European Journal of Pharmaceutical Sciences 190 (2023) 106566

Contents lists available at ScienceDirect

European Journal of Pharmaceutical Sciences

journal homepage: www.elsevier.com/locate/ejps

Review

# The Global Bioequivalence Harmonisation Initiative (GBHI): Report of the fifth international EUFEPS/AAPS conference

M. Mehta [a], B. Schug [b,*], H.H. Blume [c,d], G. Beuerle [e], W. Jiang [a], J. Koenig [f], P. Paixao [g], N. Tampal [a], Y.-C. Tsang [h], J. Walstab [b], R. Wedemeyer [b], J. Welink [i]

[a] U.S. Food and Drug Administration (FDA), Silver Spring, MD, USA
[b] SocraTec R&D GmbH, Oberursel/Erfurt, Germany
[c] SocraTec C&S GmbH, Oberursel, Germany
[d] Frankfurt Foundation Quality of Medicines, Frankfurt/Main, Germany
[e] Teva, Ulm, Germany
[f] Federal Institute for Drugs and Medical Devices (BfArM), Bonn, Germany
[g] Research Institute for Medicines (iMed.ULisboa), Faculty of Pharmacy, University of Lisbon, Portugal
[h] Apotex Inc., Toronto, Canada
[i] Medicines Evaluation Board, Utrecht, the Netherlands

ARTICLE INFO

*Keywords:*
Adaptive design
Bioequivalence
Fed/fasted
Highly variable drugs
Locally acting locally applied
Modelling
Narrow therapeutic index drug
Replicate design
Scaling
Topical product

ABSTRACT

The series of conferences of the Global Bioequivalence Harmonisation Initiative (GBHI) was started in 2015 by the European Federation for Pharmaceutical Sciences (EUFEPS). All GBHI meetings so far were co-organised together with the American Association of Pharmaceutical Scientists (AAPS). Beginning with the 3rd workshop US-FDA joined as co-sponsor – to support global harmonisation of regulatory recommendations for bioequivalence (BE) assessment.

At the 5th GBHI conference, the following BE topics were intensively discussed, and the following main conclusions were drawn:

(1) Statistical considerations for BE assessment in specific situations covering scaling approaches for highly variable drug (HVD) products, two-stage adaptive design and opportunities of modelling and simulation to support BE: even though special BE study concepts like adaptive designs are not often used in practise so far, a majority of the workshop participants were in favour of a more frequent application of such approaches. The regulatory conditions relevant in this context need further concretisation and harmonisation between the regions. Moreover, modelling and simulation were considered as a promising and evolving approach, also for BE development programmes.

(2) Fed versus fasting conditions in BE trials: Findings that BE between generic products could be confirmed only after fasted administration but failed under fed conditions seem more an exception than the rule. Obviously, BCS class IV compounds are most problematic in this context. Differences in critical excipients such as surfactants or pH-modifiers may be relevant reasons for different sensitivity for interactions in fasted versus fed conditions. Consequently, such deviations in composition of generic preparations should be avoided. Moreover, confirmation of BE may be generally difficult comparing different dosage forms, such like capsules versus tablets, especially in fed state.

(3) BE assessment of locally acting drug products applied topically to the skin: Appropriateness and potential benefit of *in-vitro* tests as alternatives to clinical efficacy studies have been comprehensively discussed. In addition to the already well-established *in-vitro* release and permeation tests, other techniques were suggested, e.g., Raman spectroscopy or dermal open flow microperfusion. Validation of those methods is challenging and, despite significant progress already achieved during previous years, more research is needed before they may be fully accepted for regulatory purposes.

(4) BE evaluation of narrow therapeutic index (NTI) drugs: The discrepancies amongst regulatory agencies in necessity of tighter BE acceptance ranges, the recommendations for inclusion of peak and total drug exposure into BE assessment with more restrictive criteria and the importance of comparison of the product-related within-

* Corresponding author.
*E-mail address:* barbara.schug@socratec-pharma.de (B. Schug).

https://doi.org/10.1016/j.ejps.2023.106566
Received 11 July 2023; Accepted 14 August 2023
Available online 15 August 2023

subject variability for NTI drugs were debated. Arguments in favour and against the different approaches were presented and discussed but need further consideration before harmonisation can be achieved.

The highly interactive meeting and extensive exchange between regulators and scientists from industry and academia resulted in useful progress in open BE issues and supported the goal of science-driven harmonisation.

## 1. Introduction

This report summarises the discussion of bioequivalence (BE) issues at the fifth Global Bioequivalence Harmonisation Initiative (GBHI) conference, held 28–29 September 2022 in Amsterdam, The Netherlands, and provides a further step forward towards harmonised criteria of BE assessment. The GBHI series of international conferences fosters an open discussion of viewpoints amongst pharmaceutical scientists from regulatory agencies, academia and industry. The specific workshop format provides scientists from academia and industry the opportunity to contribute actively to the regulatory process by presenting experimental results which are considered relevant to solve open issues. The initiative – by providing a scientific exchange platform – already significantly contributed to data exchange and more harmonised BE assessment around the globe (Blume et al., 2021; Chen et al., 2018, 2019; Mehta et al., 2020). Therefore, the conference aims at contributing scientifically also to official harmonisation processes like the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH).

Since the preceding GBHI conference held in 2019, the first harmonised BE guidance that has been developed under ICH, i.e., draft guidance on bioequivalence for immediate-release (IR) solid oral dosage forms, M13A, was endorsed in December 2022 (ICH, 2022a). Several topics discussed at the earlier GBHI conferences, including BE of oral IR drug products under fasting and/or fed conditions, prodrugs and compounds with pre-systemic extraction (moieties to be measured: parent versus metabolite), exclusion of pharmacokinetic (PK) data in BE assessment, were timely to support development of M13A.

For the 2022 GBHI conference, the International Scientific Planning Committee, composed of representatives from academia, industry, contract research organisations, and experts of the European Health Authorities and the US-Food and Drug Administration (US-FDA), selected the following BE topics for which realistic chances of harmonisation based on the scientific state of the art exist or is worth to strive for: (1) Statistical considerations for BE assessment in specific situations covering a) two-stage, adaptive and replicate design and b) the possibility to use modelling and simulation in the context of BE assessment, (2) Fed versus fasting conditions in BE trials: current status and new insights (3) BE assessment of locally acting drug products applied topically to the skin, and (4) BE evaluation of narrow therapeutic index drugs (NTIs).

Participants entered in a lively exchange and re-evaluation of certain topics that had already been discussed in previous GBHI workshops – however, without final conclusions on all open questions such as the necessity of BE studies in fed and fasted state for IR solid oral dosage forms and statistical issues around BE evaluation of highly variable drug (HVD) products. Moreover, the new topics including PK modelling approaches and BE assessment of topical drug products and NTIs were intensively discussed. In addition to and in support of the discussions within the talks, a representative picture of the different viewpoints was captured by giving the audience the chance to respond to specific questions of each session via an online voting tool. These survey results are presented in the **Supplemental Table**.

This report summarises the presentations and discussions, highlights the recommendations for harmonisation and identifies still unresolved differences on the discussed topics.

## 2. BE topics

### 2.1. Statistical considerations for BE assessment in specific situations

#### 2.1.1. Scaling approaches for highly variable drug (HVD) products

Regulatory agencies have strived for alternative statistical solutions to overcome the biometrical challenge associated with BE assessment of HVD products, which involves inappropriately high sample sizes when applying conventional BE statistical approaches. HVD products by definition have an estimated within-subject variability (WSV) of ≥30% for the PK parameters area under the curve (AUC) and/or maximum concentration ($C_{max}$) (Midha et al., 2005; Shah et al., 1996). HVDs generally are expected to come along with a wide therapeutic index, i.e., the reference product[1] was demonstrated to be safe and efficacious despite high intraindividual PK variability in pivotal trials. Most HVDs fall into BCS class II (=low aqueous solubility–high intestinal permeability) or BCS class IV (=low aqueous solubility–low intestinal permeability). Typical characteristics of such compounds include extensive first-pass effect, low bioavailability and high lipophilicity, leading to low plasma concentrations eventually resulting in high WSV (summarised by Davit et al., 2012). The large sample sizes required to satisfy the conventional BE acceptance range for the 90% confidence interval (CI) of the geometric mean ratio (GMR) of 80.00–125.00% result in ethical and economical concerns.

To maintain reasonable sample sizes for BE studies of HVD products, scaling approaches allowing for "widening" of BE limits are generally accepted by regulatory agencies, but so far, the accepted approaches still differ worldwide.

The different scaling approaches for HVD products have been already discussed in the 2nd GBHI conference (Chen et al., 2019) and were revisited and supplemented by new analyses and aspects.

*2.1.1.1. Regulatory recommendations.* Recommendations on BE assessment of HVD products have been described previously (Chen et al., 2019) and are summarised in the following:

A scaling approach generally involves an estimate of the WSV for the reference product and, therefore, a BE study with (semi-)replicate design in which at least the reference product is given more than once to the subjects, i.e., either a three-period (i.e., partially replicated) or four-period (i.e., fully replicated) crossover scheme can be applied. Although US-FDA, European Medicines Agency (EMA) and Health Canada (HC) take a "scaled average bioequivalence" approach and allow scaling starting with a shown WSV of 30%, their recommended scaling approaches are not harmonised (Table 1).

All three agencies have defined a point estimate constraint (PEC) on the GMR, which should be within 80.00–125.00%. However, the PK parameters for which scaling is allowed and the methods of scaling are different: the US-FDA recommends scaling for both AUC and $C_{max}$. The EMA applies the scaling approach only for $C_{max}$, while for AUC, the conventional average BE assessment needs to be applied. HC recommends scaling for AUC alone, whilst for $C_{max}$ only the PEC applies

[1] For US-FDA, the term reference product as used in this paper means either the reference listed drug, which is the listed drug identified by US-FDA as the drug product upon which an applicant relies in seeking approval of its ANDA, or the reference standard, which is the drug product selected by US-FDA that an applicant seeking approval of an ANDA must use in conducting an in-vivo bioequivalence study required for approval. See 21 CFR 314.3.

**Table 1**
Comparison of bioequivalence criteria for highly variable drug products.

| | US-FDA | EMA | HC |
|---|---|---|---|
| Recommended criteria for AUC | Mixed (ABE or RSABE) | ABE | Mixed (ABE or ABEL) |
| Recommended criteria for $C_{max}$ | Mixed (ABE or RSABE) | Mixed (ABE or ABEL) | PEC |
| Scaling factor (regulatory constant) | 0.25 | 0.294 | 0.294 |
| Cap maximum BE margin | No | 50% CV | 54.7% CV |
| PEC for AUC and $C_{max}$: GMR contained in [0.8, 1.25] | Yes | Yes | Yes |

ABE, average bioequivalence; ABEL, average bioequivalence expanding limit; AUC, area under the curve; BE, bioequivalence; $C_{max}$, maximum concentration; CV, coefficient of variation; EMA, European Medicines Agency; GMR, geometric mean ratio; HC, Health Canada; PEC, point estimate constraint; RSABE, reference-scaled average bioequivalence; US-FDA, US-Food and Drug Administration (EMA, 2010; Health Canada, 2018; US-FDA, 2021).

(EMA, 2010; Health Canada, 2018; US-FDA, 2021).

The US-FDA developed a "reference-scaled average bioequivalence" (RSABE) approach, whereby the BE acceptance limits are scaled to the WSV of the reference product. In this scenario, the observed WSV of the reference product in a specific BE study is compared to a cut-off value of 0.294 (corresponding to a WSV of approximately 30%), above which reference scaling is used. The US-FDA recommends 0.25 as the scaling factor (regulatory constant). The extent to which the BE limits can be widened is not defined (US-FDA, 2021).

In contrast, EMA and HC use "average bioequivalence with expanding limits" (ABEL), if the determined WSV of reference is ≥30%. Thereby, the expanding limits are based on the observed WSV in the BE study; they can be widened to a maximum of 69.84–143.19% ("upper cap" affiliated with 50% WSV) according to EMA or to a maximum of 66.7–150.0% according to HC ("upper cap" affiliated with 57.4% WSV). The regulatory constant is set to 0.294 (EMA, 2010; Health Canada, 2018).

#### 2.1.1.2. *Invited contributions and discussion*

*New insights on the consumer risk (type-I error) and assessment of the relevance of different BE scaling approaches.* Although statistical models differ between regulatory authorities, a general concern of the scaling methods used is the inflation of type-I error rate in the neighbourhood of 25% (for US-FDA) or 30% WSV (for EMA and HC) of the reference product: if a drug product is falsely classified as HVD product, the chance of passing BE increases when the scaled approach instead of the correct average BE approach is used, translating into an inflated type-I error (=increased consumer risk).

An analysis of US-FDA applying the three agencies' (FDA, EMA, HC) BE criteria to generic HVD products approved by US-FDA in the past years shows that the three criteria were consistent in passing a majority of the BE trials, despite their differences in the PK parameters to which scaling is applied. US-FDA had the highest and EMA had the lowest passing rate. The observed discrepancy in passing BE results amongst the three agencies was mostly driven by whether scaling was applied to AUC or not and less by the difference in the scaling methods used.

According to another analysis of US-FDA, for a sample size of $n = 24$, the type-I error rate at the respective cut-off (25 or 30%) of WSV of the reference product (maximum consumer risk) was lower when applying US-FDA's RSABE method compared to the ABEL approach used by EMA or HC – while maintaining a higher power of the study due to the application of RSABE to both AUC and $C_{max}$. This conclusion was challenged, however, by the audience where it was argued that the consumer risk in real life using the RSABE criteria can approach about 13% for a sample size of $n = 24$ or even be higher for larger sample sizes (Schütz et al., 2022). Similar estimations of type-I error reaching 7–8% with the EMA procedure and 13–18% with the US-FDA approach were reported previously (e.g., Muñoz et al., 2016). FDA clarified that the reported type-1 error rate for the US-FDA approach is not its true consumer risk, because the simulation did not distinguish the population parameter within-subject standard deviation of the reference product (σWR) and the observed sample estimate sWR when evaluating the type-I error rate for US-FDA's criteria (RSABE).

The US-FDA suggested that EMA and HC may adopt US-FDA's approach of applying Howe's Approximation to incorporate the uncertainty in the observed sample estimate $s_{WR}$ (i.e., by using the upper limit of the 90% CI for $s_{WR}$ based on Chi-square distribution). This would decrease the type-I error rate, while maintaining almost identical power as the current ABEL criteria. For US-FDA, implementation of an upper cap to scaling similar to EMA and HC was proposed. However, whether a constraint is necessary for WSV of reference beyond which widening of the BE limits is not recommended, is still a matter of debate. The main argument against "capping" is that the PEC of 80.00–125.00% for the GMR will dominate properties in case of large variability, therefore, capping presumably plays a minor role.

From the audience it was argued that the approach of empirically iteratively adjusting α would maintain the consumer risk at the desired level of ≤5% under certain conditions (Molins et al., 2017; Ocaña and Muñoz, 2019). It was further mentioned that only "average bioequivalence" with fixed widened limits would always maintain the consumer risk at ≤5%, however, this concept was not elaborated further in the discussions (for details see Schütz et al., 2022).

Whereas the US-FDA recommends scaling for both $C_{max}$ and AUC, in the EMA approach, scaling is only allowed for $C_{max}$ (EMA, 2010; US-FDA, 2021). EMA's scientific rationale is the following: generally, $C_{max}$ is more variable than AUC, therefore, for this parameter, the need for a solution on how to deal with high variability is greater. Also, AUC is mostly the more relevant efficacy parameter, leading to the decision not to allow widening of BE limits for AUC. Instead, the FDA approach is based on the rationale that per se the high variability of a PK parameter is not of concern for using a scaled approach, as these drugs in general are efficacious and safe over a broad therapeutic range.

#### 2.1.1.3. *Conclusions and suggestions.*

Overall, it is desirable to find an optimal harmonisation strategy that can increase the concordant approval rate among the different agencies, minimize the inconsistency with historically approved products, and at the same time reduce the consumer risk.

When asked about their preferences in a survey (see **Supplemental Table**), participants were equally split on the question of whether RSABE or ABEL is the preferred statistical approach. A vast majority of the attendees preferred applying scaling procedures also to AUC, obviously supporting the concept that HVDs need to have a broad therapeutic range. Application of widening to AUC would be a major step towards increasing the concordance rate among US-FDA, HC and EMA. The majority of meeting participants think that the regulatory agencies should give a clear recommendation on the method to be used for BE evaluation of HVD products. Some participants suggested that recommendations on the scaling method and the PK parameters for which scaling should be allowed may be included in product-specific guidances (PSGs). This proposal would allow for a scientifically based BE approach considering the pharmacological aspects of a certain drug or drug class and the clinical relevance of the endpoint.

Notably, the topic of statistical assessment of studies with HVD products will be addressed in tier 3 of the development of ICH M13 guidance on Bioequivalence for Immediate-Release Solid Oral Dosage Forms (ICH, 2022a), expected to start in 2024.

### 2.1.2. *Two-stage, adaptive design*

Two-stage, adaptive designs in the context of BE trials are considered for situations in which reliable estimates of the intraindividual variability of drug products or for the actual difference between products are

not available. The overall concept of a two-stage trial is to adapt the final sample size based on variabilities determined within the same trial, which may be preferable compared to variability assessments in separate pilot studies – from a scientific perspective, as well as considering economical and ethical aspects. In general, such an approach could, among other benefits, allow to reduce sample sizes in comparison to the common combinations of pilot and pivotal trials.

*2.1.2.1. Regulatory recommendations.* Two-stage design approaches, comprising adaptive and group-sequential designs, are in general acceptable for regulatory agencies, at least for EMA, HC and US-FDA (EMA, 2010; Health Canada, 2018; US-FDA, 2019). With respect to the choice of the statistical method, the agencies appear to be flexible as long as the overall type-I error can be appropriately controlled (EMA, 2010; Health Canada, 2018; US-FDA, 2019). Group-sequential and adaptive studies have been used, of which several designs have been published, for example by (Fuglsang, 2014; Maurer et al., 2018; O'Brien and Fleming, 1979; Pocock, 1977; Potvin et al., 2008).

*2.1.2.2. Invited contributions and discussion*

*Knowledge on two-stage designs and novel approaches for evaluation and adjustment of type-I error rate.* A two-stage design provides greater flexibility for studies, thereby bringing ethical, administrative and economic benefits. Inconsiderate use of two-stage approaches may, however, inflate the consumer risk. It is common for all guidances, that appropriate control of the type-I error is expected, however, acceptable approaches for presentation of this feature are not universally defined.

The US-FDA presented survey results of BE studies with two-stage designs in generic drug submissions from January 1996 to December 2020: within this period, only 14 applications were submitted using the two-stage design approach, highlighting the limited use of this possibility. Potvin's method C was used in two-thirds (64%, $N = 9$) of studies, followed by group-sequential design (22%, $N = 3$) and Potvin's method B (7%, $N = 1$), besides one trial with a flawed design (7%). Overall, studies for 10 drug products were assessed as acceptable.

To better visualize the probability of type-I error inflation of Potvin's methods, the US-FDA developed a tool that should help applicants and regulatory agencies evaluating the type-I error inflation in adaptive-design studies. For the BE design space (point estimate, intra-subject variability and initial sample size in stage 1), the type-I error inflation was divided into the following regions: "tolerant zone" (type-I error $\leq$0.05), "negligible zone" (>0.05–0.052), and "unfavourable zone" (>0.052). The results demonstrate that the "unfavourable zone" covers a small area of the BE design space affiliated to low initial sample sizes of 20 or less and variabilities below 30% for all three methods. However, its size increases vastly as the target GMR deviates further from 1 (e.g., from 0.95 to 0.90). When evaluating the above-described BE studies of generic drug products, none of these landed in the "unfavourable zone" and only two studies landed in the "negligible zone" (Jiang et al., 2021).

Furthermore, some novel approaches in adaptive designs focusing on alpha adjustment and parallel-design studies were presented. In general, simulations were and are utilised to investigate the control of the type-I error for adaptive designs such as two-stage designs in BE. However, as a result of the development of methods for replicate, crossover designs based on inverse-normal combination such as Maurer's method (Maurer et al., 2018), simulations are no longer required. These "exact" methods control the type-I error in the strict sense, i.e., are based on analytical proof. They offer a high degree of flexibility: various futility criteria can be applied and minimum/maximum sample sizes for the second stage can be defined.

For studies with parallel design, no such "exact" methods are available at present, since inclusion of unequal group sizes in the first stage and heterogeneity make the development difficult and no solution has become available until now. Nevertheless, as for cross-over designs, software tools are freely available allowing for the required thorough investigation of the type-I error for applicable scenarios in a trial (e.g., Power2Stage).

*2.1.2.3. Conclusions and suggestions.* Although the two-stage adaptive design approach is described in regulatory guidelines, it is not used commonly. Unfortunately, the ambiguous description of two-stage design in the European guideline contributes to the weariness of sponsors with respect to this type of study in Europe.

Although the preferred approach for two-stage BE studies was not particularly discussed, the meeting participants were generally in favour of such studies. The participants agreed that two-stage designs should be used more frequently for BE assessment due to ethical and economic advantages such as an expected decrease in sample size, especially as there is a considerable chance of declaring BE in the first stage. Around two-thirds of participants of the survey (see **Supplemental Table**) also supported acceptance of simulations in discussion of two-stage designs. It should be underlined that using simulations is no trivial task due to the inherent limitation that it is not feasible to exhaust all possible combinations of point estimates, variability and sample sizes for a design.

### 2.2. Opportunities of "modelling and simulation" to support BE

BE assessment focuses on non-compartmental metrics such as $C_{max}$ and AUC. These PK parameters are influenced by many factors such as product differences, study design and study execution. Mathematical models either characterise PK exposures accounting for the physiologic perspective (physiology-based PK [PBPK] model) or are empirical approaches developed by data fitting (semi-mechanistic and population PK [PopPK] models). PBPK models are very complex, involving many system- and drug-related parameters, and can be applied, even when no *in-vivo* data are available, in a "bottom-up" approach to predict *in-vivo* PK responses. However, PBPK modelling techniques may also consider "top-down" or "middle-out" approaches by combining some data fitting from different origins (*in vivo, in vitro* or even *in silico*). This contrasts greatly with PopPK models that are basically built based on clinical data. The possibility of modelling and simulating different products or routes of administration will depend on whether the *in-vitro* data used as input can be assumed to be representative of *in-vivo* input processes (Tsamandouras et al., 2015).

For generic drug product development, opportunities for modelling and simulation approaches include supporting BE assessment to address changes in recruitment, large sample sizes or ethical concerns (e.g., patient studies, long study durations or studies with safety concerns), and as a basis for waivers of additional BE studies (e.g., fed conditions for IR products, coadministration with proton pump inhibitors, BCS waivers in certain situations). Furthermore, modelling approaches may be useful to circumvent unnecessary/insensitive clinical endpoint (CE) studies and to support *in-vitro* approaches for generic drug approval, e.g., for locally acting drug products, by taking advantage of virtual BE simulations.

In comparison, current regulatory applications of "modelling and simulation" for model-informed drug development (MIDD) for new drug products mainly include the evaluation of alternative dosing regimens or dose adjustments for specific populations to inform the product labelling (e.g., switch between administration routes, up-/down-titration, prolonged use, renal/hepatic impairment) and *in-vitro in-vivo* correlation (IVIVC) models to support definition of specifications and manufacturing variations.

However, US-FDA and other regulatory authorities are investing on building expertise in this field. The status-quo including concrete examples and future directions on how modelling and simulation approaches may support BE were discussed.

*2.2.1. Regulatory recommendations*

The experience of the EMA with "modelling and simulation" to

support BE (model-integrated evidence) is limited, predominantly due to lack of a central inventory of such cases that may be discussed on a national level. Challenges in BE trials, that may be a reason for consideration of a modelling approach, are addressed on a case-by-case basis within the European Union. Use of model-informed BE is in principle possible from the EMA's point of view, despite methodological challenges in their use are acknowledged and should be addressed in advance of the submission via scientific advice or qualification procedures (EMA, 2020). According to the impact on the regulatory decision, the EMA generally differentiates between "low", "medium" and "high impact" models, in that evidence of BE by a PK model replacing a trial clearly belongs to the "high impact" class (EMA, 2011). The EMA did not receive any submissions for approval based on model-based BE so far and therefore has not yet established a policy. Available guidelines focus on reporting of the results of such studies (EMA, 2007), IVIVC for modified-release formulations with a focus on innovator products/development of a modified-release product (EMA, 2014) or drug-drug interactions (EMA, 2012). The new EMA Methodology Working Party (MWP) will leverage cross-disciplinary expertise.

Investment of the US-FDA on building know-how on "modelling and simulation" is reflected by a few published PSGs implementing additional measures understood to be of importance for modelling and a multitude of research projects and grants, e.g., on modelling and simulation for long-acting injectables and locally acting drug products. The US-FDA's multidisciplinary group for Quantitative Methods and Modelling is dedicated to support regulatory policy through scientific research and maintenance of a modelling knowledge base. It further contributes to worldwide harmonisation of regulatory recommendations through scientific exchange by hosting routine generic drug and pharmacometrics cluster meetings. The ultimate goal is virtual BE, i.e., model-integrated evidence serving as pivotal information for generic drug approval (Zhao et al., 2019). However, so far, there are no established acceptance criteria for adequate model performance by the agency. In case of complex products without specific guidance, applicants are encouraged to seek scientific advice within the pre-ANDA programme (US-FDA, 2022b).

### 2.2.2. Invited contributions and discussion

Examples were presented that demonstrate the potential of using modelling approaches in facilitating generic drug development. Further examples underlined also that modelling can provide relevant insight into trial outcomes including support for complicated cases such as very high incidence of missing data and risk-based assessment of inconclusive outcomes of comparative trials or differences in product characteristics.

*2.2.2.1. Model-integrated evidence for BE: example diclofenac sodium topical gel, 1%.* A recent example on the utility of "modelling and simulation" in supporting generic drug development is the approval of a generic diclofenac sodium topical gel, 1%, for treatment of patients with osteoarthritis of the knee by the US-FDA in 2019 (Tsakalozou et al., 2021). For semisolid preparations of diclofenac, the recommended BE approach by the US-FDA usually comprises a PK BE trial in healthy individuals plus a comparative CE BE trial in patients (US-FDA, 2018a). In lieu of the comparative CE BE trial, the applicant developed a dermal PBPK model characterising the relationship between systemic exposure of diclofenac in plasma and local exposure in the skin and synovial fluid to demonstrate BE of the two drug products at the site of action. The applicant obtained the agency's feedback on the intended approach during the pre-ANDA scientific advice programme. Validation of the model using literature data from eleven dermatological products revealed that the model appropriately captured the local and systemic bioavailability for compounds with varying physicochemical properties and PK characteristics. After refinement of the model by the agency, the model was considered suitable to virtually determine drug exposure at the site of action, supporting the drug approval in lieu of a comparative CE BE trial. This was the first case of a generic drug approval by US-FDA, in which a PBPK model in combination with a PK BE trial supported the assessment of equivalence for a topical product and helped avoiding a comparative CE BE trial. Since comparative CE BE trials are considered as the least sensitive approach to detect formulation differences between a generic and a reference product, applicants are encouraged to use alternative approaches as described in this example (Tsakalozou et al., 2021).

*2.2.2.2. Modelling to reduce burden of conducting long clinical endpoint equivalence studies: example levonorgestrel intrauterine system.* Since intrauterine systems (IUS) for contraception such as the levonorgestrel (LNG)-containing IUS deliver the drug locally in the uterus, the site of action, a conventional PK-based BE approach is not sufficient for demonstration of equivalence. Further, the five-year application period of the IUS made the BE assessment of complete drug release challenging in that such a long trial was not feasible for generic development. Thus, US-FDA utilized a modelling and simulation approach to shorten the recommended *in-vivo/ex-vivo* clinical study from 5 years to 1 year. The simulation results revealed that having the 90% CI of the residual LNG amount after 1 year within the BE limit of 95.00–105.26% would ensure that the residual LNG amount after 5 years lies within the conventional BE range of 80.00–125.00% (Sharan et al., 2022). Consequently, according to the current PSG of LNG IUS, BE may be based on residual amount of LNG at month 12, besides comparative physicochemical and mechanical properties and comparable *in-vitro* drug release over 5 years (US-FDA, 2020b).

*2.2.2.3. Model-based assessment of relevant PK metrics for BE: example methylphenidate extended-release formulations.* For methylphenidate, the close relationship of plasma exposure and clinical effect allowed for the development of a mathematical model that predicted PK and PD changes in response to formulation variations for different extended-release products. The virtual BE simulation results revealed that partial AUC (pAUC) parameters are more sensitive to detect clinically relevant differences between two formulations than the conventional $C_{max}$ and AUC measures (Kimko et al., 2012). Consequently, the US-FDA incorporated the recommendation to additionally assess three defined pAUC metrics in the PSG of methylphenidate (US-FDA, 2018b).

*2.2.2.4. Application of modelling in the real world of setting dissolution specifications.* To illustrate the potential of modelling for waiving additional BE studies or justifying observed differences in dissolution profiles, its potential use in the setting of dissolution specifications for bioinequivalence risk assessment was presented. A common approach is the use of data from intravenous and/or absorption PK data as obtained from submissions for new drug applications and/or literature to validate a PBPK model. The PBPK model is subsequently used to define acceptable limits for dissolution profile difference and to support specifications. In general, predicting PK data from different formulations or batches can offer valuable data to gain model credibility. Whilst differences *in vitro* can be justified in case of equivalent performance *in vivo* (concept of side batches), occurrence of differences *in vivo* may involve implementation of model-based approaches such as IVIVC and/or PBPK. IVIVC has been well established (regulatory frameworks available, including validation), but PBPK may provide more opportunities (also extrapolation outside the available data).

### 2.2.3. Conclusions and suggestions

"Modelling and simulation" is considered a promising and evolving approach within BE development programmes, and regulators are building up their expertise in this field. At present, regulatory agencies have not established acceptance criteria for model qualification, which likely would facilitate development and use of such approaches. The majority of meeting participants (75%, see **Supplemental Table**) were

in favour of encouraging model-based BE decisions for complex generic drug products. Generic applicants are encouraged to propose virtual BE approaches and to discuss these with the regulatory agencies via scientific advice. Modelling approaches are indispensable in modernising generic drug development.

In November 2022, a new ICH M15 guideline on Model-Informed Drug Development General Principles was endorsed by the ICH Management Committee (*post-meeting note*). This new guideline is expected to harmonise documentation standards, model development, data used in the analysis, model assessment and its applications (ICH, 2022b).

### 2.3. Fed versus fasting conditions in BE trials: current status and new insights

Should BE assessment be requested after both, fasted and fed administration in case of IR oral dosage forms? This question has already been investigated and discussed in previous GBHI conferences (Blume et al., 2021; Chen et al., 2018). However, some essential aspects remained open and are still waiting for further harmonisation based on experimental evidence, such as (Blume et al., 2021):

- comparability and deviations in gastrointestinal (GI) performance between different oral dosage forms, e.g., tablets vs. capsules or orodispersible/orally disintegrating tablets,
- pharmacokinetic characteristics of compounds critical for (or levelling out) food effects initiated by changes in disintegration/dissolution,
- physiological differences between fasted and fed state: impact of mixing with meal in the stomach, changes in gastric emptying, gastric residence and physical stress.

It was the intention of this session to share experimental evidence relevant in this context and to discuss options for further harmonisation of the existing regulations.

#### 2.3.1. Regulatory recommendations and where do we stand after GBHI-1 and GBHI-4?

While in most cases EMA requests BE assessment only under one administration condition (normally fasting), US-FDA more generally recommends BE studies in both, fasted and fed state (EMA, 2010; US-FDA, 2021). According to the European guidelines, BE studies with IR preparations should be conducted considering the administration conditions suggested in the labelling of the reference product. Only in case of medicinal products with "specific formulation characteristics", such as microemulsions or solid dispersions, studies in fasted as well as in fed state are expected (EMA, 2010). This approach has been accepted by a number of other regulatory agencies, e.g., Australia, Brazil, Canada, Japan, Malaysia, Russia, and Thailand. Moreover, EMA and US-FDA developed PSGs for those drug products for which BE study conditions are suggested that deviate from the normal case.

In general, studies under fasting conditions were found more sensitive to detect formulation differences than studies under fed conditions (DeRosa, 2019). However, experimental data suggest that certain excipients may affect disintegration of the drug products differently in fed and fasted state. Hence, such excipient-food interactions were identified as potential reason for deviating BE outcome under fasting vs. fed conditions. Retarded disintegration (in fed state) may result in delayed onset of absorption and/or reduced drug concentrations at the absorption site and thus, deviating peak and/or total exposure (Blume et al., 2021).

A survey conducted by US-FDA had elucidated that drug products containing BCS class IV drugs are obviously more sensitive for a different outcome of BE studies under fasting and fed conditions. This finding is in line with earlier conclusions that highly soluble drug compounds (BCS class I and III) are less sensitive to food effects (Blume et al., 2021; DeRosa, 2019). However, uncertainty remained, whether such sensitivity is only related to the drug itself or is also dependent on the formulation properties of the drug product.

Main goals for the discussion in this session were therefore to identify formulation factors and characteristics of excipients which may impact the rate and extent of drug absorption differently in fed and fasted state. Moreover, it was discussed which GI physiology aspects with a possible impact on the *in-vivo* drug release performance of IR preparations might be identified. Finally, the PK properties of the drug compounds should be critically examined how far they make drug products more or less sensitive to changes in disintegration and/or dissolution in the GI tract.

#### 2.3.2. Invited contributions and discussion

*2.3.2.1. Changes in GI physiology after fasted and fed administration of IR drug products.* Food intake leads to relevant changes of the physiological conditions in the GI tract that may impact the systemic exposure and PK profile of a drug compound. Critical parameters include GI transit times, availability/volume, composition, pH, viscosity, surface tension of luminal fluids, as well as hydrodynamics and gastric/intestinal peristalsis. Relevance of these parameters for the systemic exposure depends on the properties of the drug compound and the drug formulation, e.g., the availability and composition of luminal fluids will be significant, if solubility is the limiting factor for drug absorption.

When administering IR drug products, the conditions in the stomach are of primary importance. In this context, it was discussed how long it normally will take to achieve a fasted state again after ingestion of a meal. The common advice for fasted intake to administer a drug product 1 h before or 2 h after a meal will not suffice – neither after a standard high-fat, high-calorie meal nor after a moderate meal of 400–500 kcal. After ingestion of heavy meals, six hours or more are required to achieve fasted conditions again. Even in case of fasted administration of drug products together with water, certain changes in gastric physiology will be initiated, i.e., increase of volume and pH while temperature will be reduced. However, these changes will disappear within about 30 min (Koziolek et al., 2019).

Secretion of gastric juice and composition of GI fluids are largely affected by the type of meals taken, i.e., its content of fat, carbohydrates, and texture, as well as the co-administered liquids. In this context, it is important to consider that water does not mix well with the viscous chyme inside the stomach, but that water may be transported along the stomach wall and emptied more rapidly in the small intestine, a phenomenon sometimes named "stomach road".

Consequently, if drug products deviate in disintegration or dissolution in the fed stomach, the more rapidly dissolved drug compound could be emptied together with the co-swallowed water via the "stomach road" earlier and thus, may be absorbed more quickly than the drug entrapped in the chyme. Hence, the "stomach road" effect is a main contributor to variability in the onset of drug concentrations occurring in plasma after administration of drug products in the fed state – with differences depending on the initial localisation of the dosage form within the stomach (Koziolek et al., 2016; Schick et al., 2019).

The practical relevance of the "stomach road" effect was demonstrated in a study comparing the PK properties of a fast disintegrating and dissolving tablet (FDDT) containing acetylsalicylic acid (ASA) with those of the regular ASA tablet (RT). After fed administration of both products to 30 healthy subjects, time to peak concentration ($t_{max}$) was significantly shorter and $C_{max}$ significantly higher in case of the FDDT compared to RT, whereas total exposure (AUC) was similar in both cases. Moreover, FDDT was also investigated in fasted state (RT only fed), and under these conditions an increased $C_{max}$ was observed while $t_{max}$ was only negligibly shorter. *In-vitro* data obtained by use of a specific biorelevant dissolution testing equipment (GastroDuo) indicated a faster drug release and improved gastric emptying for the FDDT, whereas for the RT incomplete gastric emptying was suggested. Thus, the earlier $t_{max}$ of the FDDT under fed conditions compared to the RT is

presumably related to the "stomach road" effect that was enabled by the very rapid release of ASA from the FDDT (Schick et al., 2020).

Luminal pH is a further physiological parameter impacting the systemic exposure especially in case of drug compounds with pH-dependent dissolution properties. Even though pH values measured in fasted or fed state cover a similar range, more time is available for drug dissolution after fed administration due to prolonged gastric residence time and intestinal transit. Thus, intake of weakly basic compounds after a meal may result in higher oral bioavailability. In this context, it should also be considered that the composition of the intestinal fluids will change between fasted and fed state, in particular, concentrations of bile salts, phospholipids and cholesterol will be increased after meal ingestion with potential impact on the solubility of the administered drug compounds (Koziolek et al., 2019).

*2.3.2.2. Different performance of oral IR drug products in fasted and fed state.* In general, all described changes in GI physiology are more important for drug products with modified-release characteristics than for IR drug products. However, such changes may also initiate differences in disintegration and drug release of IR solid oral dosage forms and, consequently, impact their bioavailability differently between two drug products.

This had been demonstrated in a study with four caffeine-containing IR solid oral dosage forms (one film-coated tablet and three hard capsules) which all released the drug rapidly *in vitro* with almost 100% dissolved within 15 min under compendial conditions. *In vivo*, however, initial disintegration and onset of absorption (determined as occurrence of caffeine in saliva) was found to be different between the drug products, with more pronounced deviations in fed state. In this context, conditions differing in fasted and fed state, i.e., temperature, hydrodynamics and mechanical stress in the GI tract, are obviously important factors impacting the disintegration (and dissolution) of the products (Sager et al., 2019).

It was concluded that fasted and fed conditions may impact disintegration and/or release from certain solid oral dosage forms (e.g., tablets vs. capsules) differently. As such deviations may affect bioavailability of the drug, BE demonstrated after fasted administration might not be extrapolated to the fed state and vice versa.

*2.3.2.3. Predictability of in-vitro tests for BE in fasted versus fed state.* Relevance of comparative *in-vitro* dissolution tests to predict BE in fasted state is well established considering the BCS-based biowaiver concept (ICH, 2021). However, as discussed in the last GBHI conference in 2019 (Blume et al., 2021), this cannot always guarantee BE between two drug products also in the fed state, especially as long as excipients with gelling properties (e.g., croscarmellose sodium) or swellable fillers (e.g., microcrystalline cellulose) are contained in only one of the drug products (Zaheer and Langguth, 2018, 2019). Delayed disintegration of tablets has also been observed *in vivo* after ingestion of liquid meals due to precipitation of a protein film on the surface of tablets (Abrahamsson et al., 2004), however this phenomenon should not be different between different tablets under the same study conditions.

Additional comprehensive *in-vitro* comparisons have been proposed, e.g., by use of the bio-relevant GI transfer (BioGIT) model, a three-compartment *in-vitro* apparatus that simulates drug transfer from the stomach through the upper small intestine. This model could be used for assessment of the impact of dose and formulation on early exposure to drug compounds with low solubility (Kourentas et al., 2018).

Due to the large complexity and variability of physiological parameters after food intake, *in-vitro* methodologies to evaluate intragastric disintegration/dissolution, the GI transfer process and drug dissolution in the upper small intestine need to be developed further. Moreover, in-silico modelling approaches may be promising for predicting fed/fasted comparisons for BE assessment (Wagner et al., 2021).

*2.3.2.4. New food trends: possible impact on GI function.* Even though not primarily relevant for BE assessment (where standardised meals are used in all study periods) but more related for defining best administration conditions for drug products, an interesting additional aspect of the discussion was focussed on the question how different types of food may influence GI function and thereby may impact the *in-vivo* performance of orally administered drug products.

Food processing and additives may exhibit a profound influence on the GI environment and may, thus, be critical for the *in-vivo* performance of drug products. The consumption of ultra-processed food – containing lots of additives such as emulsifiers (surfactants), sugars, preservatives, stabilisers – is increasing (Monteiro et al., 2011) and was associated with the occurrence of irritable bowel syndrome in a large cohort study involving 33,342 participants (Schnabel et al., 2018). This finding is not surprising, since many food additives, e.g., sugars, salt, emulsifiers and organic solvents, induce or are associated with an increased intestinal permeability (Lerner and Matthias, 2015).

This is only one example that highlights the impact of changing dietary habits on the physiology of the GI tract, which may eventually lead to altered *in-vivo* performance of orally administered drug products. Whether this may induce also formulation-related differences in food interactions has not been systematically investigated so far and may necessitate future scientific investigations.

*2.3.3. Conclusions and suggestions*

The eligibility and relevance of the question whether assessment of BE in fasted and in fed state should be recommended for all IR drug products was confirmed by a survey of generic IR product development programmes with 82 drug products in 204 applications: in only 8.3% of all cases, BE shown under fasting conditions could not be confirmed after fed administration. The majority of the failed cases (8 products in 17 applications) concerned compounds of BCS class IV (Blume et al., 2021; DeRosa, 2019). It should be noted that not confirming BE does not automatically mean that the drug products are bio**in**equivalent. Hence, convincing scientific arguments might be needed to support the regulatory concept requesting BE studies in fasted and fed state for all IR drug products.

The discussion revealed that a decision tree for the necessity of fed and fasted studies should be developed. In this context, the applicant would need to justify any omission of a fed study considering the risk(s) involved, e.g., due to the presence of potentially interacting excipients such as surfactants or pH modifiers, but also general properties of the dosage form. It seems recommendable that drug products should not deviate qualitatively in these excipients and should exhibit quantitatively (very) similar composition.

In the conference survey, half of the respondents reported that they had experienced cases of drug products that were bioequivalent under fasting but not under fed conditions. Amongst them, half declared that they knew cases of capsules being not bioequivalent under fed conditions only. Also, in this survey, it remained unclear whether the drug products being not bioequivalent were indeed bio**in**equivalent or just that BE was not proven (e.g., due to inadequate sample size).

Common expectation is that only the situation in the stomach is essential for the *in-vivo* performance of IR solid oral dosage forms considering that disintegration of the drug product and drug release are terminated before gastric emptying. After fed administration, however, this may not be the case, thus leading to potentially different behaviour of different drug products, e.g., hard capsules and tablets. Given these assumptions, BE conclusions from fasted studies might not be extrapolatable to the situation after administration in fed state. Consequently, additional BE studies after fed administration may be necessary in case of such deviating oral dosage forms.

As far as the biopharmaceutical and pharmacokinetic properties of the drug substances are concerned, poor solubility seems critical and in particular BCS class IV compounds should be of concern. Whether (very) slowly absorbable compounds may be less critical for such differences

between fasted and fed administration needs further elucidation. Moreover, pH-dependent solubility may be critical, specifically in case of different salts with a different pH dependency of solubility or the use of different pH-modifying excipients.

The discussions and outcome of this and earlier GBHI conferences, which also covered the topic of BE of oral IR drug products under fasting and/or fed conditions, may be taken into consideration when reviewing the draft ICH M13A Guideline which has been published shortly after the GBHI conference (ICH, 2022a).

Finally, it was suggested during the conference that regulatory authorities should develop PSGs to define individual approaches for specific cases that do not fit into the overall regulatory recommendations.

### 2.4. BE assessment of locally acting drug products applied topically to the skin

Locally acting topical drug products, allowing for targeted application of a drug compound to a specific area of skin, encompass a variety of dosage forms ranging from simple solutions to complex formulations and delivery systems. Unlike drug products that are indicated for systemic action, systemic PK studies are generally not relied on for establishing equivalence of these locally acting products, since systemic drug exposure may often not reflect the exposure at the site of action. Thus, historically, comparative CE BE studies have been used in general to demonstrate equivalence of topical products, despite their disadvantages of high costs and effort due to large sample sizes and low sensitivity for potential product-related differences (Lionberger, 2008; Miranda et al., 2018a; Novakovic et al., 2019). This negatively impacts the availability of topical generics.

Thus, the development of suitable pharmacodynamic (PD) or local PK-based methods for equivalence assessment is highly warranted. With the exception of the vasoconstrictor assay for topical corticosteroid preparations, previous approaches to the validation of PD endpoints were not very effective. Instead, alternative *in-vitro* approaches and *in-vivo* methods investigating local PK in the skin are currently moving to the forefront of scientific activities and are expected to be implemented step-by-step into regulatory recommendations for equivalence assessment of locally acting topicals (Miranda et al., 2018a, 2018b).

However, the broad variety of diseases, divergent sites of actions due to a broad spectrum of disease locations within the different layers of the skin, disease- and time-related changes of the application site itself, and a broad variety of formulation properties contribute to the difficulties of developing and validating alternative methods for product characterisation.

In this context, a complete waiver of any type of *in-vivo* studies based on *in-vitro*-release- and *in-vitro*-permeation testing (IVRT/IVPT) was discussed elaborating the challenges of adequate validation of these methods. Furthermore, promising methods characterising local PK after topical application including Raman spectroscopy and dermal open flow microperfusion (dOFM) were evaluated for their suitability for equivalence assessment.

#### 2.4.1. Regulatory recommendations

The regulatory draft recommendations of the EMA and the US-FDA regarding equivalence assessment of locally acting topical products are in principle comparable. The “gold standard” method for assessing equivalence of topical generic products still relies on comparative CE BE studies (Miranda et al., 2018b).

For simple formulations such as solutions, suspensions or single-phase gels, equivalence with respect to quality may be sufficient, and an *in-vivo* BE study may thus be waived (EMA, 2018; US-FDA, 2011, 2020a).

A sole, and so far, unique example for codified use of PD endpoints for equivalence assessment of topical corticosteroids is the so-called vasoconstriction assay. This is a PD assay in healthy human subjects, relying on the efficacy-correlated vasoconstrictive properties of corticosteroids and the resulting skin blanching effect and is accepted by several international regulatory agencies including the EMA and the US-FDA (EMA, 2018; US-FDA, 2022f).

Based on a review of all PSGs for locally acting topical dermatological products issued by US-FDA as of August 2022, the following BE recommendations are made: PD or comparative CE BE study (56%), either *in-vitro* or PD/CE BE study (23%), either waiver or PD/CE BE study (16%), and PK-based/other (5%). Although PD or CE BE studies are requested for demonstration of therapeutic equivalence in most cases, this demonstrates the growing acceptance by US-FDA of alternative characterisation-based BE approaches in lieu of CE BE studies.

For Europe, no such clear picture on the extent of acceptance of alternative approaches of BE assessment including *in-vitro* methods is available. Nevertheless, multiple cases exist in which equivalence was as least partially demonstrated by using *in-vitro* assessments such as IVPT, indicating that acceptance for the *in-vitro* BE approach is also increasing in Europe (Miranda et al., 2018b). In 2018, a draft guideline on the Quality and Equivalence of locally applied and locally acting topical products was released by the EMA for consultation and feedback from relevant stakeholders. In principle, the guideline considers equivalence in quality alone not sufficient, but requires comparative permeation kinetics and PD studies to conclude on therapeutic equivalence. However, the required approach should also consider complexity of the formulation, where for simple formulations also acceptance based on equivalence with respect to quality could be considered. This guidance should be used to develop and justify topical product-specific BE protocols, and describes how *in-vitro* (e.g., IVPT) and *in-vivo* models (e.g., tape stripping, PK studies, PD studies including vasoconstriction assay for corticosteroids) may substitute for CE studies. According to this draft, these alternative methods may only support the claim of therapeutic equivalence for drug products with the same application mode and when the risks of inequivalence to the patient are minimal (no narrow therapeutic index drug, no systemic toxicity that cannot be considered equivalent in conventional BE, no unknown path to the side of action) (EMA, 2018). The meeting revealed that obviously significant changes are to be expected for the next guideline version to be published.

Whereas the majority of regulatory agencies, including those of Europe, USA, Canada and Australia, in principle demand the demonstration of therapeutic equivalence for topical generics, the regulatory agency of Brazil, ANVISA, strongly relies on *in-vitro* information for topical generic products (for drugs without a systemic effect, same excipients in the same quantities and same physical/chemical/microstructural and microbiological parameters) (Miranda et al., 2018b). Only one exception from the biowaiver exists, that was introduced into their new guidance very recently: BE of topical corticosteroids has to be assessed *in-vivo* using the vasoconstriction assay (ANVISA, 2022).

#### 2.4.2. Invited contributions and discussion

*2.4.2.1. Characterisation-based approach as a valuable alternative to clinical data for establishing equivalence of topical products.* In recent years, research initiated by US-FDA has led to advances in characterisation-based *in-vitro* methods. The modular characterisation-based BE approach of the US-FDA generally encompasses qualitative (Q1) and quantitative (Q2) sameness of inactive ingredients, physicochemical/structural (Q3) similarity (US-FDA, 2022e), and IVRT (US-FDA, 2022d). Depending on the complexity of the formulation, IVPT or another bio-relevant performance test (US-FDA, 2022c), and for a few products, *in-vivo* systemic PK studies may be additionally requested for establishing BE (e.g., PSG doxepin cream US-FDA, 2022a).

As a representative example for a complex semisolid topical product, the development of the characterisation-based BE approach for topical acyclovir cream was presented (FDA Award U01-FD005223): the characterisation of different reference products approved in different

countries revealed that rheological and IVRT data were very sensitive to formulation changes. Furthermore, IVPT data suggested that bioavailability (based on comparison of flux values from the IVPT comparison) is correlated with physicochemical characteristics/microstructure (Q3). These results led to the current draft PSG in which the characterisation-based approach including Q1/Q2 sameness, Q3 similarity, IVRT and IVPT has been implemented in option I for BE evaluation (*in-vitro* option) as an alternative to option II (comparative CE BE study) (US-FDA, 2016). Comparable outcomes, however, based on smaller data basis, were also presented for products containing lidocaine and prilocaine.

*2.4.2.2. In-vitro approaches for equivalence assessment: IVRT, IVPT.* Test protocols are provided in the EMA draft guideline for IVRT and IVPT (EMA, 2018).

IVRT evaluates the rate and extent of drug release from a product in a diffusion cell apparatus using an inert membrane that is not rate-limiting to drug release. This test does not model *in-vivo* performance, but release rate represents a critical quality attribute of a drug product. Conditions for the IVRT test as defined by the EMA and according to USP Chapter 1724 differ considerably, e.g., with respect to amount of sample in donor chamber, temperature, amount of drug compound to be released, and equivalence acceptance limits, i.e., 90% CI ratio of test and reference slopes that should lie within 90–111% according to EMA and within 75–133.3% according to USP.

It was noted that the expected fraction of drug compound released during the test and the CI of 90–111% need to be reviewed in the context of the underlying theoretical assumptions taking into account topical generic products that have already been approved as well as the corresponding topical reference products. Since semisolid dosage forms exhibit intrinsic variability, it was proposed that more reasonable equivalence criteria need to be applied to generic products (Miranda et al., 2020; Pleguezuelos-Villa et al., 2019).

For IVPT, a diffusion cell apparatus with *ex-vivo* adult human skin as a membrane is used to measure the rate and extent of permeation of drug compound through the skin. Also, for IVPT, differences exist between the recommendations of the EMA and US-FDA (draft acyclovir guidance), e.g., with respect to duration, amount of sample in donor chamber, sample size and equivalence acceptance criteria. The EMA defines a conventional BE limit of 80–125% with the possibility of widening to a maximum of 69.84–143.19% in case of a high WSV observed with low-strength and limited-diffusion drug products and if clinically justified. According to US-FDA, the within-reference variability is used to determine, if average BE (limit 80.00–125.00%) or scaled average BE analysis is applied (EMA, 2018; US-FDA, 2016).

A particular area of concern is the inherent variability in human skin and the challenges that this variability is likely to pose when progressing IVPT for BE testing. In the largest analysis of IVPT data to date, 2400 skin samples excised from 112 female donors were retrospectively analysed. In the study, inter-individual variation amounted to 37.6% but was not as high as intra-individual variation ($38.3\% \leq CV \leq 115.7\%$) (Meidan and Roper, 2008). In this context, the required number of skin donors and number of replicates per skin sample was discussed and has to be considered in light of the difficulties in sourcing human tissues in different regions. Whereas the EMA states that the number of skin samples should not be less than 12 with at least 2 replicates per donor (EMA, 2018), in the IVPT assay described in the draft acyclovir guidance, the number of donors is not specified, but at least 4 replicates per donor should be assessed (US-FDA, 2016). Other guidelines mentioning dermal absorption studies state that 8 replicates from at least 4 donors should be studied (OECD, 2019).

It was pointed out that further discussion may be needed with regard to the applicable acceptance limits and the statistical approaches. In addition, discussion may be needed regarding the part of the skin to be used, as this is not specified in the guidelines, although large differences with respect to skin characteristics such as thickness of epidermis at different parts of the body exist. This may contribute to the large variability seen in IVPT results. To overcome the issue of limited sample sizes, mainly caused by limited availability of adequate tissue, extrapolation of the data to larger virtual sample sizes using modelling approaches (see Section 2.2) was suggested.

To date, only few applicants used IVPT as part of their BE programme. Overall, there was consensus amongst the participants that the current proposed equivalence criteria for *in-vitro* tests, adopted from the criteria applied for oral products, are arbitrary and not scientifically justified, highlighting the need for more data exchange amongst scientists, industry and regulatory agencies. Furthermore, the participants stressed that harmonisation of *in-vitro* tests is highly warranted.

*2.4.2.3. Skin-PK-based methods: Raman spectroscopy and open flow microperfusion as promising BE approaches.* Assessing the equivalence of locally acting topical drug products *in vivo* is challenging because of technical difficulties associated with measuring and comparing drug concentrations in the skin. However, approaches characterising PK within the target organ skin may be the most accurate and sensitive way to demonstrate cutaneous BE. The currently discussed methods differ with regard to the skin layer in which the quantification of the drug compound is performed.

For example, the minimally invasive tape stripping with adhesive tape is solely suited for assessment of drug concentration in the stratum corneum (Escobar-Chávez et al., 2008). However, although included in the EMA guideline (EMA, 2018), this method is still not fully accepted due to the difficulties in demonstrating reproducibility; thus, in case this method shall be regularly applied, more research is needed to overcome the hurdles (Miranda et al., 2018a).

Confocal Raman spectroscopy is a non-invasive method for assessment of drug concentrations in the upper epidermis. This method is based on inelastic light scattering after illumination of the sample by a monochromatic laser. The resulting Raman spectrum entails information about the drug product's molecular composition, structure and interactions. Prerequisite is thus a polarisable molecule that is Raman-active. The evaluation needs to consider that no absolute concentration can be measured (Caspers et al., 2001; Miranda et al., 2018a). Results from several research projects initiated by US-FDA, at the moment with excised skin samples from donors, demonstrated that Raman spectroscopy represents a promising skin-PK-based technique for BE assessment: e.g., for different reference products of topical acyclovir cream approved in different countries, the congruence of Raman spectroscopy data with IVPT results could be demonstrated (FDA Award U01-FD005226). Overall, preliminary *in-vitro* data with multiple molecules suggest that comparison of cutaneous PK is feasible using the technique. As this method is fully non-invasive, the application is considered desirable but the future scope of work needs to focus on method validation strategies.

DOFM is an advanced invasive continuous dermal sampling technology, in which the probe is placed in the dermis and continuously perfused by dermal interstitial fluid which is collected over a defined period of up to 48 h. In contrast to the microdialysis technique, where the sample and the dermal fluid are separated by a semi-permeable membrane, in dOFM, the sample is in direct contact with the interstitial fluid (Anderson et al., 1991; Bodenlenz et al., 2017). Such continuous dermal sampling methods for assessment of drug concentrations in the dermis enable a head-to-head comparison of a topical test product to its reference in the same subject.

A variety of drug compounds ranging from small lipophilic molecules to large antibodies can be monitored with dOFM in the dermis of healthy subjects or patients (Bodenlenz et al., 2016, 2012, 2017; Dragatin et al., 2016). To limit variability in results, rigorous standardisation of clinical trials using dOFM is crucial, e.g., with respect to insertion of probes, minimisation of trauma formation, dosage application, probe depth and flow rate.

Suitability of this method for BE assessment was demonstrated by F. Sinner and colleagues in multiple clinical studies that were conducted in cooperation with US-FDA over nearly 10 years up to now: for example, in a study in 20 healthy subjects with two topical acyclovir cream preparations that were known to be non-equivalent, dOFM sampling was performed over 36 h with 12 probes per subject. The study failed to show BE for the two different preparations but showed BE of the drug product to itself, demonstrating sensitivity of the method for the hydrophilic compound acyclovir (Bodenlenz et al., 2017). Likewise, in a similarly designed dOFM study in 20 healthy subjects with a topical lidocaine/prilocaine cream and an approved generic version, BE could not only be demonstrated for the product itself but also to the marketed generic product. Thus, the results indicate that the method might be also reliable for assessment of BE of drug products containing moderately lipophilic compounds such as lidocaine and prilocaine (Tiffner et al., 2020).

In summary, dermal continuous sampling techniques such as dOFM have advanced over the last decade. They have demonstrated adequate sensitivity and robustness to evaluate BE of two drug products for different classes of topical drugs in multiple clinical studies using a reasonable number of subjects. One of the main challenges for implementation as a regulatory method might be the burden for the subjects and the need for conducting a pilot study to verify pivotal study parameters including dose applied, application time and concentration range. As a matter of fact, the highly sophisticated practical requirements so far limit the application of such methods to few highly specialised laboratories and so far prevent widespread use, which would be a requirement for adoption as a standard procedure worldwide.

#### 2.4.3. Conclusions and suggestions

To promote the availability of high-quality and affordable generic versions of topical products, robust BE approaches without the necessity of time- and resource-intensive comparative CE BE studies are needed. There was consensus on the lower discriminative power of comparative CE equivalence studies compared with *in-vitro* or PK BE studies. However, it was concluded that *in-vitro* studies may be overdiscriminating in some cases and that the detected differences might not be clinically relevant. In the survey, more than half of the respondents knew of generic drug applications asking for a waiver of a CE BE study: in half of these cases the applicant was successful and in the other half, a CE BE study was demanded in the end.

Several alternative methods have been reported that may allow for more straight-forward, cost-effective and robust BE assessment. There was consensus that the *in-vitro* tests IVRT and IVPT are commonly accepted by regulatory agencies. From a technical aspect, the recommendations for validation and the criteria allowing for a waiver of any *in-vivo* study need to be harmonised.

The discussion further showed that so far amongst the PD methods the vasoconstriction assay for topical corticosteroid preparations is a fully accepted one, but no other candidate for PD assessment was elucidated.

Promising *in-vivo* BE methods focus on assessment of PK within the target organ. It became obvious that Raman spectroscopy has to be further optimised before being regularly implemented. On the other hand, validation of dOFM has made significant progress, and the potential of detecting product-related differences has been demonstrated by a series of studies. However, more effort may be needed with respect to laboratory transfer to allow for common application.

A modular approach where drug compound/drug product properties, results from *in-vitro* techniques and, depending on the compound and indication, additional *in-vivo* data may be requested, is currently realised in the PSGs of US-FDA. This approach could be considered adequate against the broad spectrum of diseases and their localisation within the target organ skin.

Overall, the considerably different regulatory criteria for *in-vitro* tests and arbitrary BE criteria will have to be overcome by more intensive data exchange amongst scientists, industry and regulatory agencies, thereby fostering scientifically based decisions.

### 2.5. BE evaluation of narrow therapeutic index drugs

No globally applicable regulatory definition for narrow therapeutic index (NTI) drugs is available, and case-by-case decisions are usually required, in several cases by means of PSGs. In general, the ratio between the therapeutic/effective dose and the toxic dose (TD50/ED50) of these compounds is small. Consequently, for NTI drugs, sometimes also referred to as “critical dose drugs”, small differences in doses and consequently plasma concentrations can result in an insufficient therapeutic response or onset of adverse effects, and individual titration of dose and often also therapeutic drug monitoring are needed. Typical examples of NTI drugs include vitamin K antagonists, digoxin/digitoxin, and some antiepileptic drugs (Gozzo et al., 2022).

To face the potential ineffectiveness and adverse effects related to differences in plasma concentrations, many regulatory agencies defined stricter BE criteria for products with NTI drugs than the conventional BE limit of 80.00–125.00% for the 90% CI of the test-to-reference GMR. However, the narrower acceptance ranges as well as the PK parameters to which they need to be applied are not harmonised.

The regulatory recommendations for BE assessment of products containing NTI drugs in Europe, the USA and Japan were presented and possible alternative approaches for statistical evaluation of BE proposed. Furthermore, the challenges for development of affordable generic NTI drug products from an industry’s perspective were discussed.

#### 2.5.1. Regulatory recommendations

The BE criteria applied by US-FDA, EMA and the Japanese regulatory authority PMDA are summarised in Table 2.

According to the EMA, criteria for considering a drug as an NTI drug have not been properly defined, and the decision must be made case by case based on clinical considerations. As the European regulations allow for different procedures for market authorisation, in particular national procedures are possible, assessment of a drug product may even differ between member states. For some drug products such as those containing levothyroxine or ciclosporin, PSGs are available (EMA, 2022).

**Table 2**
Comparison of bioequivalence criteria for narrow therapeutic index drug products.

| | US-FDA | EMA | PMDA |
|---|---|---|---|
| Study design | Fully replicate, 4-way crossover | 2-way crossover | 2-way crossover |
| Approval criteria for AUC | RSABE plus unscaled average BE limit (80.00–125.00%) | 90.00–111.11% | 80.00–125.00% |
| Approval criteria for $C_{max}$ | plus<br>variability comparison: upper limit of 90% CI of the ratio of the within-subject SD of test to reference $\leq$2.5 | 80.00–125.00%<br>If particular importance of $C_{max}$: 90.00–111.11% | 80.00–125.00% |
| PEC for AUC and $C_{max}$ | No | No | GMR contained in [0.9, 1.11] |

AUC, area under the curve; BE, bioequivalence; $C_{max}$, maximum concentration; EMA, European Medicines Agency; GMR, geometric mean ratio; PEC, point estimate constraint; PMDA, Pharmaceuticals and Medical Devices Agency; RSABE, reference-scaled average bioequivalence; SD, standard deviation; US-FDA, US-Food and Drug Administration (EMA, 2010; PMDA, 2020b; US-FDA, 2021).

The BE acceptance interval for AUC should be tightened to 90.00–111.11% and, if $C_{max}$ is of particular importance, e.g., for safety or efficacy, the tightened acceptance limit shall also apply for this parameter (EMA, 2010).

The US-FDA has a cross-disciplinary working group to classify NTI drug products based on considerations as delineated in Yu et al. (2015). The BE determinations of NTI drug products are reflected in their respective PSGs. In general, the US-FDA recommends using a scaling approach with a four-way, fully replicate, crossover design study in which the WSVs of the test and reference product can be computed. In most cases, NTI drug products are characterised by low to medium WSV. The BE limits for both parameters, AUC and $C_{max}$, are scaled based on the determined WSV of the reference product (RSABE, see also Section 2.1.1.1). The generic product should pass both the reference-scaled limits and the unscaled average BE limits of 80.00–125.00% (primary BE criterion). By applying the unscaled BE criterion, it is defined that the scaling effectively stops at a WSV of 21.42% with the maximum acceptance range of 80.00–125.00%. Furthermore, the WSVs of test and reference product should be comparable, i.e., the upper limit of the 90% CI of the ratio of the within-subject standard deviation of test to reference product should be $\leq$2.5 (secondary BE criterion) (US-FDA, 2021; Yu et al., 2015).

In Japan, a list of NTI drugs is provided within the PMDA guideline for BE studies for different strengths of oral solid dosage forms (PMDA, 2020a). The requirements for BE of NTI drug products that apply to both AUC and $C_{max}$, although not explicitly stated in the guideline (PMDA, 2020b), are derived as a combination of the two possible BE criteria applicable to IR products as follows: the 90% CIs of the test-to-reference GMR should lie within the conventional BE acceptance range of 80.00–125.00%. Additionally, a PEC is defined, i.e., the estimate of the test-to-reference GMR should lie within the range of 90.00–111.11%. However, applicants are strongly encouraged to use the face-to-face consultancy service of the agency to discuss their intended BE approach.

### 2.5.2. *Invited contributions and discussion*

*2.5.2.1. Current BE criteria for NTI drugs.* As a result of earlier discussions by an advisory committee, US-FDA developed and introduced the concept of tightening the acceptance range based on the observed WSV of the reference product and the BE criterion for variability comparison (Yu et al., 2015). The variability comparison was considered relevant due to clinical concerns that an increased WSV of the test compared to the reference product and thus the larger variation in systemic exposure may increase the likelihood of therapeutic failures and/or adverse reactions. Simulation results also demonstrated that the primary criterion of scaled BE limits plus capping at 80.00–125.00% alone is insufficient to fail BE studies with large WSV differences ($\sigma_{WT}/\sigma_{WR} = 2$) between test and reference, if the GMR is close to 1 (Jiang et al., 2015). The established RSABE criteria (regulatory constant $\sigma_{W0}$ of 0.10 and $\Delta$ of "1/0.9″) imply acceptance limits of 90.00–111.11% at a reference WSV of 10%, that are comparable to common tightened acceptance ranges of other authorities. The implemented approach of scaling and comparison of variabilities involves a fully replicate design. This was considered additionally beneficial as fixed tightened limits were discussed as being too conservative in case of WSVs larger than 10%, resulting in high sample sizes or limited power. An additional criterion of the CI including 100% was rejected, as power would unreasonably be compromised. That is, in case of negligible deviations of the point estimate from 100%, higher sample sizes would decrease power and counterintuitively result in a penalty for the trial.

In a clinical study with the NTI drug warfarin administered to 10 healthy subjects on three separate study days, individual WSVs for both enantiomers (R- and S-warfarin) ranged from 3.7 to 19.1% (AUC) and from 2.5 to 16.2% ($C_{max}$). Two BE tests were performed on a WSV distribution obtained by bootstrapping 1000 replicates of the clinical data, yielding passing rates of 95–97% for the primary criterion (mean comparison) and 84–87% for the secondary criterion (WSV comparison). The passing rate for WSV comparison was lower than expected for an NTI drug tested against itself; however, it appears reasonable given the small sample size of 10 subjects (in comparison to commonly applied sample sizes of around $n = 24$). The WSV comparison was thought to provide further assurance of BE (Jayachandran et al., 2019).

A major drawback of the EMA criteria is the low study power in case of larger WSVs: the RSABE criterion of US-FDA (without 80.00–125.00% capping) ensures a close to 100% passing rate when the reference product is compared to itself or an identical generic product. In contrast, the passing rate under the tightened average BE limits of 90.00–111.11% (EMA) decreases significantly when the WSV of reference increases, with a statistical power of <30% in case the WSV approaches 25% (Jiang et al., 2015). Consequently, large sample sizes are needed for BE studies of such drug products, thereby increasing the burden for the development of generics.

*2.5.2.2. Alternative strategies for BE evaluation of NTI drug products.* Regarding the US-FDA's approach it was criticised that the applied BE criterion is a disaggregated criterion and that the acceptable range for the BE in variability is debatable. Thus, a simple 90% confidence region approach (i.e., an aggregated criterion) was proposed instead, in which a 90% confidence region is constructed for the difference of means of test and reference and the ratio of test to reference variability. BE is claimed, if the 90% CI of the combined metrics is totally within that BE acceptance region. Similar to the US-FDA's approach, a fully replicate crossover design for assessment of WSV for reference and test product is recommended.

Regarding the European approach it was pointed out that due to the low WSV that is normally expected to be present for NTI drug products, the so-called generic drifting can be more problematic resulting in the request for tighter acceptance criteria. However, the consequence of very large sample sizes in case of increased WSV, e.g., for tacrolimus, colchicine, ciclosporin or levothyroxine, is recognised. Therefore, a refined statistical approach was proposed that could help in harmonising the BE criteria of EMA and US-FDA. In the proposed BE strategy using, e.g., a semi-replicate, three-way crossover study design, the 80.00–125.00% BE acceptance range is narrowed based on the determined WSV of the reference product similar to the concept described for HVD products (see Section 2.1.1.1). The acceptance range is narrowed in case the WSV is <30%, down to 90–111% for WSV <13.93%. As has been demonstrated for NTI drugs like tacrolimus and colchicine with a predicted moderate WSV of 20–30%, the proposed approach involves smaller sample sizes compared to the current EMA approach without an expected decrease in safety (Paixão et al., 2022a). It was recognized, though, that this method is also associated with an increase in the type-I error (<7%) (Paixão et al., 2022b).

*2.5.2.3. Challenges in global development of generic NTI drug products – an industry's perspective.* The challenges from an industry's perspective are the lack of a definition of NTI drugs using clear-cut criteria and consequently, the lack of a universally accepted list of NTI drugs or existence of deviating lists in different countries, and the lack of harmonised BE criteria for NTI drug products. An extensive debate is ongoing, especially at the regulatory level, on defining criteria for NTI drugs, and low to moderate WSV ($\leq$30%) has been proposed as one criterion. The development of generics is further complicated due to changing recommendations for NTI drug products and delay in availability of up-to-date lists or PSGs. Also, recommendations of different authorities regarding BE demonstration of NTI drug products may show similarities but small to fundamental differences. A rough grouping could be considered into countries using a scaling approach (USA and China), countries applying an approach combining standard acceptance range and a PEC (Japan) and countries applying tightened but constant limits to AUC and,

sometimes, $C_{max}$ (Europe and Canada).

Further, application of tightened BE acceptance limits of 90.00–111.11% can – especially for NTI drug products with moderate to high WSV observed in one specific application – result in large, required sample sizes that are no longer reasonable for a comparative BE study. Comparison of the required sample sizes for NTI drug products with moderate to high WSV in one specific application such as those containing tacrolimus and ciclosporin revealed potentially far larger sample sizes when using the EMA approach of fixed narrower limits compared to the scaling approach of US-FDA (Paixão et al., 2022a).

Therefore, the following suggestions have been made that would facilitate the development of affordable high-quality generics of NTI drug products: definition of criteria for NTI drugs, timely provision of NTI drug lists or PSGs and use of a harmonised BE scaling approach for AUC and $C_{max}$.

#### 2.5.3. Conclusions and suggestions

Since the development of affordable generic versions of NTI drug products is particularly challenging, harmonisation of definitions and regulatory BE recommendations is key to facilitate that process. Major objectives for harmonisation are the definition of clear-cut criteria for NTI drugs and the implementation of a scientifically justified BE approach, enabling reasonable sample sizes for BE studies.

It was discussed that the "one-size-fits-all" approach with narrowed BE limits of 90.00–111.11% is arbitrary, not scientifically justifiable and may often not be optimal. Instead, a scaled BE approach did not face substantial opposition in general, although details of the procedure were not discussed in detail outside the different presentations. The majority of respondents of the survey (**Supplemental Table**) considered the suggested scaling approach without WSV comparison published by (Paixão et al., 2022a) as a good compromise between the criteria of EMA and US-FDA. A further discussion concentrated on whether the narrowed BE limits should apply to both PK parameters, AUC and $C_{max}$, and the majority of respondents favoured the EMA approach demanding the tighter BE limits depending on the clinical significance of the parameter. Respondents of the survey were equally split regarding the need for WSV comparison between test and reference product. It was previously mentioned that the comparison of WSV is not meaningful, especially if it is low (Endrenyi and Tothfalusi, 2013). A strong argument against this recommendation is that the complex fully replicate study design would no longer be necessary, if the WSV comparison is skipped. So, whether the variability comparison between test and reference formulations is necessary remains debatable.

As has been stated for HVD products, the topic of statistical assessment of studies with NTI drug products will be addressed in Tier 3 of the development of ICH M13 guidance on Bioequivalence for Immediate-Release Solid Oral Dosage Forms (ICH, 2022a), expected to start in 2024. The in-depth scientific discussions during this conference will help to direct focused harmonisation efforts within the ICH M13 expert working group.

## 3. Overall conclusion and perspective

The primary goal of the Global Bioequivalence Harmonisation Initiative (GBHI) is the facilitation of science-driven recommendations in the field of BE assessment. The prerequisites to achieve this goal are the definition of open issues and problems to be solved, scientific research and the exchange of scientific findings amongst the relevant stakeholders including academia, industry, and regulatory agencies. In that perspective, GBHI conferences will continue to provide a scientific platform, enabling open exchange between scientists, industry and regulators, thereby fostering the elaboration and harmonisation of scientifically justified recommendations and eventually facilitating the development of high-quality, affordable generic products.

Relevant progress in harmonisation of BE criteria and standards is currently expected from the ICH M13 initiative. Ahead of ICH, the GBHI conferences provide insight into the underlying scientific approaches and help identifying the scientific needs, which persist and require clarification before regulatory standards can be defined.

The recent conference helped identifying open issues in the field of specific statistical approaches covering HVD products, adaptive designs and modelling/simulation approaches, study recommendations referring to fed/fasted state, NTI drugs and topical products. Many of the topics discussed at this conference, specifically statistical approaches for HVD products, adaptive designs, and NTI drugs were timely and will support the development of ICH M13C (Tier 3 topics) which is anticipated to start in 2024.

Continuous scientific effort and progress is needed to further support the international scientific harmonisation process.

## Disclaimer & acknowledgements

Opinions expressed in this article are those of the authors and do not necessarily represent the views or policies of their affiliated organisations/agencies. The authors are grateful to the Frankfurt Foundation Quality of Medicines for funding open access of this article. This research did not receive any further specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Declaration of Competing Interest

None.

## Data availability

No data was used for the research described in the article.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejps.2023.106566.

## References

Abrahamsson, B., Albery, T., Eriksson, A., Gustafsson, I., Sjöberg, M., 2004. Food effects on tablet disintegration. Eur. J. Pharm. Sci. 22, 165–172.

Anderson, C., Andersson, T., Molander, M., 1991. Ethanol absorption across human skin measured by *in vivo* microdialysis technique. Acta Derm. Venereol. 71, 389–393.

ANVISA, 2022. RDC N° 742, criteria for relative bioavailability/bioequivalence (BD/BE) and pharmacokinetic studies (https://alimentusconsultoria.com.br/resolucao-rdc-no-742-de-10-de-agosto-de-2022-anvisa/, accessed 11 Nov 2022).

Blume, H.H., Mehta, M., Beuerle, G., Dorantes, A., Hempel, G., Jiang, W., Kovar, A., Lee, J., Potthast, H., Schug, B., Seidlitz, A., Tampal, N., Tsang, Y.C., Walstab, J., Welink, J., 2021. The Global Bioequivalence Harmonisation Initiative (GBHI): report of EUFEPS/AAPS fourth conference. Eur. J. Pharm. Sci. 167, 105987.

Bodenlenz, M., Dragatin, C., Liebenberger, L., Tschapeller, B., Boulgaropoulos, B., Augustin, T., Raml, R., Gatschelhofer, C., Wagner, N., Benkali, K., Rony, F., Pieber, T., Sinner, F., 2016. Kinetics of Clobetasol-17-propionate in psoriatic lesional and non-lesional skin assessed by dermal open flow microperfusion with time and space resolution. Pharm. Res. 33, 2229–2238.

Bodenlenz, M., Höfferer, C., Magnes, C., Schaller-Ammann, R., Schaupp, L., Feichtner, F., Ratzer, M., Pickl, K., Sinner, F., Wutte, A., Korsatko, S., Köhler, G., Legat, F.J., Benfeldt, E.M., Wright, A.M., Neddermann, D., Jung, T., Pieber, T.R., 2012. Dermal PK/PD of a lipophilic topical drug in psoriatic patients by continuous intradermal membrane-free sampling. Eur. J. Pharm. Biopharm. 81, 635–641.

Bodenlenz, M., Tiffner, K.I., Raml, R., Augustin, T., Dragatin, C., Birngruber, T., Schimek, D., Schwagerle, G., Pieber, T.R., Raney, S.G., Kanfer, I., Sinner, F., 2017. Open flow microperfusion as a dermal pharmacokinetic approach to evaluate topical bioequivalence. Clin. Pharmacokinet. 56, 91–98.

Caspers, P.J., Lucassen, G.W., Carter, E.A., Bruining, H.A., Puppels, G.J., 2001. *In vivo* confocal Raman microspectroscopy of the skin: noninvasive determination of molecular concentration profiles. J. Invest. Dermatol. 116, 434–442.

Chen, M.L., Blume, H., Beuerle, G., Davit, B., Mehta, M., Potthast, H., Schug, B., Tsang, Y. C., Wedemeyer, R.S., Weitschies, W., Welink, J., 2018. The global bioequivalence harmonization initiative: summary report for EUFEPS international conference. Eur. J. Pharm. Sci. 111, 153–157.

Chen, M.L., Blume, H., Beuerle, G., Mehta, M., Potthast, H., Brandt, A., Schug, B.S., Ducharme, M., Endrenyi, L., Gallicano, K., Schuirmann, D., Welink, J., 2019.

Summary report of second EUFEPS/AAPS conference on global harmonization in bioequivalence. Eur. J. Pharm. Sci. 127, 24–28.
Davit, B.M., Chen, M.L., Conner, D.P., Haidar, S.H., Kim, S., Lee, C.H., Lionberger, R.A., Makhlouf, F.T., Nwakama, P.E., Patel, D.T., Schuirmann, D.J., Yu, L.X., 2012. Implementation of a reference-scaled average bioequivalence approach for highly variable generic drug products by the US food and drug administration. AAPS J. 14, 915–924.
DeRosa, G., 2019. FY 2019 generic drug regulatory science initiatives public workshop: reducing the burden of proof – re-evaluating the necessity of fed bioequivalence studies, https://www.fda.gov/news-events/fy-2019-generic-drug-regulatory-science-initiatives-public-workshop-05012019-05012019, accessed 29 Oct 2020.
Dragatin, C., Polus, F., Bodenlenz, M., Calonder, C., Aigner, B., Tiffner, K.I., Mader, J.K., Ratzer, M., Woessner, R., Pieber, T.R., Cheng, Y., Loesche, C., Sinner, F., Bruin, G., 2016. Secukinumab distributes into dermal interstitial fluid of psoriasis patients as demonstrated by open flow microperfusion. Exp. Dermatol. 25, 157–159.
EMA, 2007. Guideline on reporting the results of population pharmacokinetic analyses, CHMP/EWP/185990/06, (https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-reporting-results-population-pharmacokinetic-analyses_en.pdf, accessed 17 Oct 2022).
EMA, 2010. Guideline on the investigation of bioequivalence (https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-bioequivalence-rev1_en.pdf, accessed 06 Oct 2022).
EMA, 2011. Presentation: role of modelling and simulation in regulatory decision making in Europe (https://www.ema.europa.eu/en/documents/presentation/presentation-role-modelling-simulation-regulatory-decision-making-europe_en.pdf, accessed 17 Oct 2022).
EMA, 2012. Guideline on the investigation of drug interactions, CPMP/EWP/560/95/Rev. 1 Corr. 2** (https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-drug-interactions-revision-1_en.pdf, accessed 17 Oct 2022).
EMA, 2014. Guideline on the pharmacokinetics and clinical evaluation of modified release dosage forms, EMA/CPMP/EWP/280/96 Corr1 (https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-pharmacokinetic-clinical-evaluation-modified-release-dosage-forms_en.pdf, accessed 17 Oct 2022).
EMA, 2018. Draft guideline on quality and equivalence of topical products (https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-quality-equivalence-topical-products_en.pdf, accessed 10 Nov 2022).
EMA, 2020. Qualification of novel methodologies for drug development: guidance to applicants. EMA/CHMP/SAWP/72894/2008 (https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/qualification-novel-methodologies-drug-development-guidance-applicants_en.pdf, accessed 13 Dec 2022).
EMA, 2022. Product-specific bioequivalence guidance (https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-guidelines/clinical-pharmacology-pharmacokinetics/product-specific-bioequivalence-guidance, accessed 16 Nov 2022).
Endrenyi, L., Tothfalusi, L., 2013. Determination of bioequivalence for drugs with narrow therapeutic index: reduction of the regulatory burden. J. Pharm. Pharm. Sci. 16, 676–682.
Escobar-Chávez, J.J., Merino-Sanjuán, V., López-Cervantes, M., Urban-Morlan, Z., Piñón-Segundo, E., Quintanar-Guerrero, D., Ganem-Quintanar, A., 2008. The tape-stripping technique as a method for drug quantification in skin. J. Pharm. Pharm. Sci. 11, 104–130.
Fuglsang, A., 2014. Sequential bioequivalence approaches for parallel designs. AAPS J. 16, 373–378.
Gozzo, L., Caraci, F., Drago, F., 2022. Bioequivalence, drugs with narrow therapeutic index and the phenomenon of biocreep: a critical analysis of the system for generic substitution. Healthcare (Basel 10.
Health Canada, 2018. Comparative bioavailability standards: formulations used for systemic effects (https://www.canada.ca/content/dam/hc-sc/migration/hc-sc/dhp-mps/alt_formats/pdf/prodpharma/applic-demande/guide-ld/bio/comparative-bioavailability-standards-formulations-used-systemic-effects.pdf, accessed 06 Oct 2022).
ICH, 2021. Guidance for industry - M9 biopharmaceutics classification system-based biowaivers (https://www.fda.gov/media/148472/download, accessed 09 Nov 2022).
ICH, 2022. ICH harmonised guideline: bioequivalence for immediate-release solid oral dosage forms M13A, Draft dated 20 Dec 2022 (https://database.ich.org/sites/default/files/ICH_M13A_Step2_draft_Guideline_2022_1125.pdf, accessed 20 Feb 2023).
ICH, 2022. M15: model-informed drug development general principles guideline (https://database.ich.org/sites/default/files/ICH_M15_ConceptPaper_Final_2022_1102.pdf, accessed 08 June 2023).
Jayachandran, P., Okochi, H., Frassetto, L.A., Park, W., Fang, L., Zhao, L., Benet, L.Z., 2019. Evaluating within-subject variability for narrow therapeutic index drugs. Clin. Pharmacol. Ther. 105, 411–416.
Jiang, W., Makhlouf, F., Schuirmann, D.J., Zhang, X., Zheng, N., Conner, D., Yu, L.X., Lionberger, R., 2015. A bioequivalence approach for generic narrow therapeutic index drugs: evaluation of the reference-scaled approach and variability comparison criterion. AAPS J. 17, 891–901.
Jiang, X., Yuzhuo, P., Sun, W., 2021. A novel visual tool for the “unfavorable zone” of Potvin’s methods for adaptive sequential design for pharmacokinetic bioequivalence studies, AAPS eposter 10/20/21, W4104V (https://posters.aaps.org/aaps/2021/2021-aaps-pharmsci-360/334242/xiaojian.jiang.a.novel.visual.tool.for.the.unfavorable.zone.of.potvins.methods.html?f=menu%3D6%2Abrowseby%3D8%2Asortby%3D2%2Amedia%3D3%2Ace_id%3D2047%2Aot_id%3D26424%2Amarker%3D1401, accessed 13 Oct 2022).
Kimko, H., Gibiansky, E., Gibiansky, L., Starr, H.L., Berwaerts, J., Massarella, J., Wiegand, F., 2012. Population pharmacodynamic modeling of various extended-release formulations of methylphenidate in children with attention deficit hyperactivity disorder via meta-analysis. J. Pharmacokinet. Pharmacodyn. 39, 161–176.
Kourentas, A., Vertzoni, M., Barmpatsalou, V., Augustijns, P., Beato, S., Butler, J., Holm, R., Ouwerkerk, N., Rosenberg, J., Tajiri, T., Tannergren, C., Symillides, M., Reppas, C., 2018. The BioGIT system: a valuable *in vitro* tool to assess the impact of dose and formulation on early exposure to low solubility drugs after oral administration. AAPS J. 20, 71.
Koziolek, M., Alcaro, S., Augustijns, P., Basit, A.W., Grimm, M., Hens, B., Hoad, C.L., Jedamzik, P., Madla, C.M., Maliepaard, M., Marciani, L., Maruca, A., Parrott, N., Pávek, P., Porter, C.J.H., Reppas, C., van Riet-Nales, D., Rubbens, J., Statelova, M., Trevaskis, N.L., Valentová, K., Vertzoni, M., Čepo, D.V., Corsetti, M., 2019. The mechanisms of pharmacokinetic food-drug interactions - a perspective from the UNGAP group. Eur. J. Pharm. Sci. 134, 31–59.
Koziolek, M., Grimm, M., Schneider, F., Jedamzik, P., Sager, M., Kühn, J.P., Siegmund, W., Weitschies, W., 2016. Navigating the human gastrointestinal tract for oral drug delivery: uncharted waters and new frontiers. Adv. Drug. Deliv. Rev. 101, 75–88.
Lerner, A., Matthias, T., 2015. Changes in intestinal tight junction permeability associated with industrial food additives explain the rising incidence of autoimmune disease. Autoimmun. Rev. 14, 479–489.
Lionberger, R.A., 2008. FDA critical path initiatives: opportunities for generic drug development. AAPS J. 10, 103–109.
Maurer, W., Jones, B., Chen, Y., 2018. Controlling the type I error rate in two-stage sequential adaptive designs when testing for average bioequivalence. Stat. Med. 37, 1587–1607.
Mehta, M., Blume, H., Beuerle, G., Tampal, N., Schug, B., Potthast, H., Jiang, W., Wilson, C., 2020. The global bioequivalence harmonisation initiative: report of EUFEPS/AAPS third conference. Eur. J. Pharm. Sci. 151, 105383.
Meidan, V.M., Roper, C.S., 2008. Inter- and intra-individual variability in human skin barrier function: a large scale retrospective study. Toxicol. *In Vitro* 22, 1062–1069.
Midha, K.K., Rawson, M.J., Hubbard, J.W., 2005. The bioequivalence of highly variable drugs and drug products. Int. J. Clin. Pharmacol. Ther. 43, 485–498.
Miranda, M., Cova, T., Augusto, C., Pais, A., Cardoso, C., Vitorino, C., 2020. Diving into batch-to-batch variability of topical products-a regulatory bottleneck. Pharm. Res. 37, 218.
Miranda, M., Sousa, J.J., Veiga, F., Cardoso, C., Vitorino, C., 2018a. Bioequivalence of topical generic products. Part 1: where are we now? Eur. J. Pharm. Sci. 123, 260–267.
Miranda, M., Sousa, J.J., Veiga, F., Cardoso, C., Vitorino, C., 2018b. Bioequivalence of topical generic products. Part 2. Paving the way to a tailored regulatory system. Eur. J. Pharm. Sci. 122, 264–272.
Molins, E., Cobo, E., Ocaña, J., 2017. Two-stage designs versus European scaled average designs in bioequivalence studies for highly variable drugs: which to choose? Stat. Med. 36, 4777–4788.
Monteiro, C.A., Levy, R.B., Claro, R.M., de Castro, I.R., Cannon, G., 2011. Increasing consumption of ultra-processed foods and likely impact on human health: evidence from Brazil. Public Health Nutr. 14, 5–13.
Muñoz, J., Alcaide, D., Ocaña, J., 2016. Consumer’s risk in the EMA and FDA regulatory approaches for bioequivalence in highly variable drugs. Stat Med 35, 1933–1943.
Novakovic, J., Szirtes, J., Fields, A., Tsang, Y.C., 2019. Clinical endpoint bioequivalence studies are not sensitive: a perspective from generic drugs. Clin. Pharmacol. Ther. 105, 295–297.
O’Brien, P.C., Fleming, T.R., 1979. A multiple testing procedure for clinical trials. Biometrics 35, 549–556.
Ocaña, J., Muñoz, J., 2019. Controlling type I error in the reference-scaled bioequivalence evaluation of highly variable drugs. Pharm. Stat. 18, 583–599.
OECD, 2019. OECD environment, health and safety publications series on testing and assessment No. 156 - guidance notes on dermal absorption (https://www.oecd.org/chemicalsafety/testing/Guidance%20Notes%20Dermal%20Absorption%20156_Oct2019_clean.pdf, accessed 14 Nov 2022).
Paixão, P., Guerreiro, R.B., Silva, N., Blake, K., Bonelli, M., Morais, J.A.G., Garcia Arieta, A., Gouveia, L.F., 2022a. A proposed approach for the determination of the bioequivalence acceptance range for narrow therapeutic index drugs in the European Union. Clin. Pharmacol. Ther. 111, 470–476.
Paixão, P., Silva, N., Guerreiro, R.B., Blake, K., Bonelli, M., Morais, J.A.G., García-Arieta, A., Gouveia, L.F., 2022b. Evaluation of a proposed approach for the determination of the bioequivalence acceptance range for narrow therapeutic index drugs in the European Union. Pharmaceutics 14.
Pleguezuelos-Villa, M., Merino-Sanjuán, M., Hernández, M.J., Nácher, A., Peris, D., Hidalgo, I., Soler, L., Sallan, M., Merino, V., 2019. Relationship between rheological properties, *in vitro* release and *in vivo* equivalency of topical formulations of diclofenac. Int. J. Pharm. 572, 118755.
PMDA, 2020. Guideline for bioequivalence studies for different strengths of oral solid dosage form - english translation of attachment 2 of PSEHB/PED notification No. 0319-1: (https://www.nihs.go.jp/drug/be-guide(e)/2020/GL2_diffrent_strengths_2020.pdf, accessed 16 Nov 2022).
PMDA, 2020. Guideline for bioequivalence studies of generic products - english translation of attachment 1 of PSEHB/PED notification No. 0319-1: (https://www.nihs.go.jp/drug/be-guide(e)/2020/GL1_BE_2020.pdf, accessed 16 Nov 2022).
Pocock, S.J., 1977. Group sequential methods in the design and analysis of clinical trials. Biometrika 64, 191–199.
Potvin, D., DiLiberti, C.E., Hauck, W.W., Parr, A.F., Schuirmann, D.J., Smith, R.A., 2008. Sequential design approaches for bioequivalence studies with crossover designs. Pharm. Stat. 7, 245–262.

Sager, M., Schick, P., Mischek, M., Schulze, C., Hasan, M., Kromrey, M.L., Benameur, H., Wendler, M., Tzvetkov, M.V., Weitschies, W., Koziolek, M., 2019. Comparison of *in vitro* and *in vivo* results using the GastroDuo and the salivary tracer technique: immediate release dosage forms under fasting conditions. Pharmaceutics 11, 0659.

Schick, P., Sager, M., Voelker, M., Weitschies, W., Koziolek, M., 2020. Application of the GastroDuo to study the interplay of drug release and gastric emptying in case of immediate release Aspirin formulations. Eur. J. Pharm. Biopharm. 151, 9–17.

Schick, P., Sager, M., Wegner, F., Wiedmann, M., Schapperer, E., Weitschies, W., Koziolek, M., 2019. Application of the GastroDuo as an *in vitro* dissolution tool to simulate the gastric emptying of the postprandial stomach. Mol. Pharm. 16, 4651–4660.

Schnabel, L., Buscail, C., Sabate, J.M., Bouchoucha, M., Kesse-Guyot, E., Allès, B., Touvier, M., Monteiro, C.A., Hercberg, S., Benamouzig, R., Julia, C., 2018. Association between ultra-processed food consumption and functional gastrointestinal disorders: results from the French NutriNet-Santé Cohort. Am. J. Gastroenterol. 113, 1217–1228.

Schütz, H., Labes, D., Wolfsegger, M.J., 2022. Critical remarks on reference-scaled average bioequivalence. J. Pharm. Pharm. Sci. 25, 285–296.

Shah, V.P., Yacobi, A., Barr, W.H., Benet, L.Z., Breimer, D., Dobrinska, M.R., Endrenyi, L., Fairweather, W., Gillespie, W., Gonzalez, M.A., Hooper, J., Jackson, A., Lesko, L.J., Midha, K.K., Noonan, P.K., Patnaik, R., Williams, R.L., 1996. Evaluation of orally administered highly variable drugs and drug formulations. Pharm. Res. 13, 1590–1594.

Sharan, S., Choi, S., Zou, Y., Wang, Y., Kim, M.J., Fang, L., Choi, S., Makhlouf, F., Grosser, S.C., Zhang, X., Zhao, L., 2022. Application of modeling and simulation to identify a shortened study duration and novel bioequivalence metric for a long-acting intrauterine system. AAPS J. 24, 63.

Tiffner, K.I., Ramezanli, T., Birngruber, T., Lackner, B., Raml, R., Bodenlenz, M., Schwagerle, G., Kainz, S., Gander, E., Raney, S.G., Sinner, F., 2020. A clinical study to assess the bioequivalence of lidocaine and prilocaine topical drug products using dermal open flow microperfusion, AAPS, poster 923442 (https://posters.aaps.org/aaps/2020/eposters/309717/katrin.tiffner.a.clinical.study.to.assess.the.bioequivalence.of.lidocaine.and.html?f=listing%3D3%2Abrowseby%3D8%2Asortby%3D1%2Amedia%3D1, accessed 15 Nov 2022).

Tsakalozou, E., Babiskin, A., Zhao, L., 2021. Physiologically-based pharmacokinetic modeling to support bioequivalence and approval of generic products: a case for diclofenac sodium topical gel, 1. CPT Pharmacometrics Syst. Pharmacol. 10, 399–411.

Tsamandouras, N., Rostami-Hodjegan, A., Aarons, L., 2015. Combining the 'bottom up' and 'top down' approaches in pharmacokinetic modelling: fitting PBPK models to observed clinical data. Br. J. Clin. Pharmacol. 79, 48–55.

US-FDA, 2011. Draft guidance on clobetasol propionate (aerosol, foam/topical) (https://www.accessdata.fda.gov/drugsatfda_docs/psg/Clobetasol_Propionate_fmaerosol_21142_RC2-11.pdf, accessed 10 Nov 2022).

US-FDA, 2016. Draft guidance on acyclovir (cream, topical) (https://www.accessdata.fda.gov/drugsatfda_docs/psg/Acyclovir_topical%20cream_RLD%2021478_RV12-16.pdf, accessed 13 Nov 2022).

US-FDA, 2018. Draft guidance on diclofenac sodium (gel, topical) (https://www.accessdata.fda.gov/drugsatfda_docs/psg/Diclofenac%20Sodium_draft_Topical%20gel_RLD%2022122_RC07-18.pdf, accessed 02 Nov 2022).

US-FDA, 2018. Draft guidance on methylphenidate hydrochloride (extended release tablet, oral) (https://www.accessdata.fda.gov/drugsatfda_docs/psg/Methylphenidate%20Hydrochloride_draft_Oral%20tab%20ER_RLD%2021121_RC07-18.pdf, accessed 02 Nov 2022).

US-FDA, 2019. Adaptive designs for clinical trials of drugs and biologics - guidance for industry (https://www.fda.gov/media/78495/download, accessed 13 Oct 2022).

US-FDA, 2020. Draft guidance on clindamycin phosphate (swab, topical) (https://www.accessdata.fda.gov/drugsatfda_docs/psg/PSG_050537-Swa.pdf, accessed 10 Nov 2022).

US-FDA, 2020. Draft guidance on levonorgestrel (intrauterine device) (https://www.accessdata.fda.gov/drugsatfda_docs/psg/PSG_021225.pdf, accessed 03 Nov 2022).

US-FDA, 2021. Guidance for industry - bioequivalence studies with pharmacokinetic endpoints for drugs submitted under an ANDA. Draft guidance (https://www.fda.gov/media/87219/download, accessed 13 Oct 2022).

US-FDA, 2022. Draft guidance on doxepin hydrochloride (cream, topical) (https://www.accessdata.fda.gov/drugsatfda_docs/psg/PSG_020126.pdf, accessed 13 Nov 2022).

US-FDA, 2022. Guidance for industry - formal meetings between FDA and ANDA applicants of complex products under GDUFA (https://www.fda.gov/media/107626/download, accessed 01 Nov 2022).

US-FDA, 2022. Guidance for industry - *in vitro* permeation test studies for topical drug products submitted in ANDAs (draft guidance) (https://www.fda.gov/regulatory-information/search-fda-guidance-documents/in-vitro-permeation-test-studies-topical-drug-products-submitted-andas, accessed 06 June 2023).

US-FDA, 2022 Guidance for industry - *in vitro* release test studies for topical drug products submitted in ANDAs (draft guidance) (https://www.fda.gov/regulatory-information/search-fda-guidance-documents/in-vitro-release-test-studies-topical-drug-products-submitted-andas, accessed 06 June 2023).

US-FDA, 2022e. Guidance for industry - physicochemical and structural (Q3) characterization of topical drug products submitted in ANDAs (draft guidance) (https://www.fda.gov/regulatory-information/search-fda-guidance-documents/physicochemical-and-structural-q3-characterization-topical-drug-products-submitted-andas, accessed 06 June 2023).

US-FDA, 2022f. Guidance for industry - topical dermatologic corticosteroids: *in vivo* bioequivalence, draft guidance (https://www.fda.gov/media/162457/download, accessed 13 Dec 2022).

Wagner, C., Kesisoglou, F., Pepin, X.J.H., Parrott, N., Emami Riedmaier, A., 2021. Use of physiologically based pharmacokinetic modeling for predicting drug-food interactions: recommendations for improving predictive performance of low confidence food effect models. AAPS J. 23, 85.

Yu, L.X., Jiang, W., Zhang, X., Lionberger, R., Makhlouf, F., Schuirmann, D.J., Muldowney, L., Chen, M.L., Davit, B., Conner, D., Woodcock, J., 2015. Novel bioequivalence approach for narrow therapeutic index drugs. Clin. Pharmacol. Ther. 97, 286–291.

Zaheer, K., Langguth, P., 2018. Formulation strategy towards minimizing viscosity mediated negative food effect on disintegration and dissolution of immediate release tablets. Drug Dev. Ind. Pharm. 44, 444–451.

Zaheer, K., Langguth, P., 2019. Designing robust immediate release tablet formulations avoiding food effects for BCS class 3 drugs. Eur. J. Pharm. Biopharm. 139, 177–185.

Zhao, L., Kim, M.J., Zhang, L., Lionberger, R., 2019. Generating model integrated evidence for generic drug development and assessment. Clin. Pharmacol. Ther. 105, 338–349.