibration models were tested to analyze the relationship between the amount of fluconazole (in mg) and the instrumental response (the chromatographic peak area) and decide which model is most appropriate. Two models were relevant to perform the calibration and back calculate the concentrations of the validation standards: the simple linear regression and the regression through the origin fitted to the 62.5 mg (125%) calibration standard. The obtained accuracy profiles are presented in Figure 3. Table 2 presents the method's validation criteria obtained with these models. As shown in Figure 3 and Table 2, the method of dissolution of fluconazole is valid with these two



**Figure 5.** Accuracy profiles obtained after application of a linear regression model forced through the origin and fitted only with the highest level of the calibration standards for Impurity A (a), Impurity C (b), and Impurity D (c). The plain line is the relative bias, the dashed lines are the 95%  $\beta$ -expectation tolerance limits and the dotted curves represent the acceptance limits ( $\pm 15\%$ ). The dots represent the relative back-calculated concentrations of the validation standards.

**Table 3.** Performance criteria of method validation obtained with a linear regression model forced through the origin and fitted only with the highest level of the calibration standards for Impurity A (a), Impurity C (b), and Impurity D (c)

a) Impurity A			
<b>Response function</b>			
Y = b x		<u>Run 1</u>	<u>Run 2</u>
	Slope b	4095	4164
			<u>Run 3</u>
			4395
<b>Trueness</b>			
		<u>Absolute bias, % w/w (Relative, %)</u>	
80 %		0.96 (1.2)	
90%		0.18 (0.2)	
100%		-0.20 (-0.2)	
110 %		0.66 (0.6)	
120%		0.72 (0.6)	
<b>Precision</b>			
		<u>Repeatability (RSD, %)</u>	<u>Intermediate precision (RSD, %)</u>
80 %		0.2	2.6
90%		0.3	0.6
100%		1.3	1.4
110 %		0.8	0.9
120%		0.9	1.2
<b>Accuracy</b>			
		<u>β-expectation tolerance limit (% w/w) ; relative β-expectation tolerance limit (%)</u>	
80 %		[70.64 , 91.28] ; [-11.7 , 14.1]	
90%		[88.02 , 92.34] ; [-2.2 , 2.6]	
100%		[95.80 , 103.8] ; [-4.2 , 3.8]	
110 %		[107.91 , 113.52] ; [-1.9 , 3.2]	
120%		[116.04 , 125.40] ; [-3.3 , 4.5]	
<b>Linearity</b>			
		Interval (% w/w)	80 – 120
		Slope	0.9999
		Intercept	0.1194
		r <sup>2</sup>	0.9922
<b>LOD</b>			
<b>LOQ</b>			
			0.49 % w/w
			Lower LOQ (% w/w) = 80
			Upper LOQ (% w/w) = 120
b) Impurity C			
<b>Response function</b>			
Y = b x		<u>Run 1</u>	<u>Run 2</u>
	Slope b	2211	2272
			<u>Run 3</u>
			2464
<b>Trueness</b>			
		<u>Absolute bias, % w/w (Relative, %)</u>	
80 %		-2.16 (-2.7)	
90%		-2.16 (-2.4)	
100%		-2.90 (-2.9)	
110 %		-3.41 (-3.1)	
120%		-4.56 (-3.8)	
<b>Precision</b>			
		<u>Repeatability (RSD, %)</u>	<u>Intermediate precision (RSD, %)</u>
80 %		2.0	2.0
90%		2.1	2.2
100%		0.4	1.5
110 %		1.2	1.9
120%		0.8	1.4
<b>Accuracy</b>			
		<u>β-expectation tolerance limit (% w/w) ; relative β-expectation tolerance limit (%)</u>	
80 %		[73.20 , 82.40] ; [-8.5 , 3.0]	
90%		[82.35 , 93.33] ; [-8.5 , 3.7]	

(Continues)

**Table 3.** (Continued)

a) Impurity A			
<b>Response function</b>			
Y = b x		<u>Run 1</u>	<u>Run 2</u>
	Slope b	4095	4164
100%		[89.90 , 104.20] ;[-10.1 , 4.2]	
110 %		[98.89 , 114.18] ;[-10.1 , 3.8]	
120%		[109.2 , 121.68] ;[-9.0 , 1.4]	
<b>Linearity</b>			
	Interval (% w/w )	80% – 120%	
	Slope	0.9393	
	Intercept	0.7538	
	r <sup>2</sup>	0.9851	
<b>LOD</b>			
<b>LOQ</b>			
		3.47 % w/w	
		Lower LOQ (% w/w) = 80	
		Upper LOQ (% w/w) = 120	
b) Impurity D			
<b>Response function</b>			
Y = b x		<u>Run 1</u>	<u>Run 2</u>
	Slope b	1.0629E + 04	1.0893E + 04
<b>Trueness</b>			
		Absolute bias, % w/w (Relative, %)	
80 %		0.24 (0.3)	
90%		1.98 (2.2)	
100%		3.49 (3.5)	
110 %		4.29 (3.9)	
120%		2.88 (2.4)	
<b>Precision</b>			
		<u>Repeatability (RSD, %)</u>	<u>Intermediate precision (RSD, %)</u>
80 %		0.3	2.2
90%		0.3	1.1
100%		2.0	3.1
110 %		0.7	1.2
120%		0.6	4.9
<b>Accuracy</b>			
		<u>β-expectation tolerance limit (% w/w) ; relative β-expectation tolerance limit (%)</u>	
80 %		[71.44 , 89.04] ;[-10.7 , 11.3]	
90%		[87.30 , 96.66] ;[-3.0 , 7.4]	
100%		[92.60 , 114.40] ;[-7.4 , 14.4]	
110 %		[109.56 , 119.24] ;[-0.4 , 8.4]	
120%		[94.20 , 151.68] ;[-21.5 , 26.4]	
<b>Linearity</b>			
	Interval (% w/w )	80% – 110%	
	Slope	1.080	
	Intercept	-1.368	
	r <sup>2</sup>	0.9597	
<b>LOD</b>			
<b>LOQ</b>			
		4.02% w/w	
		Lower LOQ (% w/w) = <b>80</b>	
		Upper LOQ (% w/w) = 110	

calibration models: simple linear regression and regression through the origin fitted to the 62.5 mg (125%). As in both cases it is guaranteed that each future results will be within the +/- 15% acceptance limits, the analyst is free to select the most simple one. In this case the calibration model that would finally be selected would be the one using the regression through the origin fitted to the 62.5 mg (125%) as it is the most efficient situation.

With this example, we demonstrated:

- The applicability of the accuracy profile as an adequate validation approach for the dissolution test.
- The possibility to use new factors to set the series, others than the one usually used such as days, equipments or operators. For this dissolution test, a combination of two factors was selected: days (at three levels) and breaker (at two levels) as part

of equipment). As discussed by Rozet *et al.*,<sup>[10]</sup> this concept of series or runs which can take several dimensions must include the appropriate factors for the analytical method and in the case of dissolution tests breakers are important components of the equipments.

- The problem of lack of acceptance criteria in the guidelines for some tests has been discussed (here dissolution test) and acceptance limits for the dissolution tests have been proposed.

#### Determination of methyl parahydroxybenzoate content (preservative) in syrup by HPLC

The accuracy profile approach<sup>[6,11,12]</sup> has also been used to evaluate the validity of the HPLC-UV method for the determination of methyl parahydroxybenzoate in syrup. The accuracy acceptance limits were set at  $\pm 10\%$  and the minimum probability to obtain each future result generated by this method within the  $\pm 10\%$  acceptance limits is set at 95.0%. Three calibration models were investigated. They are the simple linear regression, the linear regression forced through the origin (0) and fitted only with the 100% calibration level or fitted only with the 140% calibration level. Accuracy profiles obtained with these calibration curves for the validation standards are shown in Figure 4.

All three calibration curves allow the method to be considered as valid over the whole concentration range investigated. Indeed, the 95% tolerance intervals are fully included within the  $\pm 10\%$  acceptance limits in all cases. The final calibration curve selected was thus the calibration curve forced through the origin and fitted only with the 140% level of the calibration standards as it is one of the simplest calibration model, it provides the least bias and leads to the least extrapolation of results.

With this example the applicability of the accuracy profile approach for the validation of an assay aiming at quantifying a preservative in a drug product was illustrated.

#### Quantification of known impurities in amoxicilline tablets

Several calibration models were tested to find the adequate response function. The linear regression model was suitable to perform the calibration and back calculate the concentrations of the validation standards as shown by the accuracy profile of Figure 5. Table 3 presents the validation criteria obtained for the analysis of the impurities. The acceptance limits were set at  $+ / -15\%$  and the risk at 5%.

##### Impurity A & C

The accuracy profiles obtained for impurity A and C (Figures 5a and 5b) show that the method is valid and allows to quantify these impurities with adequate reliability over the whole concentration range investigated. Indeed, these two profiles can ensure that each future result obtained for these two analytes will be within the acceptance limits of  $\pm 15\%$  with 95% probability.

##### Impurity D

The accuracy profile of impurity D (Figure 5c) shows that the tolerance interval of the highest level of the validation standard is not inside the  $\pm 15\%$  acceptance limits. We can conclude that the method is only valid on the interval [80–110%].

This example highlights the fact that the  $\beta$ -expectation tolerance interval approach can be used successfully for the validation of quantitative impurity assays.

## Conclusions

Through the different examples presented in this work, the use of the accuracy profile as a validation approach for different matrices and different techniques and for different categories of assays commonly encountered in the pharmaceutical industry has been shown. Its applicability as an efficient validation methodology in the pharmaceutical industry is demonstrated. The accuracy profile approach assesses the fitness for purpose of these methods for their future routine application.

With the accuracy profile approach, the most important in the validation is the intended objective of this method in routine and nothing else; this is based on comparing the predictive distribution of the results to the acceptance limits to decide about the validity of a method.

With these examples, we also demonstrated:

- The flexibility to consider the matrix effect and the possibility of any analytical method to be useful even with its matrix effect.
- The concept of series or runs must incorporate the appropriate factors for the analytical method and must take several dimensions in addition of the classical factors such day/equipment / operator.

Moreover, we have underlined the problem of the absence in the guidelines of acceptance criteria for several tests: quantitative impurities tests, dissolution tests and determination of preservative content, etc. We have thus proposed several acceptance criteria for some of these tests.

## Acknowledgements

The authors are very grateful to the anonymous reviewers for providing important comments that led to significant improvements of this article. A research grant from the Belgium National Fund for Scientific Research (FRS-FNRS) to E. Rozet is gratefully acknowledged.

## REFERENCES

- [1] P. Hubert, P. Chiap, J. Crommen, B. Boulanger, E. Chapuzet, N. Mercier, S. Bervoas-Martin, P. Chevalier, D. Grandjean, P. Lagorce, M.C. Laparra, M. Laurentie, J.C. Nivet. The SFSTP guide on the validation of chromatographic methods for drug bioanalysis: From the Washington Conference to the laboratory. *Anal. Chim. Acta* **1999**, 391, 135.
- [2] B. Boulanger, P. Chiap, W. Dewe, J. Crommen, P. Hubert. An analysis of the SFSTP guide on validation of chromatographic bioanalytical methods: Progresses and limitations. *J. Pharm. Biomed. Anal.* **2003**, 32, 753.
- [3] P. Hubert, J.-J. Nguyen-huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P. Compagnon, W. Déwe, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat. Validation of quantitative analytical procedures, harmonization of approaches. *STP Pharma. Prat.* **2003**, 13, 101.
- [4] G. Gonzalez, M.A. Herrador. Accuracy profiles from uncertainty measurements. *Talanta* **2006**, 70, 896.
- [5] S. Bervoas-Martin, B. Boulanger, E. Chapuzet, P. Chevalier, P. Chiap, D. Grandjean, P. Hubert, P. Lagorce, M. Lallier, M.-Ch. Laparra, M. Laurentie, N. Mercier, C. Nivet. New strategy for the validation of chromatographic

- bio-analytical methods. Report of a SFSTP Commission. *STP Pharma. Prat.* **2000**, 10, 12.
- [6] P. Hubert, J.-J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.-A. Compagnon, W. Dewe, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat. Harmonization of strategies for the validation of quantitative analytical procedures. A SFSTP proposal? Part I. *J. Pharm. Biomed. Anal.* **2004**, 3, 579.
- [7] M. Feinberg, B. Boulanger, W. Dewé, P. Hubert. New advances in method validation and measurement uncertainty aimed at improving the quality of chemical data. *Anal. Bioanal. Chem.* **2004**, 380, 502.
- [8] P. Hubert, E. Rozet, B. Boulanger, W. Dewe, M. Laurentie, N. Dubois, C. Charlier, M. Feinberg. Harmonisation des stratégies de validation et estimation de l'incertitude associée dans le cadre de l'accréditation des laboratoires d'essais. *Acta Clin. Belg.* **2006**, 61, 54.
- [9] P. Hubert, E. Rozet, B. Boulanger, W. Dewé, M. Laurentie, N. Dubois, C. Charlier, M. Feinberg. Synchronization of validation and estimation strategies of doubt associated as part of the accreditation in trial laboratories. *Acta Clin. Belg.* **2006**, 61, 54.
- [10] E. Rozet, A. Ceccato, C. Hubert, E. Ziemons, R. Oprean, S. Rudaz, B. Boulanger, P. Hubert. An analysis of recent pharmaceutical regulatory documents on analytical method validation. *J. Chromatogr. A* **2007**, 1158, 111.
- [11] P. Hubert, J.-J. Nguyen-Huu, B. Boulanger, E. Chapuzet, N. Cohen, P.-A. Compagnon, W. Dewé, M. Feinberg, M. Laurentie, N. Mercier, G. Muzard, L. Valat, E. Rozet. Harmonization of strategies for the validation of quantitative analytical procedures. A SFSTP proposal—part III. *J. Pharm. Biomed. Anal.* **2007**, 45, 82.
- [12] P. Hubert, J.-J. Nguyen-Huu, B. Boulanger, E. Chapuzet, N. Cohen, P.-A. Compagnon, W. Dewé, M. Feinberg, M. Laurentie, N. Mercier, G. Muzard, L. Valat, E. Rozet. Harmonization of strategies for the validation of quantitative analytical procedures. A SFSTP proposal—part II. *J. Pharm. Biomed. Anal.* **2007**, 45, 70.
- [13] E. Rozet, C. Hubert, A. Ceccato, W. Dewé, E. Ziemons, F. Moonen, K. Michail, R. Wintersteiger, B. Streel, B. Boulanger, P. Hubert. Using tolerance intervals in pre-study validation of analytical methods to predict in-study results - The fit-for-future-purpose concept. *J. Chromatogr. A* **2007**, 1158, 126.
- [14] P. Hubert, J.-J. Nguyen-Huu, B. Boulanger, E. Chapuzet, N. Cohen, P.-A. Compagnon, W. Dewé, M. Feinberg, M. Laurentie, N. Mercier, G. Muzard, L. Valat, E. Rozet. Harmonization of strategies for the validation of quantitative analytical procedures: a SFSTP proposal part IV. examples of application. *J. Pharm. Biomed. Anal.* **2008**, 48, 760.
- [15] A. Bouabidi, E. Rozet, M. Fillet, E. Ziemons, E. Chapuzet, B. Mertens, R. Klinkenberg, A. Ceccato, M. Talbi, B. Streel, A. Bouklouze, B. Boulanger, P. Hubert. Critical analysis of several analytical method validation strategies in the framework of the fit for purpose concept. *J. Chromatogr. A* **2010**, 1217, 3180.
- [16] S. Rudaz, P. Hubert. Method validation, comparison and transfer. *J. Chromatogr. B* **2009**, 877, 2179.
- [17] G. Gonzalez, M.A. Herrador. A practical guide to analytical method validation, including measurement uncertainty and accuracy profiles. *Trends Anal. Chem.* **2007**, 26, 227.
- [18] B. Boulanger, W. Dewé, A. Gilbert, B. Govaerts, M. Maumy-Bertrand. Risk management for analytical methods based on the total error concept: Conciliating the objectives of the pre-study and in-study validation phases. *Chemometr. Intell. Lab. Syst.* **2007**, 86, 198.
- [19] P. Chiap, P. Hubert, B. Boulanger, J. Crommen. Validation of an automated method for the liquid chromatographic determination of atenolol in plasma: application of a new validation protocol. *Anal. Chim. Acta* **1999**, 391, 227.
- [20] E. Chapuzet, N. Mercier, S. Bervoas-Martin, B. Boulanger, P. Chevalier, P. Chiap, D. Grandjean, P. Hubert, P. Lagorce, M. Lallier, M.-Ch. Laparra, M. Laurentie, C. Nivet. Example of application of the new strategy proposed for the validation of chromatographic bioanalytical methods. *STP Pharma. Prat.* **2000**, 10, 79.
- [21] P. Chiap, A. Ceccato, B. Miralles Buraglia, B. Boulanger, P. Hubert, J. Crommen. Development and validation of an automated method for the liquid chromatographic determination of sotalol in plasma using dialysis and thrace enrichment on a cation-exchange pre-column as on-line sample preparation. *J. Pharm. Biomed. Anal.* **2001**, 24, 801.
- [22] B. Christiaens, P. Chiap, O. Rbeida, D. Cello, J. Crommen, P. Hubert. Fully automated method for the determination of cyproterone acetate in plasma using restricted access material for sample pre-treatment. *J. Chromatogr. B* **2003**, 795, 73.
- [23] O. Rbeida, B. Christiaens, P. Chiap, P. Hubert, D. Lubda, K. Boos, J. Crommen. Fully automated LC method for the determination of sotalol in human plasma using restricted access material with cation exchange properties for sample clean-up. *J. Pharm. Biomed. Anal.* **2003**, 32, 829.
- [24] A.C. Servais, M. Fillet, P. Chiap, W. Dewe, P. Hubert, J. Crommen. Determination of salbutamol enantiomers in human urine using heptakis(2,3-diacetyl-6-O-sulfo)-cyclodextrin in non-aqueous capillary electrophoresis. *Electrophoresis* **2004**, 25, 1632.
- [25] B. Christiaens, P. Chiap, M. Fillet, O. Rbeida, A. Ceccato, B. Streel, J.D. Graeve, J. Crommen, P. Hubert. New fully automated method for the LC-MS/MS determination of cyproterone acetate in human plasma using restricted access material for sample clean-up. *J. Chromatogr. A* **2004**, 1056, 105.
- [26] H.D. Hassonville, P. Chiap, J.F. Liegeois, B. Evrard, L. Delattre, J. Crommen, G. Piel, P. Hubert. Development and validation of a high-performance liquid chromatographic method for the determination of cyproterone acetate in human skin. *J. Pharm. Biomed. Anal.* **2004**, 36, 133.
- [27] E. Cavalier, E. Rozet, N. Dubois, C. Charlier, P. Hubert, J.-P. Chapelle, J.-M. Krzesinski, P. Delanaye. Performance of iohexol determination in serum and urine by HPLC: Validation, risk and uncertainty assessment. *Clin. Chim. Acta* **2008**, 396, 80.
- [28] E. Rozet, V. Wascotte, N. Lecouturier, V. Preat, W. Dewe, B. Boulanger, P. Hubert. Improvement of the decision efficiency of the accuracy profile by means of a desirability function for analytical methods validation - Application to a diacetyl-monoxime colorimetric assay used for the determination of urea in transdermal iontophoretic ex. *Anal. Chim. Acta* **2007**, 591, 239.
- [29] F. Boemer, V. Bours, R. Schoos, P. Hubert, E. Rozet. Analytical validation based on total error measurement and cut-off interpretation of a neonatal screening TSH-immunoassay. *J. Chromatogr. B* **2009**, 877, 2412.
- [30] M.A. Bimazubute, E. Rozet, I. Dizier, P. Gustin, P. Hubert, J. Crommen, P. Chiap. Liquid chromatographic determination of enrofloxacin in nasal secretions and plasma of healthy pigs using restricted access material for on-line sample clean-up. *J. Chromatogr. A* **2008**, 1189, 456.
- [31] B. Boulanger, E. Rozet, F. Moonen, S. Rudaz, P. Hubert. A risk-based analysis of the AAPS conference report on quantitative bioanalytical methods validation and implementation. *J. Chromatogr. B* **2009**, 877, 2235.
- [32] E. Rozet, B. Boulanger, S. Rudaz, R. Marini, E. Ziemons, P. Hubert. Total error for the validation of bioanalytical methods. *Ann. Toxicol. Anal.* **2009**, 21, 35.
- [33] C. Hubert, S. Houari, F. Lecomte, V. Houbart, C. De Bleye, M. Fillet, G. Piel, E. Rozet, P. Hubert. Development and validation of a sensitive solid phase extraction/hydrophilic interaction liquid chromatography/mass spectrometry method for the accurate determination of glucosamine in dog plasma. *J. Chromatogr. A* **2010**, 1217, 3275.
- [34] N. Dubois, B. Debrus, P. Hubert, C. Corinne. Validated quantitative simultaneous determination of cocaine opiates and amphetamines in serum by U-HPLC coupled to tandem mass spectrometry. *Acta Clin. Belg.* **2010**, 65, 75.
- [35] E. Staes, E. Rozet, B. Ucakar, P. Hubert, V. Préat. Validation of a method for the quantitation of ghrelin and unacylated ghrelin by HPLC. *J. Pharm. Biomed. Anal.* **2010**, 51, 633.
- [36] R.D. Marini, E. Rozet, C. Hubert, E. Ziemons, P. Hubert. Estimation of uncertainty from the total error strategy: Application to internal and normative methods. *Acta Clin. Belg.* **2010**, 65, 100.
- [37] E. Cavalier, E. Rozet, A. Carlisi, A.C. Bekaert, O. Rousselle, P. Hubert, J.P. Chapelle, P. Delanaye. Analytical validation of the BAP OSTASE on Liaison (DiaSorin). *Clin. Chem. Lab. Med.* **2010**, 48, 67.
- [38] R. Denooz, Z. Douamba, C. Charlie. Fatal intoxications by acenocumarol, phenprocoumon and warfarin: Method validation in blood using the total error approach. *J. Chromatogr. B* **2008**, 877, 2344.
- [39] E. Cheneau, J. Henri, Y. Pirotais, J.-P. Abjean, B. Roudaut, P. Sanders, M. Laurentie. Liquid chromatography-electrospray tandem mass spectrometric method for quantification of monensin in plasma and edible tissues of chicken used in pharmacokinetic studies: Applying a total error approach. *J. Chromatogr. B* **2007**, 850, 15.
- [40] P.E. De Pauw, R.B. Mackin, P. Goubert, C. Van Schravendijk, F.K. Gorusa. Total error profiling of a proinsulin time-resolved fluorescence immunoassay. *J. Chromatogr. B* **2008**, 877, 2403.

- [41] N. Gibelin, D. Dupont, S. Imbert, E. Rozet. Use of total error concept in the validation of viral activity in cell cultures. *J. Chromatogr. B* **2009**, *877*, 2407.
- [42] I. Marchi, J. Schappler, J.-L. Veuthey, S. Rudaz. Development and validation of a liquid chromatography-atmospheric pressure photoionization-mass spectrometry method for the quantification of alprazolam, flunitrazepam, and their main metabolites in haemolysed blood. *J. Chromatogr. B* **2009**, *877*, 2275.
- [43] B. Streeel, B. Cahay, R. Klinkenberg. Using total error concept for the validation of a liquid chromatography-tandem mass spectrometry method for the determination of budesonide epimers in human plasma. *J. Chromatogr. B* **2009**, *877*, 2290.
- [44] N. Ansermot, S. Rudaz, M. Brawand-Amey, S. Fleury-Souverain, J.-L. Veuthey, C.B. Eap. Validation and long-term evaluation of a modified on-line chiral analytical method for therapeutic drug monitoring of (R,S)-methadone in clinical samples. *J. Chromatogr. B* **2009**, *877*, 2301.
- [45] H. Kharbouche, F. Sporkert, S. Troxler, M. Augsburg, P. Mangin, C. Staub. Development and validation of a gas chromatography-negative chemical ionization tandem mass spectrometry method for the determination of ethyl glucuronide in hair and its application to forensic toxicology. *J. Chromatogr. B* **2009**, *877*, 2337.
- [46] M.A. Bimazubute, E. Rozet, I. Dizier, J.-Cl. Van Heugen, E. Arancio, P. Gustin, J. Crommen, P. Chiap. Pre-study and in-study validation of an ultra-high pressure LC method coupled to tandem mass spectrometry for off-line determination of oxytetracycline in nasal secretions of healthy pigs. *J. Chromatogr. B* **2009**, *877*, 2349.
- [47] S. Hambÿe, D. Stanicki, J.-M. Colet, E.M. Aliouat, J.J. Vanden Eynde, B. Blankert. Three optimized and validated (using accuracy profiles) LC methods for the determination of pentamidine and new analogs in rat plasma. *Talanta* **2011**, *83*, 832.
- [48] V. Varlet, E. Lagroy De Croutte, M. Augsburg, P. Mangin. Accuracy profile validation of a new method for carbon monoxide measurement in the human blood using Headspace-Gas chromatography-mass spectrometry (HS-GC-MS). *J. Chromatogr. B* **2011**, *880*, 125.
- [49] E. Rozet, R. Morello, F. Lecomte, G.B. Martin, P. Chiap, J. Crommen, K.S. Boos, P. Hubert. Performances of a multidimensional on-line SPE-LC-ECD method for the determination of three major catecholamines in native human urine: Validation, risk and uncertainty assessments. *J. Chromatogr. B* **2006**, *844*, 251.
- [50] R.D. Marini, A. Pantella, M.A. Bimazubute, P. Chiap, P. Hubert, J. Crommen. Optimisation and validation of a generic method for the LC assay of six corticosteroids and salicylic acid in dermatopharmaceutical forms. *Chromatographia* **2002**, *55*, 263.
- [51] R.D. Marini, A.C. Servais, E. Rozet, P. Chiap, B. Boulanger, S. Rudaz, J. Crommen, P. Hubert, M. Fillet. Nonaqueous capillary electrophoresis method for the enantiomeric purity determination of S-timolol using heptakis(2,3-di-O-methyl-6-O-sulfo)- $\beta$ -cyclodextrin: Validation using the accuracy profile strategy and estimation of uncertainty. *J. Chromatogr. A* **2006**, *1120*, 102.
- [52] R.D. Marini, P. Chiap, B. Boulanger, S. Rudaz, E. Rozet, J. Crommen, P. Hubert. LC method for the determination of R-timolol in S-timolol maleate: Validation of its ability to quantify and uncertainty assessment. *Talanta* **2006**, *68*, 1166.
- [53] J. Mantanus, E. Ziémons, P. Lebrun, E. Rozet, R. Klinkenberg, B. Streeel, B. Evrard, P. Huber. Moisture content determination of pharmaceutical pellets by near infrared spectroscopy: Method development and validation. *Anal. Chim. Acta* **2009**, *642*, 186.
- [54] J. Mantanus, E. Ziémons, P. Lebrun, E. Rozet, R. Klinkenberg, B. Streeel, B. Evrard, P. Hubert. Active content determination of non-coated pharmaceutical pellets by near infrared spectroscopy: Method development, validation and reliability evaluation. *Talanta* **2010**, *80*, 1750.
- [55] E. Ziémons, J. Mantanus, P. Lebrun, E. Rozet, B. Evrard, P. Hubert. Acetaminophen determination in low-dose pharmaceutical syrup by NIR spectroscopy. *J. Pharm. Biomed. Anal.* **2010**, *50*, 510.
- [56] S. Heuskin, E. Rozet, S. Lorge, J. Farmakidis, P. Hubert, F. Verheggen, E. Haubruge, J.P. Wathélet, G. Lognay. Validation of a fast gas chromatographic method for the study of semiochemical slow release formulations. *J. Pharm. Biomed. Anal.* **2010**, *53*, 962.
- [57] R.D. Marini, E. Rozet, M.L.A. Montes, C. Rohrbasser, S. Roht, D. Rhème, P. Bonnabry, J. Schappler, J.-L. Veuthey, P. Hubert, S. Rudaz. Reliable low cost capillary electrophoresis device for drug quality control and counterfeit medicines. *J. Pharm. Biomed. Anal.* **2010**, *53*, 1278.
- [58] J. Mantanus, E. Ziémons, E. Rozet, B. Streeel, R. Klinkenberg, B. Evrard, J. Rantanen, P. Hubert. Building the quality into pellet manufacturing environment - feasibility study and validation of an in-line quantitative near infrared (NIR) method. *Talanta* **2010**, *83*, 305.
- [59] T.R.M. De Beer, W.R.G. Baeyens, A. Vermeire, D. Broes, J.P. Remon, C. Vervaeet. Raman spectroscopic method for the determination of medroxyprogesterone acetate in a pharmaceutical suspension: Validation of quantifying abilities, uncertainty assessment and comparison with the high performance liquid chromatography reference method. *Anal. Chim. Acta* **2007**, *589*, 192.
- [60] S. Block, D. Brkic, P. Hubert, J. Quetin-Leclercq. A validated method for the quantification of pimarane and trachylobane diterpenes in the leaves of *Croton Zambesicus* by capillary gas chromatography. *Phytochem. Anal.* **2005**, *16*, 342.
- [61] C. Stevigny, M.C. Wautier, J.L. H. Jiwan, P. Chiap, Ph. Hubert, J. Quetin-Leclercq. Development and validation of a high-performance liquid chromatographic method for quantitative determination of aporphine alkaloids from different samples of *Cassytha filiformis* L. *Planta Med.* **2004**, *70*, 764.
- [62] V. Esters, L. Angenot, V. Brandt, M. Frédéric, M. Tits, C. Van Neruma, J.-N. Wauters, P. Hubert. Validation of a high-performance thin-layer chromatography/densitometry method for the quantitative determination of glucosamine in a herbal dietary supplement. *J. Chromatogr. A* **2006**, *1112*, 156.
- [63] E. Ziémons, V. Barillaro, E. Rozet, N. Wandji Mbakop, R. Lejeune, L. Angenot, L. Thunus, P. Hubert. Direct determination of tagitinin C in *Tithonia diversifolia* leaves by on-line coupling of supercritical carbon dioxide extraction to FT-IR spectroscopy by means of optical fibres. *Talanta* **2007**, *71*, 911.
- [64] D.S. Salome Kpoviessi, F. Bgaguidi, J. Gbenou, G. Accrombessi, M. Moudachirou, E. Rozet, P. Hubert, J. Quetin-Leclercq. Validation of a method for the determination of sterols and triterpenes in the aerial part of *Justicia anselliana* (Nees) T. Anders by capillary gas chromatography. *J. Pharm. Biomed. Anal.* **2008**, *48*, 1127.
- [65] M.H. Rafamantanana, E. Rozet, G.E. Raoelison, K. Cheuk, S.U. Ratsimamanga, P. Hubert, J. Quetin-Leclercq. An improved HPLC-UV method for the simultaneous quantification of triterpenic glycosides and aglycones in leaves of *Centella asiatica* (L.) Urb (APIACEAE). *J. Chromatogr. B* **2009**, *877*, 2396.
- [66] B. De Backer, B. Debrus, P. Lebrun. Innovative development and validation of an HPLC/DAD method for the qualitative and quantitative determination of major cannabinoids in cannabis plant material. *J. Chromatogr. B* **2009**, *877*, 4115.
- [67] W. Dewe, B. Boulanger, B. Govaerts, E. Rozet, P. Hubert. Approches par l'erreur totale en transfert analytique. *Ann. Toxicol. Anal.* **2006**, *18*, 59.
- [68] E. Rozet, W. Dewe, P. Chiap, F. Lecomte, P. Hubert. Le transfert d'une méthode de dosage automatisée de la noradrénaline dans l'urine humaine: Utilisation de l'erreur totale comme critère de décision. *Acta Clin. Belg.* **2006**, *61*, 57.
- [69] E. Rozet, B. Mertens, W. Dewe, A. Ceccato, B. Govaerts, B. Boulanger, P. Chiap, B. Streeel, J. Crommen, P. Hubert. The transfer of a LC-UV method for the determination of fenofibrate and fenofibric acid in Lidoses: Use of total error as decision criterion. *J. Pharm. Biomed. Anal.* **2006**, *42*, 64.
- [70] W. Dewé, B. Govaerts, B. Boulanger, E. Rozet, P. Chiap, P. Hubert. Using total error as decision criterion in analytical method transfer. *Chemometr. Intell. Lab. Syst.* **2007**, *85*, 262.
- [71] E. Rozet, W. Dewé, R. Morello, P. Chiap, F. Lecomte, E. Ziémons, K.S. Boos, B. Boulanger. Risk-based approach for the transfer of quantitative methods: Bioanalytical Applications. *J. Chromatogr. A* **2008**, *1189*, 32.
- [72] E. Rozet, W. Dewé, E. Ziémons, A. Bouklouze, B. Boulanger, P. Hubert. Methodologies for the transfer of analytical methods: A review. *J. Chromatogr. B* **2009**, *877*, 2214.
- [73] M. Laurentie, V. Gaudin. Use of the total error approach to evaluate the performance of a semi-quantitative immunological method (BIACORE method) for detecting sulfamethazine in bovine milk. *J. Chromatogr. B* **2009**, *877*, 2375.
- [74] M. Feinberg, J. San-Redon, A. Assié. Determination of complex polysaccharides by HPAE-PAD in foods: Validation using accuracy profile. *J. Chromatogr. B* **2009**, *877*, 2388.
- [75] M. Feinberg. Validation of analytical methods based on accuracy profiles. *J. Chromatogr. A* **2007**, *1158*, 174.



- [76] V. Gaudin, M. Laurentie. Application of total error approach to assess the performance of a biological method (ELISA) to detect nicarbazin residues in eggs. *J. Chromatogr. B* **2009**, *877*, 2358.
- [77] C. Marlet, G. Lognay. Development and validation by accuracy profile of a method for the analysis of monoterpenes in indoor air by active sampling and thermal desorption-gas chromatography-mass spectrometry. *Talanta* **2010**, *82*, 1230.
- [78] C. Hubert, E. Ziémons, E. Rozet, A. Breuer, A. Lambert, C. Jasselette, C. De Bleye, R. Lejeune, P. Hubert. Development and validation of a quantitative method for the selective determination of tin species in tin octoate by differential pulse polarography. *Talanta* **2010**, *80*, 1413.
- [79] C. Hartmann, J. Smeyers-Verbeke, D.L. Massart, R.D. McDowall. Validation of bioanalytical chromatographic methods. *J. Pharm. Biomed. Anal.* **1998**, *17*, 193.
- [80] J. Caporal-Gautier, J.M. Nivet, P. Algranti, M. Guilloteau, M. Histe, L. Lallier, J.J. N'Guyen-Huu, R. Russotto. Guide de validation analytique: Rapport d'une commission SFSTP I. méthodologie. *STP Pharma. Prat.* **1992**, *2*, 205.
- [81] D.J. Schuirmann. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J. Pharmacokinet. Biopharm.* **1987**, *15*, 657.
- [82] C. Hartmann, J. Smeyers-Verbeke, W. Penninckx, Y. Vander Heyden, P. Vankeerberghen, D.L. Massart. Reappraisal of hypothesis testing for method validation: Detection of systematic error by comparing the means of two methods or of two laboratories. *Anal. Chem.* **1995**, *67*, 4491.
- [83] R. Kringle, R. Khan-Malek, F. Snikeris, P. Munden, C. Agut, M. Bauer. A unified approach for design and analysis of transfer studies for analytical methods. *Drug Inf. J.* **2001**, *35*, 1271.
- [84] C. Hartmann, D.L. Massart, R.D. McDowall. An analysis of the Washington conference report on bioanalytical method validation. *J. Pharm. Biomed. Anal.* **1994**, *12*, 1337.
- [85] Anonymous, European Pharmacopoeia (Ph. Eur.), 7th Edition. Strasbourg, France: Council of Europe; **2011**.

## **CHAPITRE VI : Utilité des indices de capabilité pour la validation des méthodes analytiques**

*Ce chapitre correspond à l'article « Usefulness of capability indices in the framework of analytical methods validation » publié dans le journal *Analytica Chimica Acta*, Volume 714, 10 February 2012, Pages 47-56*

- 1. Synthèse et conclusion**
- 2. Article**

## 1. Synthèse et conclusion

Comparées à un processus de production, les méthodes analytiques peuvent être considérées comme des processus de petite taille. Dans les activités industrielles, il est généralement nécessaire d'obtenir des informations sur la capabilité des processus mesurant la performance du processus lors de son fonctionnement sous contrôle statistique.

Cette capabilité représente la capacité d'un procédé de mesure à respecter en permanence *un niveau de spécification* à atteindre. Elle a été initialement introduite pour les procédés de fabrication industriels, pour lesquels il existe une spécification définie a priori, comme un écart de poids maximum acceptable sur une ensacheuse. Cependant, elle peut aussi s'appliquer avec succès à des procédures d'analyse chimique.

Dans sa forme la plus intuitive, la capabilité se définit comme le rapport :

$$\text{Capabilité} = \frac{\text{Intervalle d'acceptabilité}}{\text{Dispersion}}$$

Des indices de capabilité sont calculés afin d'évaluer si le processus étudié est en mesure de fournir suffisamment d'unités conformes. Par analogie, les indices de capabilité pourraient être utilisés afin d'évaluer si la méthode analytique est en mesure de fournir suffisamment de résultats exacts et de savoir ainsi si la méthode est adaptée à son usage prévu. De plus, la capabilité des méthodes analytiques peut également être utilisée pour évaluer les risques inhérents au transfert des méthodes analytiques afin d'augmenter ou de diminuer la charge de travail du transfert de méthode.

Dans les ouvrages ou les normes consacrées au calcul de la capabilité, on distingue notamment deux principaux indices de capabilité : Cp et Cpk :

$$C_p = \frac{USL - LSL}{6\sigma} \quad \text{Eq.1}$$

$$C_{pk} = \min \left[ \frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma} \right] \quad \text{Eq.2}$$

LSL et USL sont respectivement les limites de spécification supérieures et inférieures du processus.  $\mu$  et  $\sigma$  sont la moyenne et l'écart-type du processus étudié. Il est évident que les méthodes analytiques sont sujettes à des déviations systématiques. L'indice Cp est donc peu utile car il suppose que la méthode analytique étudiée soit centrée sur la concentration de référence ou nominale de l'échantillon étudié.

Les valeurs Cpk obtenues à partir des expériences pourraient être comparées à des valeurs minimales consensuelles telles que:

- $Cpk \geq 1.0$
- $Cpk \geq 1.33$
- $Cpk \geq 2$

Par exemple, l'Annexe 2 du chapitre <1080> de la pharmacopée américaine (USP) indique que les valeurs Cpk supérieures à 1,33 montrent que le processus est adéquat pour répondre aux spécifications. Les valeurs comprises entre 1,00 et 1,33 indiquent que le processus, bien que suffisant pour répondre aux spécifications, a besoin d'un contrôle étroit. Les valeurs inférieures à 1,00 indiquent que le processus n'est pas adéquat pour répondre aux spécifications et que le processus et / ou les spécifications doivent être changés.

L'évaluation de la capabilité des méthodes analytiques peut être une méthodologie utile pour évaluer la validité de ces méthodes d'analyses. Toutefois, des précautions sur la façon de calculer les indices de capabilité doivent être prises. En effet, les formules couramment utilisées pour calculer des indices tels que la Cpk, vont fortement surestimer la capabilité réelle de ces méthodes. Surtout lors de la validation de méthodes ou de leur transfert, où seulement quelques expériences sont réalisées. L'utilisation dans ces situations des indices de capabilité courant pour déclarer une méthode comme valide ou transférable à un laboratoire receveur conduira à des décisions inadéquates.

Le but de ce travail est donc d'améliorer le calcul des indices de capabilité afin de permettre leur utilisation efficace dans les situations où de petites tailles d'échantillon sont produites, par exemple, lors de l'évaluation de la validation ou du transfert des méthodes analytiques.

Donc, un indice de capabilité amélioré, nommé CPK-tol et l'estimateur correspondant de la proportion de résultats non conformes ( $\pi$  Cpk-tol) ont été proposés. À travers des simulations de Monte-Carlo, il a été démontré l'utilisation efficace de ces indices dans l'estimation de la capabilité dans les situations de faible taille d'échantillon rencontrées lors de l'évaluation de la validité ou la transférabilité des méthodes analytiques. De plus, l'utilité de ces indices de capabilité a été illustrée par plusieurs études de cas couvrant les applications couramment rencontrées dans l'industrie pharmaceutique.

Enfin, une méthodologie pour définir la taille d'échantillon optimale nécessaire pour valider les méthodes analytiques est aussi donnée en utilisant comme métrique la capabilité.

## **2. Article**

### **Usefulness of capability indices in the framework of analytical methods validation<sup>2</sup>**

*Analytica Chimica Acta, Volume 714, 10 February 2012, Pages 47-56*

A. Bouabidi, E. Ziemons, R. Marini, C. Hubert, M. Talbi, A. Bouklouze, H. Bourichi, M. El Karbane, B. Boulanger, Ph. Hubert, E. Rozet,



## Usefulness of capability indices in the framework of analytical methods validation

A. Bouabidi<sup>a,b</sup>, E. Ziemons<sup>a</sup>, R. Marini<sup>a</sup>, C. Hubert<sup>a</sup>, M. Talbi<sup>b</sup>, A. Bouklouze<sup>c</sup>, H. Bourichi<sup>c</sup>, M. El Karbane<sup>c</sup>, B. Boulanger<sup>d</sup>, Ph. Hubert<sup>a,1</sup>, E. Rozet<sup>a,\*,1,2</sup>

<sup>a</sup> Analytical Chemistry Laboratory, CIRMA, Institute of Pharmacy, University of Liège, Liège, Belgium

<sup>b</sup> Analytical Chemistry Laboratory, University Hassan II – Mohammedia – Faculty of Sciences Ben M'Sik, Casablanca, Morocco

<sup>c</sup> Pharmacology-Toxicology Laboratory, Faculty of Medicine and Pharmacy, University Med V Soussi, Rabat, Morocco

<sup>d</sup> Arlenda SA, Liège, Belgium

### ARTICLE INFO

#### Article history:

Received 2 June 2011

Received in revised form 3 October 2011

Accepted 8 November 2011

Available online 6 December 2011

#### Keywords:

Capability

Tolerance intervals

Drug products

Method validation

Fit for purpose

### ABSTRACT

Analytical methods capability evaluation can be a useful methodology to assess the fitness of purpose of these methods for their future routine application. However, care on how to compute the capability indices have to be made. Indeed, the commonly used formulas to compute capability indices such as  $Cpk$ , will highly overestimate the true capability of the methods. Especially during methods validation or transfer, there are only few experiments performed and, using in these situations the commonly applied capability indices to declare a method as valid or as transferable to a receiving laboratory will conduct to inadequate decisions.

In this work, an improved capability index, namely  $Cpk-tol$  and the corresponding estimator of proportion of non-conforming results ( $\pi_{Cpk-tol}$ ) have been proposed. Through Monte-Carlo simulations, they have been shown to greatly increase the estimation of analytical methods capability in particular in low sample size situations as encountered during methods validation or transfer. Additionally, the usefulness of this capability index has been illustrated through several case studies covering applications commonly encountered in the pharmaceutical industry. Finally a methodology to determine the optimal sample size required to validate analytical methods is also given using the proposed capability metric.

© 2011 Elsevier B.V. All rights reserved.

### 1. Introduction

Compared to a production process, methods of analysis can be considered as shorter processes. In industrial activities, it is usually necessary to obtain some information about the process capability, that is, the performance of the process when it is operating in statistical control. Capability indices are then computed in order to evaluate whether the process under study is able to provide sufficient conforming units. By analogy capability indices could therefore be used in order to evaluate whether the analytical method only is able to provide enough conforming results and check if a method is fitted for its intended purpose. For instance, Dejaegher et al. [1], discussed the capability of HPLC and titration assays for the analysis of drug substances. Kamberi et al. [2], have

used capability indices to give recommendations for assays precision and trueness required to validate analytical methods of drug eluting stents. Additionally analytical methods capability can also be used to evaluate the risks inherent to analytical methods transfers in order to increase or decrease the method transfer workload [3]. Two main capability indices are commonly used assuming that the process is centred at the nominal dimension ( $Cp$ ) or not ( $Cpk$ ) [4,5]:

$$Cp = \frac{USL - LSL}{6\sigma} \quad (1)$$

$$Cpk = \min \left[ \frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma} \right] \quad (2)$$

where LSL and USL are the lower and upper specifications for the process, respectively.  $\mu$  and  $\sigma$  are the mean and standard deviation of the process under study.  $Cp$  is the most used capability index. It aims at measuring if the dispersion of the studied variable is more or less larger than the interval defined by the process specification limits. It measures the percentage of the specification interval used by the variable measured. However this index does not give any information about the position of the variable with respect to a target value. While it is evident that analytical methods are prone

\* Corresponding author at: Laboratory of Analytical Chemistry, Institute of Pharmacy, Université de Liège, CHU, B 36, B-4000 Liège, Belgium. Tel.: +32 4 3664320; fax: +32 4 3664317.

E-mail address: [Eric.Rozet@ulg.ac.be](mailto:Eric.Rozet@ulg.ac.be) (E. Rozet).

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> F.R.S.-FNRS Postdoctoral Researcher (Belgium).

to systematic deviations, the  $C_p$  index is useless as it assumes that the analytical method under study is centred at the reference or nominal concentration of the sample being studied.

It would be thus tempting to use the  $C_{pk}$  index to evaluate whether analytical methods are fit for their purpose and thus declare them as valid using  $C_{pk}$  values. Indeed, the  $C_{pk}$  index measures simultaneously the position and the dispersion of the variable measured. For instance, the  $C_{pk}$  values obtained from experiments could be compared to consensual minimum values such as [6]:

- $C_{pk} \geq 1.0$
- $C_{pk} \geq 1.33$
- $C_{pk} \geq 2$

For instance, the USP chapter <1080> appendix 2 expresses that  $C_{pk}$  values “exceeding 1.33 show that the process is adequate to meet specifications. Values between 1.00 and 1.33 indicate that the process, although adequate to meet specifications, will require close control. Values below 1.00 indicate that the process is not adequate to meet specifications and that the process and/or specifications should be changed” [7].

Nonetheless to compute these capability indices some key assumptions are made:

1. The data are distributed following a normal distribution.
2. The values of  $\mu$  and  $\sigma$  are supposed to be known.

Normality of the data is usually assumed and based on the expertise of the operator. Normality can also be assessed through dedicated graphs (e.g. Q–Q plots) or using statistical tests (e.g. Kolmogorov-Smirnov or Chi-square tests, . . .). However they require a large number (more than 30 results [8]) of data to be really effective. Analytical method validation studies generate far lesser amount of data. Classical interpretation of ICH Q2 guideline [9] will provide 9 results per concentration levels (3 replicates during 3 series), for the FDA guidance on bioanalytical methods validation [10] it usually leads to a minimum of 15 results per concentration level (5 replicates during 3 series). These small amounts of data generated have thus two implications that are closely linked. The first one is that a formal evaluation of normality using these data is useless or only a first gross evaluation. The second is that the other assumption made when computing capability indices is not met: the values of the mean ( $\mu$ ) and standard deviation ( $\sigma$ ) of the method are not known at the end of a validation study. Computing the  $C_{pk}$  index as given in Eq. (2) is only a poor estimate of the real capability of the method. Consequently, the fractions of non-conforming results  $\pi_{C_{pk}}$  falling outside the specifications that are estimated using these indices are only a first approximation:

$$\begin{aligned} \pi_{C_{pk}} &= P(X < LSL) + P(X > USL) \\ &= P\left(Z < \frac{LSL - \mu}{\sigma}\right) + P\left(Z > \frac{USL - \mu}{\sigma}\right) \end{aligned} \quad (3)$$

where  $Z \sim N(0, 1)$

The aim of this work is thus to enhance the computation of capability indices in order to allow their efficient use in small sample size situations occurring for example when assessing the validity or transferability of analytical methods. First the rationale and formulas to compute the proposed enhanced capability index  $C_{pk-tol}$  as well as the corresponding fraction of non-conforming results  $\pi_{C_{pk-tol}}$  will be presented. Then Monte-Carlo simulations will show the improvements realised. The usefulness of these developments will be illustrated through the validation of two original analytical methods as well as using validation data available from previously published articles. Finally a methodology to define the optimal

sample size required to validate analytical methods is also given using the proposed capability metric.

## 2. Materials and methods

### 2.1. Fluconazole

#### 2.1.1. Chemicals and reagents

All chemicals and solvents used were of analytical or HPLC grade. Fluconazole was supplied by the European Pharmacopoeia (Strasbourg, France). Potassium dihydrogen phosphate was purchased from Acros Organics (Geel, Belgium). Methanol and phosphoric acid (85%) were obtained from Merck (Darmstadt, Germany). Deionized water was generated from the Milli-Q water purifying system (Millipore, Watford, UK). Phosphate buffer was adjusted to pH 3.6 with phosphoric acid.

#### 2.1.2. Instrumentation and chromatographic conditions

Analyses were performed on an Agilent technologies HPLC 1100 series system (Hewlett-Packard, Palo-Alto, CA, USA) equipped with a solvent delivery quaternary pump G1311A, an on-line degasser G1322A, an autosampler G1313A, a column oven G1316A and a diode-array detector G1315A.

Chromatographic analysis was performed on a C18 column ( $250 \times 4.6$  mm i.d.,  $5 \mu\text{m}$  particle size) and kept at  $25^\circ\text{C}$ . The mobile phase was prepared by mixing separately measured methanol and phosphate buffer (pH 3.6) in a ratio 40:60 v/v and was degassed before use. The HPLC system was operated isocratically at a flow rate of  $1.0 \text{ mL min}^{-1}$  and the injection volume was  $20 \mu\text{L}$ . UV detection was performed at 261 nm.

#### 2.1.3. Sample preparation

A stock solution of fluconazole was prepared by accurately weighting 100.0 mg of fluconazole and diluting this in 100.0 mL of methanol. The calibration standards for fluconazole were prepared by diluting an appropriate volume of stock solution with 0.1 M HCl aqueous solution to reach the concentration levels corresponding to 125% of fluconazole. The standards were analysed in duplicate.

The validation standards were prepared similarly with addition of the excipients of the formulation of the drug product to reach four concentration levels ranging from 25% to 125% of fluconazole, i.e. 12.67, 25.52, 50.52 and 62.30 mg of fluconazole. Each validation standard was analysed four times. The validation study was performed on six different series.

### 2.2. Acetazolamide (ACZ) determination in tablets by HPLC

#### 2.2.1. Chemicals and reagents

All chemicals and solvents used were of analytical or HPLC grade. Acetazolamide (ACZ) was supplied by the European Pharmacopoeia. Triethylamine was purchased from Sigma-Aldrich (Steinheim, Germany). Acetonitrile and phosphoric acid (85%) were obtained from Merck. Deionised water was generated from the Milli-Q water purifying system (Millipore).

#### 2.2.2. Instrumentation and chromatographic conditions

The HPLC system was the same as in Section 2.1. Chromatographic analysis was achieved using a Lichrospher ODS C18 ( $250 \times 4.0$  mm i.d.,  $5 \mu\text{m}$  particle size; Agilent, Palo-Alto, CA, USA) column. The isocratic chromatographic separation was carried out using a mobile phase consisting of a mixture of 0.01 M phosphate acid buffer solution (0.676 g of phosphate acid in 1000 mL of water) adjust with triethylamine to a pH of 7.1 and acetonitrile in a proportion 85:15 (v/v), pumped at a constant flow rate of  $1.0 \text{ mL min}^{-1}$ . The column was maintained at  $25^\circ\text{C}$  and UV



absorbance was monitored at a wavelength of 266 nm. The sample injection volume was 20  $\mu\text{L}$ .

### 2.2.3. Sample preparation

The stock solutions of ACZ were prepared by weighing 100 mg, transferred to individual 100 mL volumetric flasks and diluted to volume with the mobile phase. The stock solutions were stored at 2–8 °C protected from light.

The calibration standards for ACZ were prepared by diluting an appropriate volume of stock solution with the mobile phase to reach the concentration level corresponding to 150% of ACZ. These calibration standards were analysed in triplicate. The validation standards were prepared similarly with addition of the excipients of the formulation of the drug product to reach five concentration levels ranging from 25% to 150% of ACZ reaching 0.05, 0.10, 0.15, 0.20 and 0.30  $\mu\text{g mL}^{-1}$  of ACZ. Each validation standard was analysed three times. The validation study was performed on three different days.

### 2.3. Simulations

Simulations were performed in order to estimate the performances of the different methodologies to compute the capability indices as well as the exact fraction of non-conforming results. The computations of capability indices were made according to Eq. (2), as well as the proposed modified  $Cpk\text{-tol}$  indices. The fractions of non-conforming results using each of these indices were also computed.

To achieve this, independent validation results were generated from the ANOVA I model described below:

$$X_{ij} = \delta + \phi_B + \varepsilon_W \quad (4)$$

where  $X_{ij}$  is the result of the  $j$ th measurement in series  $i$ ,  $\delta = \mu_{\text{Lab}} - \mu_T$  is the bias between the true (or reference or nominal) value of the result ( $\mu_T$ ) and the average value of the results of the laboratory ( $\mu_{\text{Lab}}$ ),  $\phi_B$  is the between series random effect supposed to be normally distributed  $N(0, \sigma_B^2)$  and  $\varepsilon_W$  is the within-series (or repeatability) random error supposed to be independent and normally distributed  $N(0, \sigma_W^2)$ .

$\mu_T$ , the true value was fixed at 0. Simulations were performed for analytical methods with biases ranging from –5 to +5% and for intermediate precision relative standard deviation ranging from 0% to 5%. The ratio  $R$  corresponding to the series-to-series variance over the within-series (or repeatability) variance was constant and set to  $R = 1$ . The number of replicates was fixed at  $I = 3$  and the number of series  $J$  used were 3 and 5. For each situation 2000 simulations were performed. The specification limits or acceptance limits were fixed at  $\pm 5\%$  around the true or nominal value as commonly adopted for the release of pharmaceutical drug products. The region of methods with true  $Cpk$  of 1, 1.33 and 2 are also computed.

### 2.4. Computations

All simulations were performed with R v2.9.1 (CRAN, <http://cran.r-project.org>). The general validation criteria were obtained using e-nova V3.0 (Arlenda, Liege, Belgium).

## 3. Results

### 3.1. Modified capability indices and fraction of non-conforming results

The core problem when using capability indices in validation studies is the lack of sufficient data to estimate precisely the mean and standard deviation of the analytical method. In order to circumvent this and to take into account the uncertainty of the analytical

method mean and standard deviation when computing  $Cpk$ , the use of tolerance interval should be preferred. Tolerance intervals allow describing the entire population of individual results based on a sample of the random variable under study (e.g. concentration of an analyte) [11–13].  $\beta$ -expectation tolerance intervals, also known as “mean coverage tolerance intervals” or “prediction intervals” are intervals for which each future results has  $\beta$  (e.g.  $\beta = 95\%$ ) probability to fall within the computed interval  $[L; U]$  [14]. Additionally, the estimation of the mean ( $\hat{\mu}$ ) and standard deviation ( $\hat{\sigma}_{IP}$ ) of analytical methods should be made following the statistical model representing the way experiments have been performed. Method validation experiments or method transfer experiments must be performed by including several series (or runs) as well as repetitions within these series to obtain reliable estimations of these parameters [11,12,15,16]. These trials are thus following a hierarchical or stratified sampling scheme that should be taken into account when computing analytical mean results and standard deviation and therefore to compute capability indices. For these stratified random sampling schemes commonly encountered during methods validations or transfers, a  $\beta$ -expectation tolerance intervals formula is given by Mee [17]:

$$[L, U] = [\hat{\mu} - k_E \hat{\sigma}_{IP}; \hat{\mu} + k_E \hat{\sigma}_{IP}] \quad (5)$$

where

$$k_E = t_{(df, (1+\beta)/2)} \sqrt{1 + \frac{J\hat{R} + 1}{J(\hat{R} + 1)}} \quad (6)$$

$$\text{with } df = \frac{(\hat{R}+1)^2}{(\hat{R}+1/J)^2 / J - 1 + (1-(1/I)/J)} \quad \text{and} \quad \hat{R} = \frac{\hat{\sigma}_B^2}{\hat{\sigma}_W^2}$$

$t_{(df, \gamma)}$  is the  $\gamma$ th percentile of a Student distribution with  $df$  degrees of freedom and  $\hat{\mu}$  is the estimated mean of the results. The intermediate precision variance can be estimated using:  $\hat{\sigma}_{IP}^2 = \hat{\sigma}_B^2 + \hat{\sigma}_W^2$ .  $\hat{\sigma}_B^2$  is the run-to-run or series-to-series variance and  $\hat{\sigma}_W^2$  is the within-run or repeatability variance obtained with random one way analysis of variance (ANOVA) methodology [18].  $J$  is the number of series performed and  $I$  the number of replicates per series.

The modified capability index proposed,  $Cpk\text{-tol}$ , is thus based on these tolerance intervals and is computed as it follows:

$$Cpk\text{-tol} = \min \left[ \frac{USL - \hat{\mu}}{t_{(df, (1+\beta)/2)} \sqrt{1 + \frac{J\hat{R} + 1}{J(\hat{R} + 1)}} \hat{\sigma}_{IP}}, \frac{\hat{\mu} - LSL}{t_{(df, (1+\beta)/2)} \sqrt{1 + \frac{J\hat{R} + 1}{J(\hat{R} + 1)}} \hat{\sigma}_{IP}} \right] \quad (7)$$

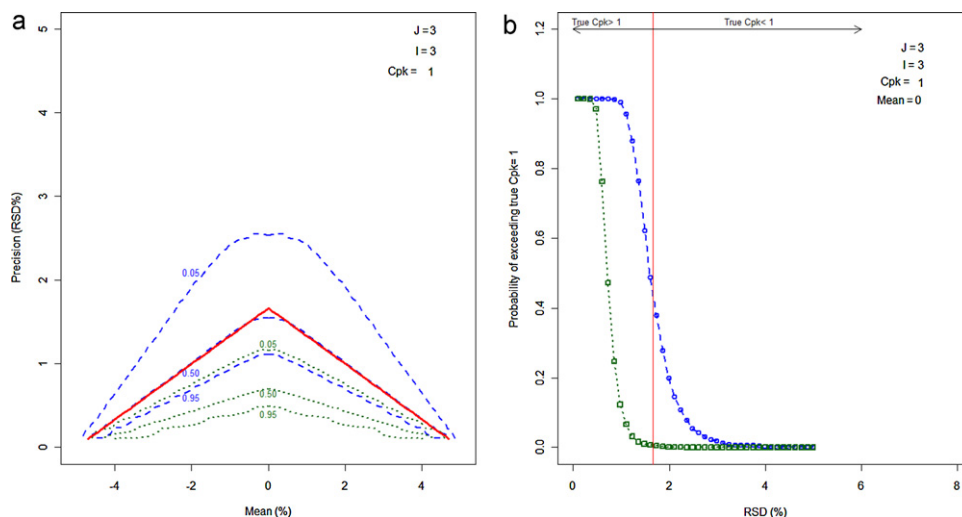
The  $Cpk$  index is computed with  $3\sigma$  at the denominator meaning that for a centred process the maximum fraction of non-conforming result is about 2700 dpm (precisely 2699.796 dpm). In order to keep this same theoretical coverage of the distribution used with the  $Cpk$  index (i.e.  $\pm 3\sigma$ ), the probability  $\beta$  of the  $Cpk\text{-tol}$  index is fixed to 0.9973.

Then the fraction of non-conforming results can be estimated by rearrangement of the expression in Eq. (5):

$$\pi_{Cpk\text{-tol}} = P[X_i < LSL] + P[X_i > USL] \\ = P \left[ t(df) < \frac{LSL - \hat{\mu}}{\hat{\sigma}_{IP} \sqrt{1 + \frac{J\hat{R} + 1}{J(\hat{R} + 1)}}} \right] + P \left[ t(df) > \frac{USL - \hat{\mu}}{\hat{\sigma}_{IP} \sqrt{1 + \frac{J\hat{R} + 1}{J(\hat{R} + 1)}}} \right] \quad (8)$$

where  $t(df)$  is a student distribution with  $df$  degrees of freedom.

$Cpk\text{-tol}$  and  $\pi_{Cpk\text{-tol}}$  therefore take into account the statistical model underlying the experiments performed during methods validation (or transfer) as well the uncertainty of the parameters. The uncertainty of the mean ( $\hat{\mu}$ ) can be obtained by simple mathematical operations and rearrangements of  $\hat{\sigma}_{IP} \sqrt{1 + (J\hat{R} + 1)/J(\hat{R} + 1)}$  to obtain the standard error of the mean  $\sqrt{I\hat{\sigma}_B^2 + \hat{\sigma}_W^2}/J$  [19]. While



**Fig. 1.** Performance of the capability indices  $Cpk$  and  $Cpk-tol$  for a true  $Cpk$  value of 1 and an experimental design of 3 runs and 3 repetitions. (a) Isoprobability contour measuring the probability that  $Cpk$  (dashed curves) and  $Cpk-tol$  (dotted curves) exceeds the true  $Cpk$  value of 1. (b) Operating characteristic curves obtained for the case of a centred analytical method (mean = 0%) depicting the probability that  $Cpk$  (dashed curves with open circles) and  $Cpk-tol$  (dotted curves with open squares) exceeds the true  $Cpk$  value of 1, the vertical line delimits methods with true  $Cpk > 1$  and true  $Cpk < 1$ .

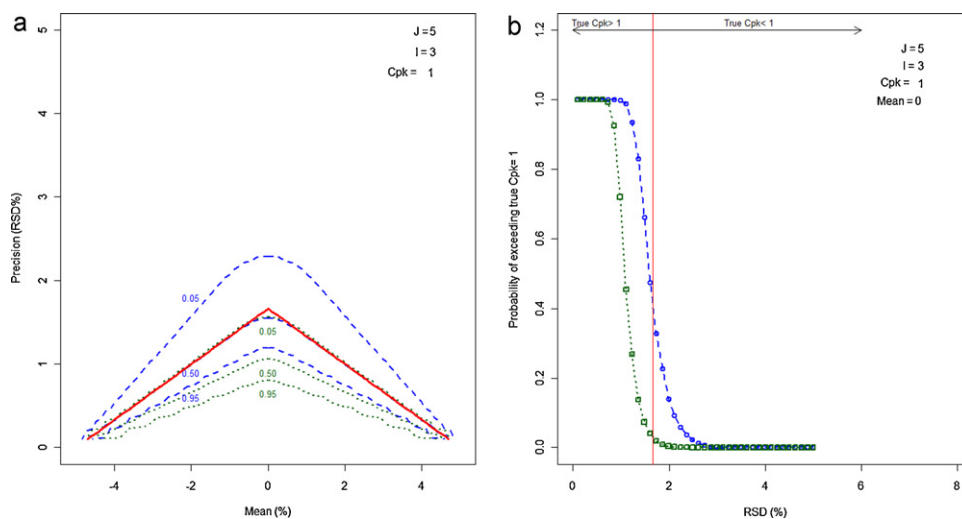
the uncertainty of the intermediate precision variance ( $\hat{\sigma}_{IP}$ ) is taken into account by the student distribution  $t(df)$  [20].

### 3.2. Simulations results

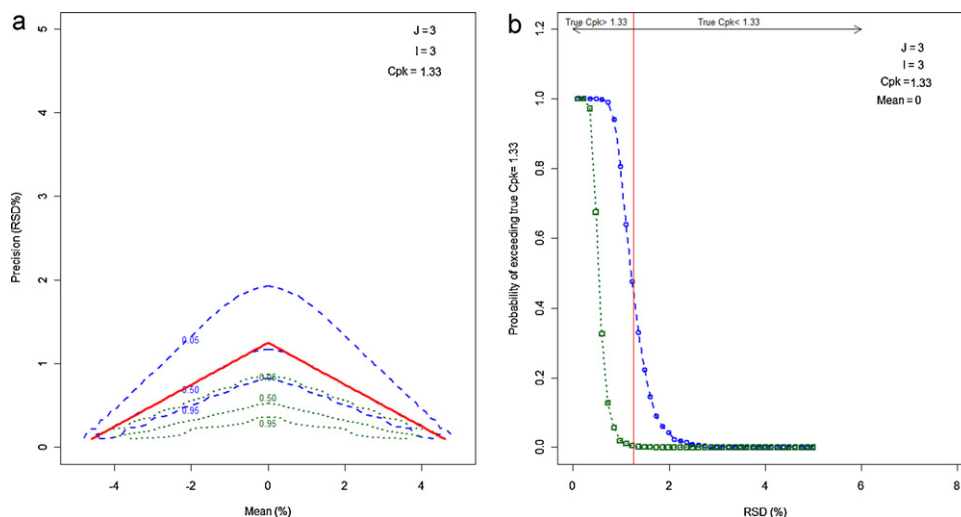
The results of the simulation study are given in Figs. 1–6 for the known capability values of 1, 1.33 and 2 for each of the combination of series and repetitions per series. In this discussion, we will focus mainly on the ability of the  $Cpk$  and  $Cpk-tol$  indices to control or manage the customer risk, i.e. the risk to conclude erroneously that the capability of the method is better than its true capability index. From Figs. 1–6a, it can be seen that the probability to exceed the true capability value (defined by the triangle in continuous line of Figs. 1–6a) by using Eq. (2) for computation of the  $Cpk$  index is almost 50%, whatever the true capability value and whatever the sample size used in the method validation. Indeed the isoprobability

curve (dashed line of Figs. 1–6a) that is almost exactly on the region which defines methods with known true  $Cpk$  value is the isoprobability curve of 50%. Figs. 1–6b give operating characteristic curves of each situation tested for methods with known biases of 0% and in function of their intermediate precision variance. The vertical line defines the region of methods with true capability indices of 1, 1.33 or 2 in Figs. 1–6b, respectively. It can be seen, for this case that the probability to exceed the true capability value is almost 50% in each simulation case studied.

By opposition when using Eq. (7) to compute  $Cpk-tol$ , Figs. 1–6a show that the probability to exceed the true capability value is extremely low as the closest isoprobability curve to the region defining methods with true capability indices of 1, 1.33 or 2 is the 10% one. Therefore there is about only 10% probability to declare a method capable when in reality it is not, i.e. the customer risk is about 10% using such a capability index compared to the 50% risk



**Fig. 2.** Performance of the capability indices  $Cpk$  and  $Cpk-tol$  for a true  $Cpk$  value of 1 and an experimental design of 5 runs and 3 repetitions. (a) Isoprobability contour measuring the probability that  $Cpk$  (dashed curves) and  $Cpk-tol$  (dotted curves) exceeds the true  $Cpk$  value of 1. (b) Operating characteristic curves obtained for the case of a centred analytical method (mean = 0%) depicting the probability that  $Cpk$  (dashed curves with open circles) and  $Cpk-tol$  (dotted curves with open squares) exceeds the true  $Cpk$  value of 1, the vertical line delimits methods with true  $Cpk > 1$  and true  $Cpk < 1$ .



**Fig. 3.** Performance of the capability indices  $Cpk$  and  $Cpk\text{-tol}$  for a true  $Cpk$  value of 1.33 and an experimental design of 3 runs and 3 repetitions. (a) Isoprobability contour measuring the probability that  $Cpk$  (dashed curves) and  $Cpk\text{-tol}$  (dotted curves) exceeds the true  $Cpk$  value of 1.33. (b) Operating characteristic curves obtained for the case of a centred analytical method (mean = 0%) depicting the probability that  $Cpk$  (dashed curves with open circles) and  $Cpk\text{-tol}$  (dotted curves with open squares) exceeds the true  $Cpk$  value of 1.33, the vertical line delimits methods with true  $Cpk > 1.33$  and true  $Cpk < 1.33$ .

observed for the classical  $Cpk$  index. Figs. 1–6b show this behaviour for the special case of truly unbiased methods (relative bias = 0%). Indeed, these figures show that the operating characteristic curve of  $Cpk\text{-tol}$  (dotted curve) crosses the line defining analytical methods with truly 1, 1.33 or 2 capability values (vertical line) around values of probability that are all lower than 10%.

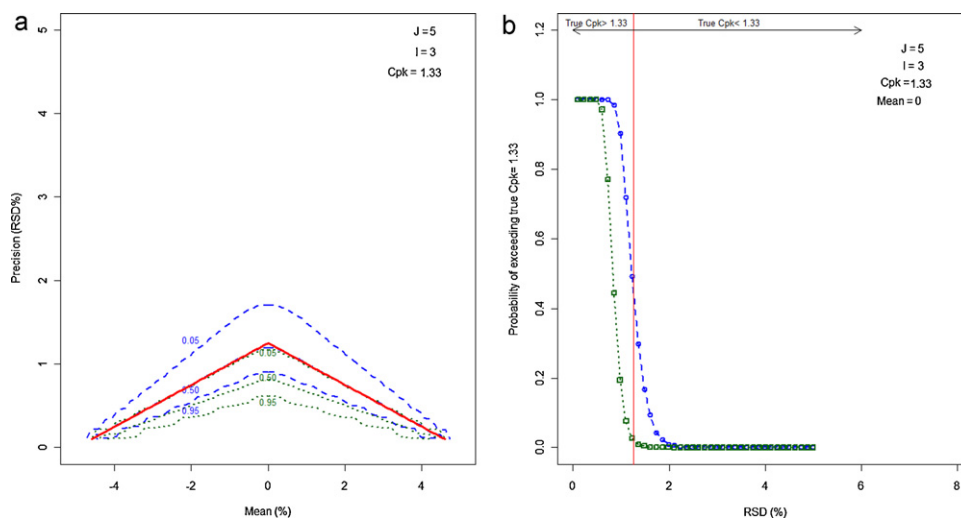
While the  $Cpk\text{-tol}$  index controls well the customer risk, the producer risk (i.e. the risk to conclude the method is not capable while it is truly capable) is relatively high. However this risk can be reduced by increasing the sample size of the method validation. This is shown by comparing Figs. 1, 3 and 5 obtained with 3 runs and 3 repetitions per run to Figs. 2, 4 and 6 obtained with 5 runs and 3 repetitions per run for true  $Cpk$  values of 1, 1.33 and 2, respectively. Finally, these simulations highlighted first the fact that using  $Cpk$  index (Eq. (2)) to decide about the validity of analytical methods is highly controversial especially when using a method validation design of 3 runs and 3 replicates. Second, they showed that using  $Cpk\text{-tol}$  to make such a decision better controls the cos-

tumer risk, thus controls the risk for patients or public health risk, while the producer risk can be modulated by increasing sample size.

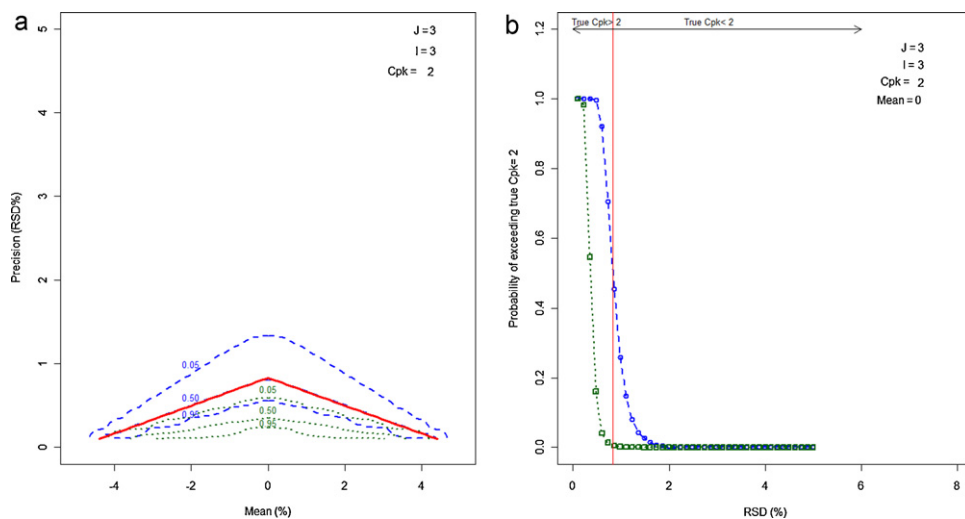
### 3.3. Application to real cases

#### 3.3.1. Case 1: fluconazole determination

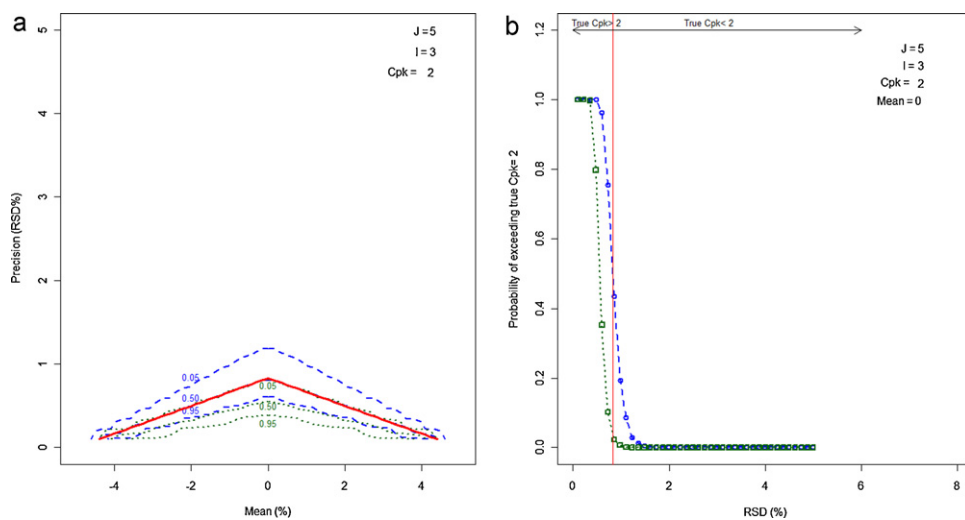
The first case studied concerns the validation of an HPLC-UV method dedicated to monitor the dissolution of fluconazole tablets. For this application, the acceptance limits [LSL; USL] have been set to [85%; 115%] where 100% is the nominal amount of fluconazole of the prepared validation standards. Table 1 summarises the general validation criteria obtained for the validation of this HPLC-UV method. Trueness expressed in terms of relative bias (%) did not exceed the value of 1%, irrespective of the concentration level. The RSD values for intermediate precision presented in Table 1 are relatively low, about maximum 4.5%. In order to demonstrate results linearity, a regression line was fitted on the



**Fig. 4.** Performance of the capability indices  $Cpk$  and  $Cpk\text{-tol}$  for a true  $Cpk$  value of 1.33 and an experimental design of 5 runs and 3 repetitions. (a) Isoprobability contour measuring the probability that  $Cpk$  (dashed curves) and  $Cpk\text{-tol}$  (dotted curves) exceeds the true  $Cpk$  value of 1.33. (b) Operating characteristic curves obtained for the case of a centred analytical method (mean = 0%) depicting the probability that  $Cpk$  (dashed curves with open circles) and  $Cpk\text{-tol}$  (dotted curves with open squares) exceeds the true  $Cpk$  value of 1.33, the vertical line delimits methods with true  $Cpk > 1.33$  and true  $Cpk < 1.33$ .



**Fig. 5.** Performance of the capability indices  $Cpk$  and  $Cpk\text{-tol}$  for a true  $Cpk$  value of 2 and an experimental design of 3 runs and 3 repetitions. (a) Isoprobability contour measuring the probability that  $Cpk$  (dashed curves) and  $Cpk\text{-tol}$  (dotted curves) exceeds the true  $Cpk$  value of 2. (b) Operating characteristic curves obtained for the case of a centred analytical method (mean = 0%) depicting the probability that  $Cpk$  (dashed curves with open circles) and  $Cpk\text{-tol}$  (dotted curves with open squares) exceeds the true  $Cpk$  value of 2, the vertical line delimits methods with true  $Cpk > 2$  and true  $Cpk < 2$ .



**Fig. 6.** Performance of the capability indices  $Cpk$  and  $Cpk\text{-tol}$  for a true  $Cpk$  value of 2 and an experimental design of 5 runs and 3 repetitions. (a) Isoprobability contour measuring the probability that  $Cpk$  (dashed curves) and  $Cpk\text{-tol}$  (dotted curves) exceeds the true  $Cpk$  value of 2. (b) Operating characteristic curves obtained for the case of a centred analytical method (mean = 0%) depicting the probability that  $Cpk$  (dashed curves with open circles) and  $Cpk\text{-tol}$  (dotted curves with open squares) exceeds the true  $Cpk$  value of 2, the vertical line delimits methods with true  $Cpk > 2$  and true  $Cpk < 2$ .

calculated amounts of the validation standards ( $Y$ ) as a function of the introduced amounts ( $X$ ) by applying a linear regression model. The equation obtained is of  $Y = -0.06430 + 1.009 X$  ( $r^2 = 0.9985$ ) for fluconazole.

**Table 1**  
Validation results of the method dedicated to the determination of fluconazole.

Trueness ( $J = 6, I = 4$ ) <sup>a</sup> (mg)		Relative bias (%)
12.67		0.9
25.52		0.2
50.52		1.0
62.30		0.7
Precision ( $J = 6, I = 4$ ) (mg)	Repeatability (RSD%)	Intermediate precision (RSD%)
12.67	0.5	3.7
25.52	0.5	4.4
50.52	0.6	2.1
62.30	0.2	0.6

<sup>a</sup>  $J$  = days (runs) and  $I$  = no. of replicates.

The fluconazole method capability is given on Table 2 for the  $Cpk$  (Eq. (2)) and  $Cpk\text{-tol}$  (Eq. (7)) indices together with their respective estimations of the proportion of non-conforming results in defects per million (dpm) units. As can be seen on this table, the  $Cpk$  values are all above 1.33 except for the second amount level tested (25.52 mg) which is however greater than 1. These values of the  $Cpk$  index show that the method is capable (or almost capable for

**Table 2**  
Capability indices ( $Cpk$  and  $Cpk\text{-tol}$ ) and proportion of non-conforming results ( $\pi_{Cpk}$  and  $\pi_{Cpk\text{-tol}}$ ) obtained for the validation of the HPLC-UV method dedicated to the determination of fluconazole.

Amount level (mg)	$Cpk$	$Cpk\text{-tol}$	$\pi_{Cpk}$ (dpm) <sup>a</sup>	$\pi_{Cpk\text{-tol}}$ (dpm) <sup>a</sup>
12.67	1.38	0.64	36.60	14123.13
25.52	1.20	0.55	321.81	28251.80
50.52	2.43	1.21	0.00	813.11
62.3	9.04	4.83	0.00	0.20

<sup>a</sup> dpm = defects per million results, i.e. proportion of non-conforming results per one million results.

**Table 3**  
Validation results of the method dedicated to the determination of acetazolamide (ACZ).

Trueness ( $J=3, I=3$ ) <sup>a</sup> ( $\mu\text{g mL}^{-1}$ )	Relative bias ( $\mu\text{g mL}^{-1}/\%$ )	
0.05	0.9	
0.10	0.9	
0.15	-0.3	
0.20	-0.8	
0.30	-0.1	
Precision ( $J=3, I=3$ ) <sup>a</sup> ( $\mu\text{g mL}^{-1}$ )	Repeatability (RSD%)	Intermediate precision (RSD%)
0.05	0.7	0.7
0.10	0.8	0.8
0.15	1.0	1.0
0.20	0.7	0.7
0.30	0.9	0.9

<sup>a</sup>  $J$  = days (runs) and  $I$  = no. of replicates.

the second level) by comparison to the common standard value of  $Cpk = 1.33$ . For the two highest amount level,  $Cpk$  values are well above 2 suggesting an excellent capability of the method at this level. The corresponding proportions of non-conforming results range from 322 dpm for the amount level of 25.52 mg to almost 0 dpm for the two highest amount levels of the validation standard levels. When looking at the values of the  $Cpk-tol$  index, it can be seen that the  $Cpk$  index over-estimates the method capability as it was illustrated with the previous simulations. Indeed the  $Cpk-tol$  values range from 0.55 to 4.83 corresponding to proportions of non-conforming results ranging from 28,252 to 0.20 dpm. This indicates that the method capability is far from being excellent using the conventional value of 1.33 as minimum acceptable criterion and suggested by the USP [7]. The method is therefore really highly capable for the highest amount levels tested that are 50.52 and 62.3 mg of fluconazole as there is less than 2700 dpm.

### 3.3.2. Case 2: ACZ determination

This second case studied is about the validation of an HPLC-UV method developed to quantify ACZ in a drug product. The acceptance limits [LSL; USL] have been set to [95%; 105%] where 100% is the nominal concentration of the ACZ validation standard concentration levels. The validation criteria obtained for the validation of HPLC-UV method for the determination of ACZ are given in Table 3. The relative bias (%) of the HPLC-UV method was of maximum 1.0%, whatever the concentration level studied. The RSD values for intermediate precision presented in Table 3 do not either exceed 1.0% illustrating the excellent precision of the HPLC-UV method. For results linearity, a regression line was fitted on the calculated amounts of the validation standards ( $Y$ ) as a function of the introduced amounts ( $X$ ) by applying a linear regression model. The equation obtained is of  $Y = 0.0003 + 0.9973 X$  ( $r^2 = 0.9997$ ) for ACZ.

Finally, the ACZ method capability is given on Table 4 for the  $Cpk$  (Eq. (2)) and  $Cpk-tol$  (Eq. (7)) indices together with their respective estimations of the proportion of non-conforming results. From Table 4, it can be observed that, the  $Cpk$  values are all well above

1.33 and above 2 for the  $0.20 \mu\text{g mL}^{-1}$  concentration level of the validation standards. This illustrates that the method is highly capable and for instance meets the expectation of the USP [7]. The maximum corresponding proportion of non-conforming results is 3.84 dpm for the concentration level  $0.15 \mu\text{g mL}^{-1}$ . According to these numbers, the risk to obtain results outside the acceptance limits of [95%; 105%] of the nominal concentration of the validation standards tested is therefore minimal. However, when using the  $Cpk-tol$  index formula (Eq. (7)), the values of this index are between 1.02 and 1.39. It only exceeds slightly the value of 1.33 for the concentration level of  $0.20 \mu\text{g mL}^{-1}$  of ACZ. From this index, the capability of the HPLC-UV method to quantify ACZ is more moderate. Nonetheless, the proportions of non-conforming results obtained from Eq. (8) are ranging from 206.7 to 1811.6 dpm. The method is therefore still relatively well capable, as the proportions of non-conforming results do not exceed the value of 2700 dpm. Again using Eq. (3) to compute the proportion of non-conforming results or Eq. (2) to compute  $Cpk$  provides a too optimistic view of the method capability.

### 3.3.3. Other applications

In order to stress the importance to use adequate capability indices to determine if analytical methods are fit for their purpose, the capability indices  $Cpk$  and  $Cpk-tol$  were also applied to several other HPLC-UV methods dedicated to the quality control of pharmaceutical formulations previously published. The different HPLC-UV method validations reviewed were dedicated to quantify loratadine [13], paracetamol and codeine [13], methyl parahydroxybenzoate [21], and to quantify valsartan and hydrochlorothiazide [21]. For all analytes except methyl parahydroxybenzoate the acceptance limits were set at [95%; 105%] of the nominal concentrations of the validation standards. For methyl parahydroxybenzoate, a preservative, the acceptance limits were set at [90%; 110%] of the nominal concentrations of the validation standards. The values of  $Cpk$ ,  $Cpk-tol$  and their respective estimations of proportion of non-conforming results are given in Table 5. As can be seen on this table, all the  $Cpk$  values computed using Eq. (2) are all except one above the value of 1.33 and 16 out of 24 are above 2 suggesting that these methods are fully capable. This is corroborated by the corresponding proportion of non-conforming results obtained using Eq. (3) that are all under 1 dpm except for three cases where the maximum value of non-conforming result is about 444 dpm. However Table 5 also stresses that these  $Cpk$  values as well as the proportion of non-conforming results are in fact far too optimistic. Indeed by looking the value of  $Cpk-tol$ , only six  $Cpk-tol$  values are above 1.33 and only half of these last ones are above 2. The maximum predicted proportion of non-conforming results using Eq. (8) is about 45,219 dpm which is about one hundred time worth than the proportion obtained previously with Eq. (3). This tempers the declaration of these methods as highly capable by comparison to the common standards values of capability index of 1.33 or 2. These examples further illustrates that using  $Cpk$  as a measure of the capability of the methods with few data such as the amount of data generated during analytical methods validation or during methods transfer can be risky and may require further investigations.

**Table 4**  
Capability indices ( $Cpk$  and  $Cpk-tol$ ) and proportion of non-conforming results ( $\pi_{Cpk}$  and  $\pi_{Cpk-tol}$ ) obtained for the validation of the HPLC-UV method dedicated to the determination of acetazolamide (ACZ).

Concentration level ( $\mu\text{g mL}^{-1}$ )	$Cpk$	$Cpk-tol$	$\pi_{Cpk}$ (dpm) <sup>a</sup>	$\pi_{Cpk-tol}$ (dpm) <sup>a</sup>
0.05	1.82	1.20	0.05	503.90
0.10	1.67	1.08	0.53	956.24
0.15	1.54	1.02	3.84	1811.59
0.20	2.15	1.39	0.00	206.70
0.30	1.85	1.22	2.96E-02	795.37

<sup>a</sup> dpm = defects per million results, i.e. proportion of non-conforming results per one million results.

**Table 5**  
Capability indices ( $C_{pk}$  and  $C_{pk-tol}$ ) and proportion of non-conforming results ( $\pi_{C_{pk}}$  and  $\pi_{C_{pk-tol}}$ ) obtained from the validation of quantitative HPLC-UV methods previously published [13,21].

Compound	Method	Concentration level (%)	$C_{pk}$	$C_{pk-tol}$	$\pi_{C_{pk}}$ (dpm) <sup>a</sup>	$\pi_{C_{pk-tol}}$ (dpm) <sup>a</sup>	Reference
Loratadine	HPLC-UV	80	6.26	4.19	0.00	0.12	
		100	1.93	1.27	0.01	384.26	
		120	1.67	0.39	0.58	25903.89	
Paracetamol	HPLC-UV	80	1.17	0.42	443.83	32458.99	[13]
		100	1.72	0.73	0.24	8177.79	
		120	5.42	3.57	0.00	0.27	
Codeine	HPLC-UV	80	1.81	0.71	0.06	9089.11	
		100	2.05	1.64	0.00	17.24	
		120	2.33	1.83	0.00	5.41	
Methyl parahydroxybenzoate	HPLC-UV	60	1.42	0.70	19.31	9910.32	
		80	2.54	1.24	0.00	844.09	
		100	3.81	1.89	0.00	128.08	
		120	3.44	0.94	0.00	2742.61	
		140	1.78	0.27	0.09	42137.08	
Valsartan	HPLC-UV	60	3.54	0.48	0.00	12237.81	[21]
		80	4.35	0.54	0.00	8616.89	
		100	3.08	0.39	0.00	18230.32	
		120	3.14	0.58	0.00	9937.07	
		140	6.13	4.02	0.00	0.12	
Hydrochlorothiazide	HPLC-UV	60	2.08	0.26	0.00	40278.59	
		80	2.09	0.26	0.00	29845.29	
		100	5.43	0.79	0.00	4427.34	
		120	4.23	1.19	0.00	1117.12	
		140	1.52	0.21	5.35	45218.58	

<sup>a</sup> dpm = defects per million results, i.e. proportion of non-conforming results per one million results.

Table 6 illustrates this difference of behaviour of  $C_{pk}$  and  $C_{pk-tol}$  applied to results obtained from an HPLC-UV method for the quantification of levonorgestrel in a polymeric matrix [22]. The results given in Table 6 are coming from the analysis of data obtained during two steps of the life cycle of an analytical method. The first step is the method validation which was performed in 4 series of 4 repetitions at three concentration levels 30, 500 and 1000 ng mL<sup>-1</sup> [22] while the second step is routine analysis where data were obtained after 21 runs of analysis that included 6 quality control (QC) samples prepared at two concentration levels: 30 and 500 ng mL<sup>-1</sup> [22]. The method validation involved thus 16 results at each concentration level, while the routine analysis amounted to 126 results at each concentration level of the QC samples. What can be seen from this table is that the over estimation obtained for the capability values by using  $C_{pk}$  (Eq. (2)) and consequently the underestimation of the fraction of non-conforming results obtained using Eq. (3) during the method validation phase is much less important when the sample size increases, such as the data obtained during the routine application of this method. Indeed the  $C_{pk}$  values of the method validation are almost two times the values of the  $C_{pk-tol}$ . While for the routine phase, this difference in capability value has almost disappeared. The difference between the two ways of estimating the proportion of non-conforming results has also greatly diminished going from the validation phase (only 16 analytical results) to the routine analyses (126 analytical results).

**Table 6**  
Capability indices ( $C_{pk}$  and  $C_{pk-tol}$ ) and proportion of non-conforming results ( $\pi_{C_{pk}}$  and  $\pi_{C_{pk-tol}}$ ) obtained from the validation and the routine use of the HPLC-UV method dedicated to quantify levonorgestrel [22].

Compound	Method life cycle phase	Concentration level (ng mL <sup>-1</sup> )	$C_{pk}$	$C_{pk-tol}$	$\pi_{C_{pk}}$ (dpm) <sup>a</sup>	$\pi_{C_{pk-tol}}$ (dpm) <sup>a</sup>	Reference
Levonorgestrel	Validation	30	1.22	0.85	252.34	4720.20	[22]
		500	3.16	1.15	0.00	1578.48	
		1000	4.69	2.11	0.00	70.61	
	Routine analysis	30	1.24	1.16	205.82	286.16	
		500	2.75	2.46	0.00	0.00	

<sup>a</sup> dpm = defects per million results, i.e. proportion of non-conforming results per one million results.

#### 4. Method validation sample size determination

Sample size determination is essential to implement an adequate method validation study. For methods validation the sample size is constituted of two elements, the number of series and the number of repetitions per series. The previous simulations (Fig. 1) have shown that both the customer risk and the producer risk are not easily reduced simultaneously. Therefore in order to define a sample size for method validation, we propose to use a design based on a compromise. To achieve this, two theoretical  $C_{pk}$  values have to be chosen. The first one  $C_{pk}^-$  is the minimum value of capability the method can have to be declared valid and the second one  $C_{pk}^+$  is the highest capability value the method should have to be declared as valid. Second, the probability  $\pi_{C_{pk}^-}$  that the method will be declared as valid for  $C_{pk}^-$  has to be defined. Usually  $\pi_{C_{pk}^-}$  will be set relatively low at, e.g. 5% or 10%. Then the probability  $\pi_{C_{pk}^+}$  that the method will be declared valid for  $C_{pk}^+$  is also specified. For  $\pi_{C_{pk}^+}$ , the probability will be set relatively high at, e.g. 90 or 95%. Then, using the simulation framework described in Section 2.3, the optimal number of series and repetitions to reach these conditions is searched for. The optimal sample size will be here defined as the minimum number of total experiments.

As an example let define  $C_{pk}^- = 0.5$  and  $C_{pk}^+ = 2.5$  with  $\pi_{C_{pk}^-} = 5\%$  and  $\pi_{C_{pk}^+} = 90\%$ . This means that the method can be declared as valid with a true capability value of 0.5 only 5 times out

**Table 7**

Optimal number of experiments to perform during method validation to ensure that  $Cpk^- = 0.5$  and  $Cpk^+ = 2.5$  with  $\pi_{Cpk^-} = 5\%$  and  $\pi_{Cpk^+} = 90\%$  for specification limits of  $\pm 5\%$ .

R	Number of repetitions	Number of series
0.5	5	3
1.0	4	5
2.0	4	7

of a hundred and a method with a true capability value of 2.5 will be declared valid 90 times out of a hundred. Table 7 provides the optimal number of experiments to perform during method validation to ensure these prespecified requirements for 3 different values of  $R = \sigma_B^2/\sigma_W^2$  with  $R \in (0.5, 1, 2)$ . As can be seen on this table, to achieve the desired capability values with the defined probability, the number of series required increases with the value of the ratio  $R$ . Indeed in such situations it is the number of series that is of crucial importance in order to estimate  $\sigma_B^2$  with satisfactory precision. Table 7 also shows that the usual minimum sample size of 3 series of 3 repetitions is generally too small. For situation where the repeatability variance  $\sigma_W^2$  is the greatest source of variability the number of repetitions should be increased to 5. When the ratio  $R$  is close to 1 or greater then it is the number of series that should be increased. The optimal number of experiments to perform in order to validate analytical methods should be estimated on a case-by-case basis based on prior information about the magnitude of the variance components (e.g. obtained through robust optimisation or pre-validation experiments). Nonetheless, by following this methodology we provide a practical framework to achieve this objective using capability metric.

## 5. Conclusion

Analytical methods capability evaluation can be a useful methodology to assess the fitness of purpose of these methods for their future routine application. However, as shown with the simulations as well as with the practical examples provided, care on how to compute the capability indices has to be made. Indeed, the commonly used formulas to compute capability indices such as  $Cpk$ , will highly overestimate the true capability of the methods. Especially during methods validation or transfer, there are only few experiments performed and, using in these situations the commonly applied capability indices to declare a method as valid or as transferable to a receiving laboratory will conduct to inadequate decisions. In this work, an improved capability index, namely  $Cpk-tol$  and the corresponding estimator of proportion of non-conforming results has been proposed. They have been shown to greatly increase the estimation of analytical methods capability in particular in low sample size situations as encountered during methods validation or transfer. Additionally, the usefulness of this capability index has been illustrated through several case studies covering applications commonly encountered in the pharmaceutical industry.

As shown in the examples of Section 3.3, all common validation criteria for quantitative assays are needed and evaluated when using  $Cpk-tol$  as decision index to assess methods validity. These required criteria are trueness, precision (repeatability and intermediate precision), valid range, results linearity, LOQ. They are those required by regulatory requirements such as ICH Q2R1 [9] or ISO 17025 for laboratories accreditations [23]. Nonetheless, methodologies for deciding whether a method is valid or not are lacking in these general guidance documents and it is the responsibility of the laboratory to choose one decision methodology. This is the purpose of this new  $Cpk-tol$  index. This capability index fully matches the validation parameters. The  $Cpk-tol$  includes both trueness and

precision (intermediate precision and hence repeatability). The protocol for method validation for end-users of the method is to realise method validation using «validation standards» at different concentration levels covering the required concentration range for the routine application of the method, each level should be repeated this procedure reproduced in several series. A more precise number of repetitions per series and series could valuably be addressed following the approach to determine sample size proposed in section 4.

In order to both control the customer risk and producer risk, the users should use the  $Cpk-tol$  index as it controls well the customer risk. And based on preliminary information about the method performance (gathered through method development or in method robust optimisation or in pre-validation studies) define the optimum sample size to reduce the producer risk (Section 4).

$Cpk-tol$  is also related to measurement uncertainty [24,25]. It includes the remaining source of systematic influence and the random influences coming from repeatability experiments, together with those included in the between series or run conditions. In addition, it includes the uncertainty of these random influences through the Student- $t$  statistics as well as the uncertainty of the remaining systematic influence as demonstrated by Feinberg et al. [19]. The proportion of non-conforming results can also be seen as the estimation of the confidence probability connected to the coverage factor used when computing expanded uncertainty (usually  $k=2$  for 95% confidence probability).

Concerning concentration levels close to the LOQ for instance in quantitative impurity assays or for trace analysis, the examples used seem to show that capability decreases when concentration levels decrease. This observation however depends on the way the specification limits are set over the concentration range. In the examples they were set in constant relative values over the concentration range. However, in other situations [10,26], they widening when concentration levels decrease, thus method capability would not necessarily seem to worsen for low concentration levels close to LOQ. Would the method precision be excessive, a possible solution apart getting back to method development, would be to consider a result as the average of the analysis of several samples preparations (at least two). Nonetheless, good design of the validation experiments and especially the selection of the sample size required to ensure the method is correctly declared valid should be planned a priori before jumping in the method validation study.

Finally method validation is only a first step in evaluating the ability of a method to provide reliable results. Assessing the fitness of purpose of analytical methods in routine analysis is still essential and mandatory [23,27]. Pre-study validation is in fact a prediction of the future fitness of purpose of the analytical method [22]. The state of validity of analytical method has still to be controlled and verified during the routine use of the method, for instance using control charts obtained from quality control samples.

## Acknowledgements

The authors are very grateful to the anonymous reviewers for providing important comments that led to significant improvements of this article. A research grant from the Belgium National Fund for Scientific Research (FRS-FNRS) to E. Rozet is gratefully acknowledged.

## References

- [1] B. Dejaegher, M. Jimidar, M. De Smet, P. Cockaerts, J. Smeyers-Verbeke, Y. Vander Heyden, J. Pharm. Biomed. Anal. 42 (2006) 155–170.
- [2] M. Kambari, H. Garcia, D.P. Feder, R.J. Rapoza, Eur. J. Pharm. Sci. 42 (2011) 230–237.
- [3] C.S. Raska, T.S. Bennett, S.A. Goodberlet, Anal. Chem. 82 (2010) 5932–5936.
- [4] S. Kotz, N.L. Johnson, J. Qual. Technol. 34 (2002) 1–19.

- [5] S. Kotz, C. Lovelace, *Process Capability Indices in Theory and Practice*, Arnold, London, UK, 1998.
- [6] C. Burgess, Aberrant or atypical results: analytical process capability, in: J. Ermer, J.H. Miller (Eds.), *Method Validation in Pharmaceutical Analysis. A Guide to Best Practice*, Wiley-Verlag, Weinheim, 2005, pp. 362–366.
- [7] USP 33 NF 28 S1, U.S. Pharmacopeia, 2007. USP–NF General Chapter <1080>.
- [8] P. Ferignac, *Revue de statistique appliquée* 10 (1962) 13–32.
- [9] International Conference on Harmonization (ICH) of Technical Requirements for registration of Pharmaceuticals for Human Use, Topic Q2 (R1): Validation of Analytical Procedures: Text and Methodology, Geneva, 2005.
- [10] Guidance for Industry: Bioanalytical Method Validation, US Department of Health and Human Services, US Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Rockville, May 2001.
- [11] Ph. Hubert, J.J. N'guyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewé, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, E. Rozet, J. Pharm. Biomed. Anal. 45 (2007) 82–96, <http://hdl.handle.net/2268/6187>.
- [12] E. Rozet, A. Ceccato, C. Hubert, E. Ziemons, R. Oprean, S. Rudaz, B. Boulanger, Ph. Hubert, J. Chromatogr. A 1158 (2007) 111–125, <http://hdl.handle.net/2268/3646>.
- [13] A. Bouabidi, E. Rozet, M. Fillet, E. Ziemons, E. Chapuzet, B. Mertens, R. Klinkenberg, A. Ceccato, M. Talbi, B. Streel, A. Bouklouze, B. Boulanger, Ph. Hubert, J. Chromatogr. A 1217 (2010) 3180–3192, <http://hdl.handle.net/2268/29471>.
- [14] G.J. Hahn, W.Q. Meeker, *Statistical Intervals: A Guide for Practitioners*, Wiley Edition, 1991, 392 p.
- [15] Ph. Hubert, J.-J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.-A. Compagnon, W. Dewé, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, J. Pharm. Biomed. Anal. 36 (2004) 579–586, <http://hdl.handle.net/2268/6169>.
- [16] Ph. Hubert, J.J. N'guyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewé, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, E. Rozet, J. Pharm. Biomed. Anal. 45 (2007) 70–81, <http://hdl.handle.net/2268/6170>.
- [17] R.W. Mee, *Technometrics* 26 (1984) 251.
- [18] S.R. Searle, G. Casella, C.E. McCulloch, *Variance Components*, Wiley, 1992.
- [19] M. Feinberg, B. Boulanger, W. Dewé, P. Hubert, Anal. Bioanal. Chem. 380 (2004) 502–514, <http://hdl.handle.net/2268/6186>.
- [20] P. Dagnelie, *Statistique théorique et appliquée. 2. Inférence statistique à une et deux dimensions*, De Boeck Université, 2006, p. 737.
- [21] A. Bouabidi, M. Talbi, A. Bouklouze, M. El Karbane, H. Bourichi, M. El Guezzar, E. Ziemons, Ph. Hubert, E. Rozet, J. Pharm. Biomed. Anal. 55 (2011) 583–590, <http://hdl.handle.net/2268/85024>.
- [22] E. Rozet, C. Hubert, A. Ceccato, W. Dewé, E. Ziemons, F. Moonen, K. Michail, R. Wintersteiger, B. Streel, B. Boulanger, Ph. Hubert, J. Chromatogr. A 1158 (2007) 126–137, <http://hdl.handle.net/2268/1810>.
- [23] International Organization of Standardization (ISO), ISO/CEI 17025: General Requirements for the Competence of Testing and Calibration Laboratories, ISO, Geneva, Switzerland, 2005.
- [24] Joint Committee for Guides in Metrology, JCGM 100:2008, Evaluation of measurement data – Guide to the expression of uncertainty in measurement (GUM), 2008 (available from the website: [www.bipm.org](http://www.bipm.org)).
- [25] Eurachem/CITAC Guide, Quantifying the Uncertainty in Analytical Measurement, second ed., 2000.
- [26] 2002/657/EC, Commission Decision of 12 August 2002 implementing Council Directive 96/23/EC Concerning the Performance of Analytical Methods and the Interpretation of Results, Off. J. Eur. Commun. L221 (2002) 8.
- [27] International Organization of Standardization (ISO), ISO: 15189 Medical Laboratories – Particular Requirements for Quality and Competence, ISO, Geneva, Switzerland, 2007.



## **CHAPITRE VII : Est-ce que les échantillons placebo imitent bien le comportement réel des lots de fabrication ?**

*Ce chapitre correspond à l'article "DO PLACEBO BASED VALIDATION STANDARDS MIMIC REAL BATCH PRODUCTS BEHAVIOUR? CASE STUDIES" publié dans le journal of Pharmaceutical and Biomedical Analysis 55 (2011) 583–590*

### **1. Synthèse et conclusion**

### **2. Article**

## 1. Synthèse et conclusion

La validation des méthodes analytiques est une étape obligatoire pour évaluer la capacité des méthodes développées à fournir des résultats exacts pour leur application de routine. La validation implique généralement des standards de validation ou des échantillons CQ (de contrôle de qualité) qui sont préparés dans un placebo ou dans une matrice reconstituée composée d'un mélange de tous les ingrédients du médicament, à l'exception de la substance active ou de la substance à analyser. Toutefois, l'un des principaux soucis de cette approche est qu'elle peut ne pas prendre en considération une source importante de variabilité qui provient du processus de fabrication du produit fini comme le mélange, la compression, le chauffage, l'intensité de la cohésion des particules, ...etc, de tous les ingrédients, y compris la substance active.

En revanche, le document ICH recommande que «la fidélité devrait être étudiée en utilisant des échantillons homogènes authentiques», et ajoute que l'utilisation d'échantillons artificiellement préparée ne doit être utilisée que "s'il n'est pas possible d'obtenir un échantillon homogène".

La question qui demeure à la fin de l'étape de validation concerne la transférabilité des performances quantitatives des standards de validation à des échantillons authentiques réels du médicament. Puisque la plupart du temps, seuls les échantillons préparés artificiellement sont utilisés comme des standards de validation dopés.

Le but de ce travail est de répondre partiellement à cette question. Pour y parvenir, trois études de cas ont été effectuées et trois méthodes d'analyse ont été validées en utilisant des placebos dopés des standards de validation à plusieurs niveaux de concentration ainsi que sur des échantillons provenant des échantillons de lots authentiques (comprimés et sirops).

D'abord une méthode spectrophotométrique UV dédiée à la quantification de la metformine dans des comprimés a été validée en utilisant des standards de validation artificiellement préparés ainsi que des comprimés réels provenant d'un seul lot industriel.

Deuxièmement, une méthode HPLC isocratique dédiée au dosage du parahydroxybenzoate de méthyle dans un sirop à base de carbocéstiène a aussi été validée en utilisant à la fois des standards de validation dopés et le sirop obtenu à partir du processus de fabrication de trois lots différents.

Enfin, une méthode gradient HPLC-UV visant à quantifier simultanément le valsartan et l'hydrochlorothiazide en comprimés a également été en même temps étudiée.

Les résultats de ce travail montrent que, selon le modèle de fonction de réponse utilisé comme courbe d'étalonnage, il y a plusieurs degrés de différences dans les résultats d'exactitude obtenus avec les deux types d'échantillons. Néanmoins, l'utilisation des placebo dopés par des standards de validation simule relativement bien le comportement quantitatif des méthodes d'analyse avec des échantillons de lots réels.

L'ajout de ces échantillons de lots réels au plan de validation peut aider l'analyste à choisir et à confirmer la courbe d'étalonnage la plus adéquate et donc à augmenter l'exactitude et la fiabilité des résultats générés par la méthode lors de son application de routine.

## **2. Article**

### **DO PLACEBO BASED VALIDATION STANDARDS MIMIC REAL BATCH PRODUCTS BEHAVIOUR? CASE STUDIES**

Journal of Pharmaceutical and Biomedical Analysis 55 (2011) 583–590

A. Bouabidi, M. Talbi, A. Bouklouze, M. El Karbane, H. Bourichi, M. El Guezzar, E. Ziemons, Ph. Hubert, E. Rozet



## Do placebo based validation standards mimic real batch products behaviour? Case studies<sup>☆</sup>

A. Bouabidi<sup>a,b</sup>, M. Talbi<sup>a</sup>, A. Bouklouze<sup>c</sup>, M. El Karbane<sup>c</sup>, H. Bourichi<sup>c</sup>, M. El Guezzar<sup>d</sup>, E. Ziemons<sup>b</sup>, Ph. Hubert<sup>b,1</sup>, E. Rozet<sup>b,\*,1,2</sup>

<sup>a</sup> Analytical Chemistry Laboratory, University Hassan II - Mohammedia - Faculty of Sciences Ben M'Sik, Casablanca, Morocco

<sup>b</sup> Analytical Chemistry Laboratory, Department of Pharmacy, CIRP, University of Liège, Liège, Belgium

<sup>c</sup> Pharmacology-Toxicology Laboratory, Faculty of Medicine and Pharmacy, University Med V Soussi, Rabat, Morocco

<sup>d</sup> Cooper Pharma, Casablanca, Morocco

### ARTICLE INFO

#### Article history:

Received 20 October 2010

Received in revised form 14 January 2011

Accepted 10 February 2011

Available online 16 February 2011

#### Keywords:

Authentic batch sample

Placebo validation standards

Validation methodology

Accuracy profile

Reconstituted validation standard

### ABSTRACT

Analytical methods validation is a mandatory step to evaluate the ability of developed methods to provide accurate results for their routine application. Validation usually involves validation standards or quality control samples that are prepared in placebo or reconstituted matrix made of a mixture of all the ingredients composing the drug product except the active substance or the analyte under investigation. However, one of the main concerns that can be made with this approach is that it may lack an important source of variability that come from the manufacturing process. The question that remains at the end of the validation step is about the transferability of the quantitative performance from validation standards to real authentic drug product samples.

In this work, this topic is investigated through three case studies. Three analytical methods were validated using the commonly spiked placebo validation standards at several concentration levels as well as using samples coming from authentic batch samples (tablets and syrups). The results showed that, depending on the type of response function used as calibration curve, there were various degrees of differences in the results accuracy obtained with the two types of samples. Nonetheless the use of spiked placebo validation standards was showed to mimic relatively well the quantitative behaviour of the analytical methods with authentic batch samples. Adding these authentic batch samples into the validation design may help the analyst to select and confirm the most fit for purpose calibration curve and thus increase the accuracy and reliability of the results generated by the method in routine application.

© 2011 Elsevier B.V. All rights reserved.

### 1. Introduction

Prior to the daily application of analytical methods in pharmaceutical laboratories for drug products release it is mandatory to validate them following for instance the ICH recommendations [1]. Analytical method validation usually involves validation standards or quality control samples that are prepared in placebo or reconstituted matrix made of a mixture of all the ingredients composing the drug product except the active substance [2–9]. However, one of the main concerns that can be made with this approach is that it may

lack an important source of variability that could come from the manufacturing process when the final pharmaceutical formulation of the product is produced. This supplemental source of variability can result from the mixing, compressing, blending, heating, intensity of cohesion of particles, and so on, of all the ingredients including the active substance performed during production.

In the other hand, the ICH guideline (Q2R1) [1] recommends that “precision should be investigated using homogeneous, authentic samples”, and adds that the use of artificially prepared samples must be used only “if it is not possible to obtain a homogeneous sample”. Therefore, the question that remains at the end of the validation step is about the transferability of the quantitative performance from validation standards to real drug product samples as most of the times only artificially prepared samples are used such as the spiked validation standards.

The aim of this paper is to answer partially this question as it has never been done to our best knowledge. To achieve this, three case studies have been performed. First a UV spectrophotometric method dedicated to the quantification of metformin in tablets

<sup>☆</sup> Work presented at the 2010 International Symposium on Drug Analysis (Drug Analysis 2010) held in Antwerp, Belgium – September 21–24, 2010.

\* Corresponding author at: Laboratory of Analytical Chemistry, Institute of Pharmacy, Université de Liège, CHU, B 36, Av. de l'Hôpital, B-4000 Liège, Belgium. Tel.: +32 4 3664320; fax: +32 4 3664317.

E-mail address: [eric.rozet@ulg.ac.be](mailto:eric.rozet@ulg.ac.be) (E. Rozet).

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> F.R.S.-FNRS Postdoctoral Researcher (Belgium).

was validated using artificially prepared validation standards as well as real tablets coming from a single produced industrial batch. Second, an isocratic HPLC method dedicated to the quantification of methyl parahydroxybenzoate content in carbocestiene syrup pharmaceutical formulation was also validated using both spiked validation standards and authentic syrup obtained from the manufacturing process of three different batches. Finally, a gradient HPLC-UV method aiming at quantifying simultaneously valsartan and hydrochlorothiazide in tablets was also similarly studied.

## 2. Material and method

### 2.1. Chemicals and solvents

Metformin hydrochloride standard reference, methyl parahydroxybenzoate standard reference, valsartan standard reference and hydrochlorothiazide standard reference were supplied by the European Pharmacopoeia (Strasbourg, France). Acetonitrile of HPLC grade was purchased from Merck (Darmstadt, Germany). Potassium dihydrogen phosphate was purchased from Acros Organics (Geel, Belgium). Sodium acetate of analytical grade was supplied from Sigma-Aldrich (Steinheim, Germany). Deionized water was generated from a Milli-Q water purifying system (Millipore, Watford, UK).

### 2.2. Apparatus

#### 2.2.1. UV determination of metformin

Determination of metformin was performed with an Agilent UV-VIS 8453E double-beam spectrophotometer (Agilent, Palo-Alto, CA, USA) at a wavelength of 232 nm using purified water as blank.

#### 2.2.2. HPLC determination of methyl parahydroxybenzoate

The HPLC system consisted in a LaChrom (Merck-Hitachi, Darmstadt, Germany) composed of a quaternary pump L-7100, an autosampler L-7200, an oven L-7360 and a DAD detector L-7455. The mobile phase consisted in a 33/67 (v/v) mixture of acetonitrile and an aqueous buffer at pH 5.0 of sodium acetate and water. The chromatographic isocratic separation was made with a Nucleosil C18 column (150 mm × 4.6 mm ID.; particle size: 5 µm; Macherey-Nagel, Hoerdt, France) and the UV detection was performed at 245 nm. The flow rate was set at 1.5 ml/min and 20 µl of each sample was injected onto the column. The analysis run time was of 15 min.

#### 2.2.3. HPLC determination of valsartan and hydrochlorothiazide

The same HPLC system as employed for the determination of methyl parahydroxybenzoate was used. The mobile phase consisted in a mixture of acetonitrile and a phosphate buffer at pH 3.0 prepared with a 1 mM potassium dihydrogen phosphate solution in deionized water. The pH was adjusted to 3.0 with phosphoric acid. The chromatographic gradient separation was made with a Zorbax-SB C18 column (50 mm × 4.6 mm ID.; particle size: 1.8 µm; Agilent) and the UV detection was performed at 273 nm. The gradient program is given in Table 1. The flow rate was set at 0.8 ml/min and 5 µl of each sample was injected onto the column.

### 2.3. Solutions

#### 2.3.1. UV determination of metformin

##### a. Sample solution

About 1/5th of the mean weight of 20 tablets is accurately weighed and transferred into a 100 ml volumetric flask. It is dissolved in purified water and diluted up to the mark. Then this

**Table 1**

Gradient program used for the HPLC-UV method dedicated to the simultaneous determination of valsartan and hydrochlorothiazide.

Time (min)	pH 3.0 phosphate buffer	Acetonitrile
0	80	20
8	80	20
10	55	45
30	55	45
31	80	20
40	80	20

solution is further filtered and about 1/100 of this solution is diluted with purified water.

##### b. Calibration and validation standards

Calibration standards and validation standards were prepared at three concentration levels: 80, 100 and 120% of the nominal target concentration of active substance in the drug product. The calibration standards prepared in purified water were analysed in duplicates during three days. Each validation standard prepared in a reconstituted matrix of the formulation (or placebo) was analysed in triplicates during the same three days, except the 100% level that was replicated 6 times each day.

#### 2.3.2. HPLC determination of methyl parahydroxybenzoate

##### a. Sample solution

An accurately weighed sample of about 1.0 g of syrup is transferred to a 100 ml volumetric flask and diluted up to the mark with methanol. The same single initial source of syrup was used for each replicate and in each series to avoid introducing additional variability potentially coming from the batch heterogeneity.

##### b. Calibration and validation standards

Calibration standards and validation standards were prepared at five concentration levels: 60, 80, 100, 120 and 140% of the target nominal concentration of methyl parahydroxybenzoate in the syrup formulation. The calibration standards prepared in methanol were analysed once during 3 days. Each validation standard prepared in a reconstituted matrix of the formulation (or placebo) was analysed in triplicates during the same three days, except the 100% level that was replicated 6 times each day.

#### 2.3.3. HPLC determination of valsartan and hydrochlorothiazide

##### a. Sample solution

Twenty tablets are crushed and a mass equivalent to the mean mass of a tablet is accurately weighed, dissolved and diluted to 100.0 ml with a 65/35 (v/v) mixture of pH 3.0 phosphate buffer and acetonitrile. For each replicate and for each series of analysis the same single initial sample of twenty tablets was used for the method validation using authentic batch samples in order to avoid introducing additional variability that may arise from the batch heterogeneity.

##### b. Calibration and validation standards

Calibration standards and validation standards were prepared at five concentration levels: 80, 90, 100, 110 and 120% of the target nominal concentration of valsartan and hydrochlorothiazide in the pharmaceutical formulation. The calibration standards prepared in a 65/35 (v/v) mixture of pH 3.0 phosphate buffer and acetonitrile were analysed in triplicates during 3 days. Each validation standard prepared in a reconstituted matrix of the formulation (or placebo) was analysed in triplicates during the same three days.

### 2.4. Validation approach

The validation approach used is the accuracy profile methodology [7–10]. This methodology is based on the concept of total

measurement error or results accuracy, i.e. the simultaneous combination of the systematic error (measured by biases, i.e. method trueness) and of the random error (measured by RSDs, i.e. method precision) [1,7–12].

What is required at the end of the validation step is to give guarantees that each of the future results that the laboratory will obtain during routine analyses will be accurate enough. Therefore, to reach this objective, instead of a whole set of statistical tests, the accuracy profile methodology uses only one statistical decision methodology, namely a  $\beta$ -expectation tolerance interval for one way ANalysis Of Variance (ANOVA) model [7–10]. This interval represents a region where it is expected to find each future result with a defined probability  $\beta$  specified by the user [13]. The following  $\beta$ -expectation tolerance interval formula is used [14]:

$$[L, U] = [\hat{\mu} - k\hat{\sigma}_{IP}; \hat{\mu} + k\hat{\sigma}_{IP}] \quad (1)$$

where  $\hat{\mu}$  is the mean of the results and  $k$  is calculated in order to have an expected proportion  $\beta$  of the population within this interval. The formula of this tolerance interval is given by:

$$k = Q_{\tau} \left( df, \frac{(1 + \beta)}{2} \right) \sqrt{1 + \frac{J\hat{R} + 1}{JL(\hat{R} + 1)}} \quad (2)$$

$$\text{with } df = \frac{(\hat{R} + 1)^2}{(\hat{R} + (1/L))^2 / (J - 1) + (1 - (1/L)) / JL} \quad \text{and} \quad \hat{R} = \frac{\hat{\sigma}_B^2}{\hat{\sigma}_W^2}$$

where  $Q_{\tau}(df, (1 + \beta)/2)$  is the  $(1 + \beta)/2$ th percentile of a Student distribution with  $df$  degrees of freedom. Then, the intermediate precision variance can be estimated using:  $\hat{\sigma}_{IP}^2 = \hat{\sigma}_B^2 + \hat{\sigma}_W^2$ .  $\hat{\sigma}_B^2$  is the run-to-run variance and  $\hat{\sigma}_W^2$  is the within-run or repeatability variance obtained with ANOVA methodology [15].  $J$  is the number of run performed and  $L$  the number of replicates per run.

#### 2.4.1. Using spiked placebo validation standards

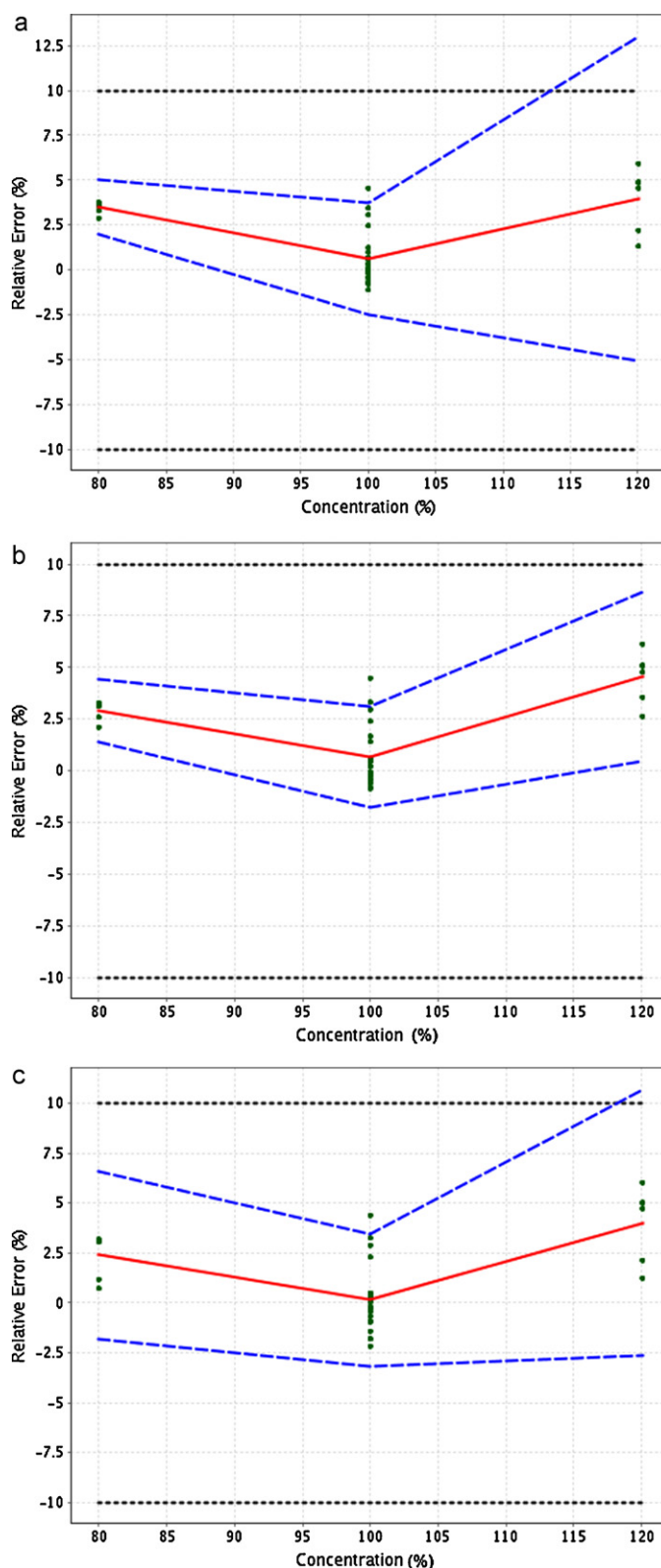
When the previously described validation approach is applied at the different concentration levels of the spiked placebo validation standards, an accuracy profile can be drawn [7–10]. Indeed the  $\beta$ -expectation tolerance interval is computed at each of these concentration levels and the method bias is estimated by deducting to each mean value the concentration of the validation standard. This profile is then compared to “a priori” defined acceptance limits. In the remaining of this work, the acceptance limits were set at  $\pm 10\%$  while the probability  $\beta$  is fixed at 95%.

#### 2.4.2. Using authentic batch samples

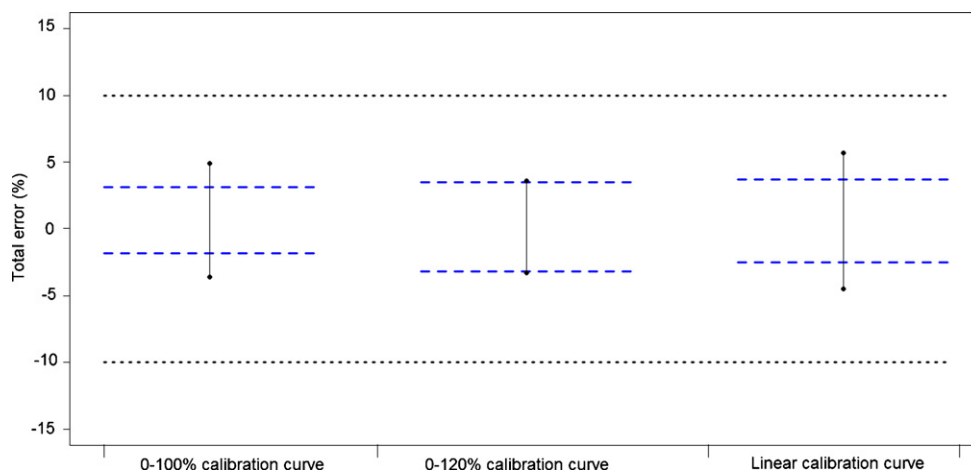
The following authentic batch samples were analysed:

- For the UV determination of metformin: tablets coming from a batch were analysed 6 times during each of the same days of the validation standards.
- For the HPLC-UV determination of methyl parahydroxybenzoate: syrups coming from three different batches were all analysed 6 times during the same days of the validation standards.
- For the simultaneous determination of valsartan and hydrochlorothiazide by gradient HPLC-UV, tablets coming from the same batch were analysed in triplicates during the same days of the validation standards.

The results obtained from the authentic batch samples are also analysed using the same  $\beta$ -expectation tolerance interval for one way ANOVA model. Indeed, the authentic batch samples are analysed in several runs and in several replicates within runs. Then, to compute total error of the authentic batch samples, the method bias ( $\delta$ ) obtained from the spiked validation standards matching



**Fig. 1.** Accuracy profiles of the UV method for the determination of metformin obtained for the spiked validation standards with the simple linear regression model (a) and with the linear model forced through the origin and fitted only with the 100.0% calibration level (b) or fitted only with the 120.0% calibration level (c). The plain line is the relative bias, the dashed lines are the 95%  $\beta$ -expectation tolerance limits and the dotted curves represent the acceptance limits ( $\pm 10\%$ ). The dots represent the relative back-calculated concentrations of the validation standards.



**Fig. 2.** Accuracy of the results obtained for the UV method for the determination of metformin for the authentic batch samples (tablets) with the simple linear regression model, with the linear model forced through the origin and fitted only with the 100.0% calibration level or fitted only with the 120.0% calibration level. The vertical continuous segments are the 95%  $\beta$ -expectation tolerance intervals obtained with the authentic batch samples, the dashed lines are the 95%  $\beta$ -expectation tolerance limits obtained with the 100% concentration level of the spiked validation standards and the dotted curves represent the acceptance limits ( $\pm 10\%$ ).

the target nominal concentration of analyte under study in the drug product is added to this distribution and graphed.

### 2.5. Computations

All data treatments were realised using e-noval software V3.0 (Arlenda, Liege, Belgium) for the accuracy profile approach and the  $\beta$ -expectation tolerance intervals computations.

## 3. Results

### 3.1. UV determination of metformin

#### 3.1.1. Spiked placebo validation standards

By using the approach proposed by Hubert et al. [8–10], the method is considered as valid within the range for which the accuracy profile is included inside the accuracy acceptance limits set at  $\pm 10\%$ . This approach gives the guarantee that each future result generated by the method will be included within  $\beta$ -expectation tolerance limits with a user defined guarantee. Here this guarantee is set at 95.0%.

Three calibration models were investigated, namely the simple linear regression, the linear regression forced through the origin (0) and fitted only with the 100% calibration level or fitted only with the 120% calibration level. Accuracy profiles obtained with these calibration curves for the validation standards are shown in Fig. 1.

As illustrated in this figure, the method is thus considered valid over the whole concentration range investigated using a calibration curve forced through the origin and fitted only with the 100% level of the calibration standards as the 95% tolerance intervals are fully included within the  $\pm 10\%$  acceptance limits (Fig. 1b). All the validation results obtained with this calibration curve are given in Table 2.

#### 3.1.2. Authentic batch samples effect

In order to assess the impact of the process on the results accuracy, 95%  $\beta$ -expectation tolerance intervals were obtained from results of the real samples coming from a manufactured batch of tablets. The same previous calibration models were tested. Fig. 2 shows added to the  $\beta$ -expectation tolerance intervals computed for the real batch samples those obtained from the 100% concentration level of the spiked placebo validation standards. The  $\beta$ -expectation tolerance intervals were computed with the same guarantee:  $\beta = 95\%$ .

As shown in Fig. 2, there is almost no difference between the tolerance intervals computed for the real samples (tablets) with the simple linear calibration model and the one forced through the origin and only the 100% calibration level. However, for these calibration models, the method seems more variable when analysing real tablets than when analysing spiked validation standards (Fig. 2). Nonetheless the quality of the results generated remains acceptable. Indeed, the tolerance intervals are still within the  $\pm 10\%$  acceptance limits, thus corroborating the conclusion made with the spiked placebo validation standards. The results of precision and accuracy obtained with the calibration curve fitted only with the 100% calibration level for the authentic batch samples are given in Table 3.

When using the linear model forced through the origin and fitted only with the 120% calibration level it can be seen in Fig. 2 that the method precision for authentic batch samples is increased. Furthermore for this calibration curve the quality of the results generated when analysing real tablets and spiked placebos are equivalent. However it has to be reminded that when using the spiked placebo validation standards, this calibration curve did not allow to obtain accurate results at the highest concentration level, thus impairing the validity of this method when referring to the ICH Q2 requirements about the range criterion [1].

### 3.2. HPLC determination of methyl parahydroxybenzoate

#### 3.2.1. Spiked placebo validation standards

The accuracy profile approach [8–10] has also been used to evaluate the validity of the HPLC-UV method for the determination of methyl parahydroxybenzoate in syrup. The accuracy acceptance limits were also set at  $\pm 10\%$  and the minimum probability to obtain each future result generated by this method within the  $\pm 10\%$  acceptance limits is set at 95.0%. Three calibration models were investigated. They are the simple linear regression, the linear regression forced through the origin (0) and fitted only with the 100% calibration level or fitted only with the 140% calibration level. Accuracy profiles obtained with these calibration curves for the validation standards are shown in Fig. 3.

All three calibration curves allow the method to be considered as valid over the whole concentration range investigated. Indeed, the 95% tolerance intervals are fully included within the  $\pm 10\%$  acceptance limits in all cases. The final calibration curve selected was thus the calibration curve forced through the origin and fitted only with the 140% level of the calibration standards as it is one of the



**Table 2**

Validation results obtained with the spiked placebo validation standards for the UV method for the determination of metformin, for the HPLC-UV method for the determination of methyl parahydroxybenzoate as well as for the HPLC-UV method for the simultaneous determination of valsartan and hydrochlorothiazide.

	Metformin	Methyl parahydroxybenzoate	Valsartan	Hydrochlorothiazide
Calibration curve	0–100% linear model	0–140% linear model	0–120% linear model	0–120% linear model
<i>Trueness</i>				
Concentration level (%)	Relative bias (%)	Relative bias (%)	Relative bias (%)	Relative bias (%)
60	–	0.5	–	–
80	2.9	0.5	0.1	0.0
90	–	–	0.2	0.9
100	0.7	–0.4	0.0	0.0
110	–	–	0.0	–0.7
120	4.6	–0.7	–0.4	–1.3
140	–	0.6	–	–
	Metformin	Methyl parahydroxybenzoate	Valsartan	Hydrochlorothiazide
Calibration curve	0–100% linear model	0–140% linear model	0–120% linear model	0–120% linear model
<i>Precision</i>				
Concentration level (%)	Repeatability/intermediate precision (RSD, %)	Repeatability/intermediate precision (RSD, %)	Repeatability/intermediate precision (RSD, %)	Repeatability/intermediate precision (RSD, %)
60	–	2.2/2.2	–	–
80	0.2/0.5	1.2/1.2	0.1/0.5	0.1/0.9
90	–	–	0.1/0.4	0.1/0.7
100	1.4/1.4	0.8/0.8	0.1/0.6	0.1/0.3
110	–	–	0.2/0.6	0.2/0.4
120	0.6/1.4	0.6/0.9	0.3/0.3	0.2/0.9
140	–	0.7/1.9	–	–
	Metformin	Methyl parahydroxybenzoate	Valsartan	Hydrochlorothiazide
Calibration curve	0–100% linear model	0–140% linear model	0–120% linear model	0–120% linear model
<i>Accuracy</i>				
Concentration level (%)	Relative 95% $\beta$ -expectation tolerance interval (%)	Relative 95% $\beta$ -expectation tolerance interval (%)	Relative 95% $\beta$ -expectation tolerance interval (%)	Relative 95% $\beta$ -expectation tolerance interval (%)
60	–	[–5.8; 6.8]	–	–
80	[1.4; 4.5]	[–3.0; 4.0]	[–2.4; 2.6]	[–4.5; 4.4]
90	–	–	[–1.8; 2.3]	[–2.7; 4.6]
100	[–1.8; 3.1]	[–2.8; 1.9]	[–3.0; 3.0]	[–1.6; 1.5]
110	–	–	[–2.5; 2.4]	[–2.0; 0.6]
120	[0.5; 8.6]	[–4.1; 2.7]	[–1.0; 0.2]	[–5.6; 3.0]
140	–	[–8.1; 9.3]	–	–

**Table 3**

Results obtained for the analysis of the authentic batch samples for the UV method for the determination of metformin, for the HPLC-UV method for the determination of methyl parahydroxybenzoate as well as for the HPLC-UV method for the simultaneous determination of valsartan and hydrochlorothiazide.

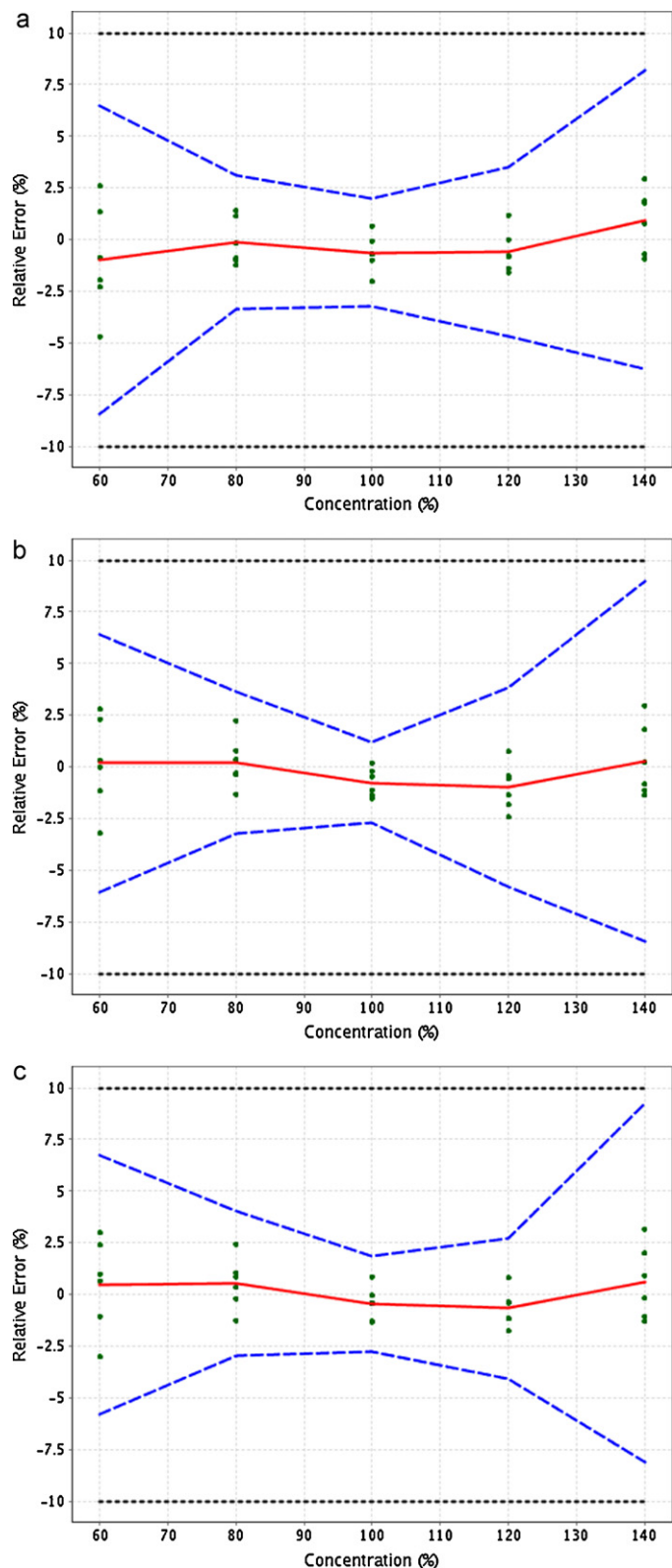
	Metformin	Methyl parahydroxybenzoate			Valsartan	Hydrochlorothiazide
Calibration curve	0–100% linear model	0–140% linear model			0–120% linear model	0–120% linear model
<i>Precision</i>		Batch 1	Batch 2	Batch 3	Batch 1	Batch 1
Repeatability (RSD, %)	0.7	0.4	0.7	0.9	0.6	0.4
Intermediate precision (RSD, %)	1.6	1.1	1.1	1.6	0.8	0.8
<i>Accuracy</i>						
Relative 95% $\beta$ -expectation tolerance interval (%)	[–3.6; 4.9]	[–5.2; 4.3]	[–4.2; 3.3]	[–6.4; 5.5]	[–2.3; 2.2]	[–3.2; 3.1]

simplest calibration model, it provides the least bias and leads to the least extrapolation of results. All the validation results obtained with this calibration curve are given in Table 2.

### 3.2.2. Authentic batch samples effect

To evaluate if the spiked validation standard mimic and predict well the quality of authentic batch sample results, samples coming from three different batches were analysed. Samples of each batch were analysed during three different days and in six replicates. The same calibration models as those evaluated with the spiked placebo standards were tested. Fig. 4 shows the 95%  $\beta$ -expectation tolerance intervals obtained with the authentic batch samples together with those obtained from the results of the analysis of the 100% concentration level spiked placebo validation standards as well as the  $\pm 10\%$  acceptance limits.

For each calibration model tested, a batch-to-batch variability can be observed, as illustrated by the various lengths of the 95%  $\beta$ -expectation tolerance intervals shown in Fig. 4. What is also visible in Fig. 4 is that the method is providing less accurate results when analysing authentic batch samples than when analysing spiked placebo validation standards. All the calibration models tested are nonetheless providing results of adequate accuracy and reliability as all 95%  $\beta$ -expectation tolerance intervals are included within the  $\pm 10\%$  acceptance limits. In this example, the linear model forced through the origin and fitted only with the 140.0% calibration level is the model providing results from authentic batch samples analysis of the best accuracy. This fully confirms the selection of this model as calibration curve with the spiked placebo validation standards. The results of precision and accuracy obtained with this calibration curve for the



**Fig. 3.** Accuracy profiles of the HPLC-UV method for the determination of methyl parahydroxybenzoate obtained for the spiked validation standards with the simple linear regression model (a) and with the linear model forced through the origin and fitted only with the 100.0% calibration level (b) or fitted only with the 140.0% calibration level (c). The plain line is the relative bias, the dashed lines are the 95%  $\beta$ -expectation tolerance limits and the dotted curves represent the acceptance limits ( $\pm 10\%$ ). The dots represent the relative back-calculated concentrations of the validation standards.

authentic batch samples are given in Table 3 for the three batches tested.

### 3.3. HPLC determination of valsartan and hydrochlorothiazide

#### 3.3.1. Spiked placebo validation standards

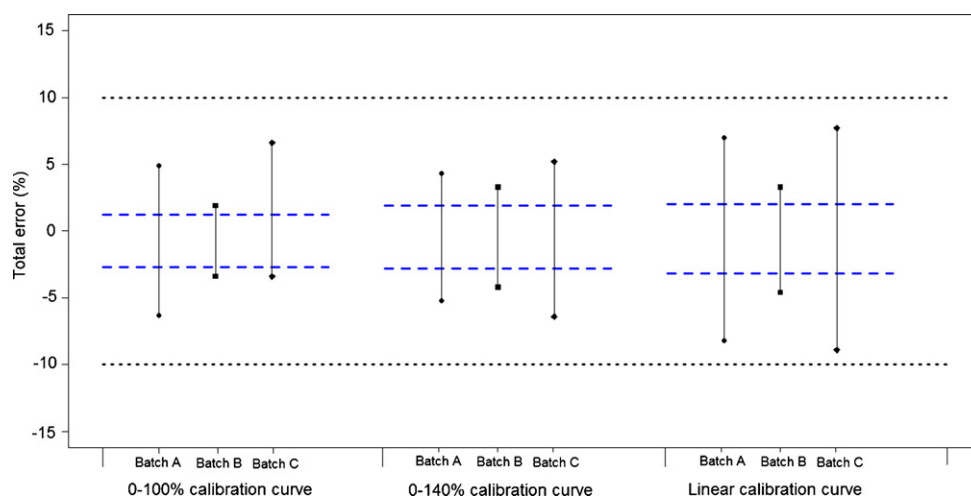
For the validation of the HPLC-UV method for the determination of valsartan and hydrochlorothiazide in tablets, the accuracy profile approach was applied with acceptance limits also set at  $\pm 10\%$  and the probability  $\beta$  is set at 95% for both analytes. The simple linear regression, the linear regression forced through the origin (0) and fitted only with the 100% calibration level or fitted only with the 120% calibration level were tested as standard curve. All these calibration curves allowed guarantee to obtain accurate results for both analytes over the whole concentration range studied. The final calibration curve selected was thus the calibration curve forced through the origin and fitted only with the 120% level of the calibration standards as it is one of the simplest calibration model, it provides the least bias and leads to the least extrapolation of results. All the validation results obtained with this calibration curve are given in Table 2 and the corresponding accuracy profiles for valsartan and hydrochlorothiazide are shown in Fig. 5a and b, respectively.

#### 3.3.2. Authentic batch samples effect

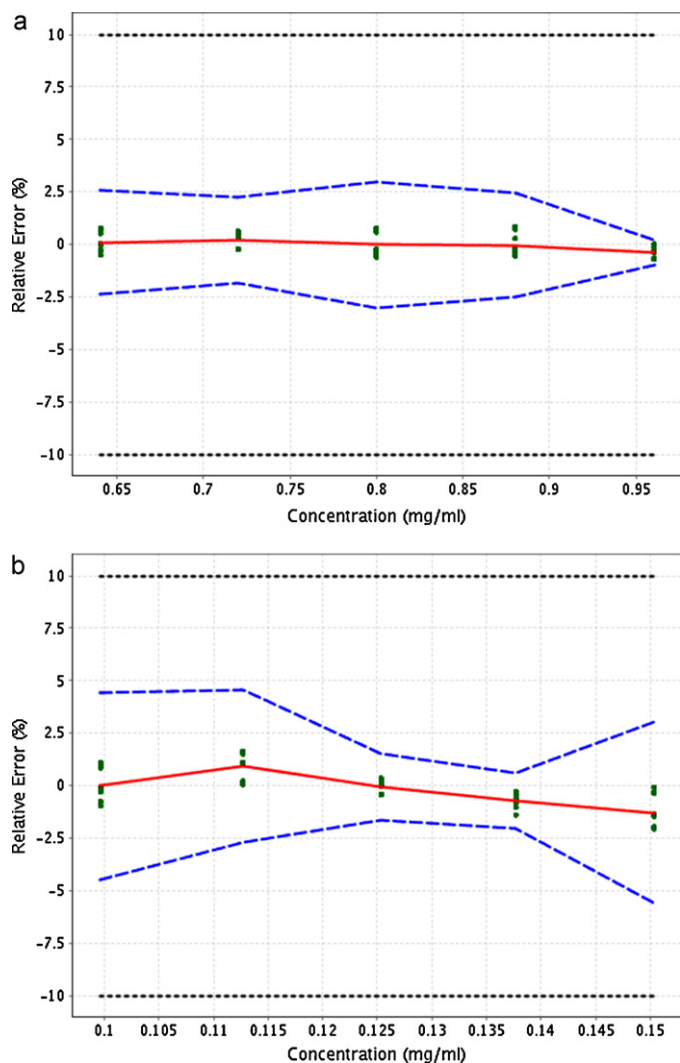
To evaluate if placebo based validation standards mimic well the behaviour of authentic batch samples, tablets from a batch were analysed during three days and in triplicate. Fig. 6 shows the 95%  $\beta$ -expectation tolerance intervals obtained with the authentic batch samples together with those obtained from the results of the analysis of the 100% concentration level spiked placebo validation standards as well as the  $\pm 10\%$  acceptance limits for both analytes and obtained with the previously selected calibration curve. As can be seen, the results obtained for valsartan from tablets are less variable than those obtained from the validation standards, while for hydrochlorothiazide it is the opposite behaviour. The results of precision and accuracy obtained with this calibration curve for the analysis of valsartan and hydrochlorothiazide in authentic batch samples are given in Table 3. Nonetheless, the 95%  $\beta$ -expectation tolerance intervals of both analytes obtained with the authentic batch samples are fully included into the  $\pm 10\%$  acceptance limits, thus confirming the validity of the method and the selection of the calibration curve.

## 4. Discussion

From the three examples shown, it can be seen that adding the analysis of authentic batch samples as soon as they are available confirms and strengthens the decision about the validity of the method. Special care has been done when analysing the authentic batch samples to reduce as much as possible the increasing variability that may arise from the batches heterogeneity by using the same initial sample of tablets or syrups for one analysis to the other, especially for the two HPLC-UV methods. It therefore annihilates the increase in variability that may be due to the batch heterogeneity. While for the UV spectrophotometric method this was not done, the increase of variability that may originate from the batch heterogeneity is moderate. Nonetheless, some differences in the methods variability were observed between the uses of the two types of samples in the three case studies. Table 4 shows the  $p$ -values of the statistical tests of differences between the intermediate precision variance obtained for each type of sample for the three methods. As can be seen, they are all non-significant ( $p$ -values  $> 0.05$ ). This suggests that the differences observed are only due to the random sampling, and not coming from a special cause,



**Fig. 4.** Accuracy of the results obtained for the HPLC-UV method for the determination of methyl parahydroxybenzoate for the authentic batch samples (syrups) with the simple linear regression model, with the linear model forced through the origin and fitted only with the 100.0% calibration level or fitted only with the 140.0% calibration level. The vertical continuous segment are the 95%  $\beta$ -expectation tolerance intervals obtained with the authentic batch samples for each of the three batches, the dashed lines are the 95%  $\beta$ -expectation tolerance limits obtained with the 100% concentration level of the spiked validation standards and the dotted curves represent the acceptance limits ( $\pm 10\%$ ).

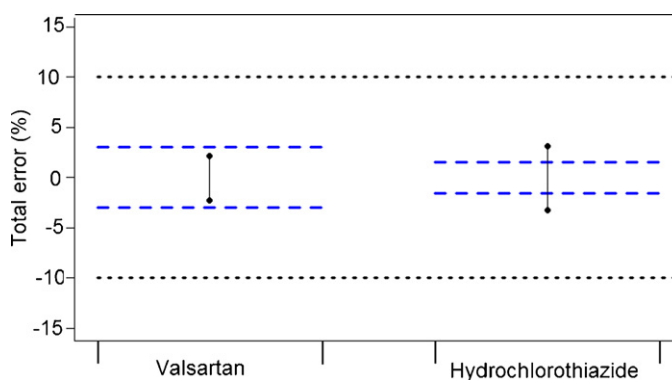


**Fig. 5.** Accuracy profiles of the HPLC-UV method for the simultaneous determination of (a) valsartan and (b) hydrochlorothiazide obtained for the spiked validation standards with the linear model forced through the origin and fitted only with the 120.0% calibration level. The plain line is the relative bias, the dashed lines are the 95%  $\beta$ -expectation tolerance limits and the dotted curves represent the acceptance limits ( $\pm 10\%$ ). The dots represent the relative back-calculated concentrations of the validation standards.

thus confirming that, for these examples, the validation standards mimicked adequately the authentic batch samples.

None of the case studies depicted in this work showed the situation where the analysis of authentic batch impairs the conclusion of validity obtained with the validation standards. In such a situation, the regulatory guideline ICHQ2R1 [11] is the rule and should be followed. As this document requires to validate the method over a concentration range (e.g. 80–120% for the active substance in a drug product), that is only achievable using validation standards, the validity of the method should be defined by the analysis of those placebo based validation standards. Similarly the selection of the adequate standard curve should be defined when using validation standards. The use of authentic batch samples is here seen as a valuable confirmation of the method validity.

However, when the analysis of the authentic batch samples does not confirm the validity of the method, investigation of the origins of this failure should be realised as it leaves doubts about the usefulness of the method. Sources of these differences may arise from the sample preparation of the authentic batch samples. Indeed, this step is not necessarily present or similar when using placebo based validation standards. For instance some sources of additional variability may come from the segregation of electro-



**Fig. 6.** Accuracy of the results obtained for HPLC-UV method for the simultaneous determination of valsartan and hydrochlorothiazide for the authentic batch samples (tablets) with the linear model forced through the origin and fitted only with the 120.0% calibration level. The vertical continuous segments are the 95%  $\beta$ -expectation tolerance intervals obtained with the authentic batch samples, the dashed lines are the 95%  $\beta$ -expectation tolerance limits obtained with the 100% concentration level of the spiked validation standards and the dotted curves represent the acceptance limits ( $\pm 10\%$ ).

**Table 4**  
*p*-Values obtained for the test of differences in variances obtained with the placebo spiked validation standards versus the authentic batch samples for each analyte studied: metformin, methyl parahydroxybenzoate, valsartan and hydrochlorothiazide.

Analyte	Method	Calibration curve	Batch no	<i>p</i> -Value
Metformin	Spectrophotometric-UV	Linear 0–100%	1	0.404
Methyl parahydroxybenzoate	HPLC-UV	Linear 0–140%	1	0.729
			2	0.712
			3	0.232
Valsartan	HPLC-UV	Linear 0–120%	1	0.365
Hydrochlorothiazide				0.152

static analytes when vigorously crushing tablets, from variability of the dissolution of the authentic samples, from the additional steps of filtration or even from inadequate homogenisation of the initial sample. Another additional source of variability could come from the matrix of the placebo based validation standards that may not match perfectly the one of the authentic batch samples.

The aim of adding to the method validation the analysis of authentic batch samples is to increase the confidence the analysts, laboratory, patients as well as regulatory bodies can have in the results generated by the analytical method under study.

## 5. Conclusions

Through the three examples presented in this work, the use of authentic batch samples (tablets or syrups) has been shown to valuably increase the reliability of the decision to declare an analytical method valid using only spiked placebo validation standards. Using authentic batch samples for the evaluation of results accuracy has been shown feasible. This study further illustrated that, depending on the type of response function used as calibration curve, there were various degrees of difference in the methods performance between the two types of samples: spiked validation standards and authentic batch samples. Thus adding these last samples into the validation design may help the analyst to select the most fit for purpose calibration curve and thus increase the reliability of the results generated by the method in routine application. Finally, using reconstituted validation standards seems a coherent practice for analytical methods validation as it has been shown in this work that they mimicked relatively well the behaviour of authentic batch samples. Although this last statement may not be applicable to all situations, the inclusion of authentic batch samples in any analytical method validation step would efficiently help analyst to answer the question about the transferability of the validity of a method from spiked validation standards to authentic batch samples.

## Acknowledgments

The authors are very grateful to the anonymous reviewers for providing important comments that led to significant improvements of this article. A research grant from the Belgium National Fund for Scientific Research (FRS-FNRS) to E. Rozet is gratefully acknowledged.

## References

- [1] International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use Topic Q2 (R1): Validation of Analytical Procedures: Text and Methodology, Geneva, 2005.
- [2] A. Bouabidi, E. Rozet, M. Fillet, E. Ziemons, E. Chapuzet, B. Mertens, R. Klinkenberg, A. Ceccato, M. Talbi, B. Streel, A. Bouklouze, B. Boulanger, Ph. Hubert, Critical analysis of several analytical method validation strategies in the framework of the fit for purpose concept, *J. Chromatogr. A* 1217 (2010) 3180–3192.
- [3] R.D. Marini, P. Chiap, B. Boulanger, S. Rudaz, E. Rozet, J. Crommen, Ph. Hubert, LC method for the determination of R-timolol in S-timolol maleate: validation of its ability to quantify and uncertainty assessment, *Talanta* 68 (2006) 1166–1175.
- [4] J. Mantanus, E. Ziémons, P. Lebrun, E. Rozet, R. Klinkenberg, B. Streel, B. Evrard, Ph. Hubert, Active content determination of non-coated pharmaceutical pellets by near infrared spectroscopy: method development, validation and reliability evaluation, *Talanta* 80 (2010) 1750–1757.
- [5] E. Ziemons, J. Mantanus, P. Lebrun, E. Rozet, B. Evrard, Ph. Hubert, Acetaminophen determination in low-dose pharmaceutical syrup by NIR spectroscopy, *J. Pharm. Biomed. Anal.* 53 (2010) 510–516.
- [6] R.D. Marini, E. Rozet, M.L.A. Montes, C. Rohrbasser, S. Roht, D. Rhème, P. Bonnabry, J. Schappler, J.-L. Veuthey, Ph. Hubert, S. Rudaz, Reliable low cost capillary electrophoresis device for drug quality control and counterfeit medicines, *J. Pharm. Biomed. Anal.* 53 (2010) 1278–1287.
- [7] E. Rozet, A. Ceccato, C. Hubert, E. Ziemons, R. Oprean, S. Rudaz, B. Boulanger, Ph. Hubert, Analysis of recent pharmaceutical regulatory documents on analytical method validation, *J. Chromatogr. A* 1158 (2007) 111–125.
- [8] Ph. Hubert, J.-J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.-A. Compagnon, W. Dewé, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, Harmonization of strategies for the validation of quantitative analytical procedures A SFSTP proposal—part I, *J. Pharm. Biomed. Anal.* 36 (2004) 579–586.
- [9] Ph. Hubert, J.J. N'guyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewé, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, E. Rozet, Harmonization of strategies for the validation of quantitative analytical procedures. A SFSTP proposal—part II, *J. Pharm. Biomed. Anal.* 45 (2007) 70–81.
- [10] Ph. Hubert, J.J. N'guyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewé, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, E. Rozet, Harmonization of strategies for the validation of quantitative analytical procedures. A SFSTP proposal—part III, *J. Pharm. Biomed. Anal.* 45 (2007) 82–96.
- [11] D. Hoffman, R. Kringle, Two-sided tolerance intervals for balanced and unbalanced random effects models, *J. Biopharm. Stat.* 15 (2005) 283–293.
- [12] D. Hoffman, R. Kringle, A total error approach for the validation of quantitative analytical methods, *Pharm. Res.* 24 (2007) 1157–1164.
- [13] G.J. Hahn, W.Q. Meeker, *Statistical Intervals: A Guide for Practitioners*, Wiley Edition, 1991, p. 392.
- [14] R.W. Mee,  $\beta$ -Expectation and  $\beta$ -content tolerance limits for balanced one-way ANOVA random model, *Technometrics* 26 (1984) 251–254.
- [15] S.R. Searle, G. Casella, C.E. McCulloch, *Variance Components*, Wiley, 1992.

## **CHAPITRE VIII : Conclusions & perspectives**

**1. Conclusions**

**2. Perspectives**

## 1. Conclusions

Dans le but d'améliorer la fiabilité des décisions fondées sur les résultats d'analyse, ce travail s'est focalisé sur une étape primordiale du cycle de vie des méthodes analytiques pour les industries pharmaceutiques, à savoir la validation analytique. Son rôle est majeur en raison de sa position stratégique dans ce cycle.

En effet, c'est la toute dernière phase où la fiabilité et donc l'exactitude des résultats qui seront générés par ces méthodes peut être évaluée avant leur application en routine. Les décisions quotidiennes prises au sujet de la libération d'un lot sur le marché, la bioéquivalence de deux formulations pharmaceutiques ou la déclaration qu'un patient est ou n'est pas en bonne santé dépend directement de la confiance qu'on a sur les résultats générés. La fiabilité est évaluée par l'exactitude des résultats en accord avec les besoins spécifiques du client ou de l'utilisateur final.

La validation est généralement réalisée en évaluant les critères habituels de validation et en les comparant à des niveaux acceptables: c'est ce qu'on appelle les approches classiques de validation des méthodes analytiques.

Par opposition, une autre approche étudiée dans ce travail est différente des autres approches classiques car avec elle, il n'est plus question d'évaluer les performances de la méthode telles que la justesse ou la fidélité, mais plutôt la fiabilité des résultats obtenus avec ces méthodes d'analyse. Les paramètres de performance de la méthode sont toujours utilisés, mais l'attention n'est plus centrée sur eux. Cette approche est connue sous le nom de profil d'exactitude ou d'erreur totale.

Même si les critères de validation sont bien connus dans les documents réglementaires tels que ceux issus de ICH (2005), de la FDA (2001), d'EURACHEM, .... une étude détaillée et comparative a été réalisée concernant leurs définitions, leurs interprétations, les plans expérimentaux et les critères d'acceptation de chaque critère et ce en raison des nombreuses incohérences disponibles dans ces documents réglementaires. En effet, ces critères doivent être évalués tout en gardant à l'esprit que l'objectif de la phase de validation

devrait être lié au véritable objectif de chaque méthode d'analyse quantitative: obtenir des résultats exacts. Parmi les premiers points principaux soulignés se trouve la divergence existante dans ces textes quant aux critères de linéarité, qui exprime la relation entre la vraie concentration en analyte dans les échantillons et les résultats obtenus en utilisant la méthode analytique, et la fonction de réponse, étant la relation entre la concentration d'analytes dans l'échantillon et le signal de réponse du détecteur utilisé. De même, le deuxième point principal est de mettre en évidence les différences entre la justesse et la fidélité de la méthode et l'exactitude des résultats. Seul ce dernier critère révèle la fiabilité des résultats.

Par ailleurs, une analyse critique des différentes approches classiques versus l'approche du profil d'exactitude a été réalisée en les appliquant à des exemples pratiques de dosage de principe actif dans le médicament.

Dans la nouvelle méthodologie de validation présentée tout au long de cette thèse, la question de la qualité des résultats futurs est soulignée, puisque l'objectif même d'une procédure d'analyse est de fournir des résultats. L'accent est donc mis sur la "satisfaction client": le profil d'exactitude consiste à fixer le pourcentage minimal prévu des résultats futurs à inclure à l'intérieur des limites d'acceptation fixées a priori. En outre, il a été démontré que cette approche peut s'appliquer à tout type de procédures analytiques rencontrées dans l'industrie pharmaceutique via différentes études de cas incluant différentes matrices existantes pour le médicament.

Considéré comme des processus de petite taille et afin de tenter de mesurer à nouveau la performance des méthodes d'analyse déjà validées par l'approche de profil d'exactitude, une étude de capabilité a été réalisée en adaptant l'indice de capabilité Cpk, utilisé couramment pour les processus de fabrication, au contexte des méthodes analytiques en proposant un nouvel indice spécifique à ces procédés.

Généralement, la validation implique des standards de validation ou des échantillons de contrôle de qualité (CQ) qui sont préparés dans un placebo ou dans une matrice

reconstituée. Toutefois, l'un des principaux soucis de cette approche est qu'elle peut ne pas prendre en considération une importante source de variabilité qui provient du processus de fabrication du produit fini. La dernière partie de ce travail a donc essayé de répondre partiellement, et pour la première fois, à la question de transférabilité des performances quantitatives des standards de validation à des échantillons authentiques réelles du médicament, puisque la plupart du temps seuls les échantillons préparés artificiellement sont utilisés comme des standards de validation dopés.

Une étape principale du cycle de vie des méthodes analytiques a été étudiée dans cette thèse. L'amélioration de la qualité prédictive de la méthodologie étudiée pour évaluer la validation de la méthode nous aidera certainement à améliorer la fiabilité des résultats obtenus par la méthode d'analyse et à accroître la confiance dans les décisions critiques subséquentes qui sont faites avec leur aide.

## 2. Perspectives

Les résultats présentés dans cette thèse de doctorat sont une petite partie de ce qui pourrait être fait dans le cadre de l'évaluation de la fiabilité des résultats obtenus par des méthodes analytiques au long de leur cycle de vie. La validation de la méthode est un bon point de départ. Néanmoins, il existe plusieurs autres questions qui n'ont pas été traitées et qui pourraient faire l'objet de recherches futures.

La continuité ciblée de cette recherche pourrait se focaliser en particulier sur l'étude de l'universalité de cette approche en l'appliquant à différentes méthodes et matrices rencontrées dans d'autres domaines d'activité tels que l'industrie agroalimentaire et autres industries de chimie.



Dans le même contexte général, d'autres perspectives potentielles de recherches peuvent être envisagées:

- Etude critique des différentes pratiques, parfois erronées, rencontrées dans les laboratoires de contrôle.
- Développer et proposer des protocoles standards pour la validation des méthodes analytiques des différents tests dans l'industrie pharmaceutique : tests des impuretés, de dissolution .....
- Proposer des limites d'acceptation de ces méthodes toujours en relation avec l'objectif des méthodes et ce afin de combler les lacunes dans les guidelines concernant ces limites
- Orienter nos recherches vers un autre type de méthodes analytiques, celles semi-quantitatives et, en particulier, vers la validation de ces méthodes ou leur transfert.

Tous ces travaux futurs, combinés avec la présente thèse, pourraient avoir une finalité commune: ouvrir la voie vers le principe de la qualité totale exigée par l'industrie, et en particulier l'industrie pharmaceutique, tout en améliorant la confiance entre le consommateur et les acteurs économiques vers les principales décisions prises à l'aide des méthodes analytiques.

## Travaux scientifiques

- Liste des publications :

1. **Critical analysis of several analytical method validation strategies in the framework of the fit for purpose concept**

*Journal of Chromatography A, Volume 1217, Issue 19, 7 May 2010, Pages 3180-3192*

2. **Flexibility and Applicability of  $\beta$ -expectation tolerance interval approach to assess the fitness of purpose of pharmaceutical analytical methods”**

*Journal of Drug Testing and Analysis, In press, DOI 10.1002/dta.1345*

3. **Usefulness of capability indices in the framework of analytical methods validation**

*Analytica Chimica Acta, Volume 714, 10 February 2012, Pages 47-56*

4. **Do placebo based validation standards mimic real batch products behaviour? Case studies**

*Journal of Pharmaceutical and Biomedical Analysis 55 (2011) 583–590*

- **Communications nationales et internationales**

Nom de congrès	Type	Date	Lieu	Organisé par
Analytical Sciences Congress – 2006	orale	01 au 03/11/2006	Marrakech, Maroc	Biocap Association, Morocco
Analytical Sciences Congress – 2009	poster	29/10/2008	Casablanca, Maroc	Biocap Association, Morocco
HPLC 2009	poster	28/06/2009 au 02/07/2009	Dresden, Germany	German Chemical Society, GDCh
DRUG ANALYSIS 2010	Poster	Sept 2010	Anvers, Belgique	Drug analysis

### Résumé:

Les résultats issus des méthodes analytiques ont un rôle essentiel dans de nombreux domaines suite aux décisions qui sont prises sur leur base, telles que la détermination de la qualité et la quantité des principes actifs, des spécialités pharmaceutiques, des nutriments ou autres échantillons tels que ceux d'origine biologique impliqués dans les études pharmacocinétiques ou de biodisponibilité et bioéquivalence. La fiabilité des résultats analytiques est primordiale dans ce contexte et ils doivent surtout être en accord avec les besoins des utilisateurs finaux, et ce pour s'assurer de la fiabilité des résultats qui seront fournis lors des analyses de routine.

Pour assurer cette fiabilité, la validation des méthodes analytiques est donc d'une importance primordiale. La validation des méthodes analytiques est une des étapes essentielles dans le processus intégral de démonstration de la fiabilité des résultats analytiques qui seront générés par ces méthodes dans leurs utilisations quotidiennes. En outre, l'interprétation et l'évaluation des résultats d'analyse au cours de cette étape-clé doivent être fondées sur des bases statistiques. Sinon, l'importance et l'utilité des résultats d'analyse obtenus pour la prise de décision restent non clairs.

On est confronté à plusieurs approches de validation analytique dans la littérature et dans les différents textes réglementaires.

L'objectif principal de cette thèse est de comparer ces différentes approches en théorie comme en pratique, afin de statuer sur la fiabilité des décisions prises, par chaque approche, au moyen des résultats obtenus durant la validation des méthodes analytiques.

Pour atteindre cet objectif, nous avons tout d'abord reprecisé l'objectif de toute méthode analytique quantitative et de sa validation. Nous avons ensuite réalisé une étude comparative des différents textes réglementaires en dégagant les erreurs et confusions incluses dans ces documents et en particulier leurs implications pratiques lors de l'évaluation de la validité d'une méthode.

Compte tenu de ces constatations, une nouvelle approche pour évaluer la validité des méthodes analytiques quantitatives a été proposée et détaillée d'un point de vue statistique. Elle se base sur l'utilisation d'une méthodologie statistique utilisant un intervalle de tolérance de type «  $\beta$ -expectation » qui a été transposée en outil de décision final appelé profil d'exactitude. Ce profil permet de garantir qu'une proportion définie des futurs résultats qui seront fournis par la méthode lors de son utilisation en routine sera bien incluse dans des limites d'acceptation fixées a priori en fonction des besoins des utilisateurs.

A cet effet, une étude comparative pratique entre cette approche et les différentes approches classiques existantes dans le domaine analytique a été réalisée. Puis une illustration de l'universalité potentielle de cette méthodologie a été faite à travers l'application de différentes méthodes d'analyse rencontrées dans le domaine pharmaceutique ainsi qu'une étude de capabilité des méthodes d'analyse, validées par cette approche, en les considérant comme des procédés à part entière et en proposant de nouveaux indices de capabilité adaptés à ce genre de procédés. Enfin, une étude de certaines pratiques rencontrées dans les laboratoires d'analyse a été faite ; celle qui consiste à l'utilisation des placebos en tant que standards de validation à la place des échantillons réels issus de la fabrication en routine.

**Abstract**

The analytical methods' results play a crucial role in many areas on the decisions made on their basis, such as determining the quality and quantity of active ingredients of drug products, the quality of food or in assessing pharmacokinetic studies or bioavailability and bioequivalence. The reliability of analytical results is paramount in this context, as it should certainly meet the needs of end users, in order to ensure the reliability of the conclusions taken with them.

To attain this reliability, the validation of analytical methods is therefore of chief importance. Validation of analytical methods is one of the essential steps in the entire process of demonstrating reliability of analytical results that will be generated by these methods in their routine use. In addition, the interpretation and evaluation of test results during this step should be based on a statistical basis. Otherwise, the importance and usefulness of the analytical results for decision making is irrelevant.

We are acquainted to several analytical validation approaches in both the literature and in various regulations' text. The main objective of this thesis is to compare these various approaches in theory and in practice, in order to determine the reliability of the decisions taken by each of them.

To achieve this goal, we first redefined the aim of any quantitative analytical method and of their validation. We then conducted a comparative study of various regulations by identifying errors and confusions included in these documents and in particular their practical implications when assessing the validity of a method.

Given these findings, a new approach for assessing the validity of quantitative analytical methods have been studied and enlarged a statistical point of view. It is based on a statistical methodology using a " $\beta$ -expectation" tolerance interval that is incorporated into final decision-making tool called accuracy profile. This profile ensures that a specified proportion of the future results that will be provided by the method during its routine use is well included in the acceptance limits fixed a priori according to the users needs.

For this purpose, a practical comparative study between this approach and different conventional approaches existing was performed. Then an illustration of the potential universality and versatility of this methodology has been made through its application of different analytical methods encountered in the pharmaceutical field. Also a new capability index has been proposed that is suitable for the assessment of an analytical method capability in the context of its validation. Finally, a study of certain practices encountered in laboratories was made, which involves the use of placebos as validation standards instead of actual samples from the manufacturing routine.