In vitro studies into establishing therapeutic bioequivalence of complex topical products: Weight of evidence

Margarida Miranda [a,b,c,d], Zoe Volmer [e], Alicia Cornick [e], Aidan Goody [e], Catarina Cardoso [c], Alberto A.C.C. Pais [b], Marc Brown [e,*], Carla Vitorino [b,d,f,*]

[a] Egas Moniz Center for Interdisciplinary Research (CiiEM), Egas Moniz School of Health & Science, Monte de Caparica, Portugal
[b] Coimbra Chemistry Center, Department of Chemistry, University of Coimbra, Portugal
[c] Laboratórios Basi, Mortágua, Portugal
[d] Faculty of Pharmacy, University of Coimbra, Portugal
[e] MedPharm Ltd, Surrey Research Centre, Guildford, Surrey, UK
[f] Polo das Ciências da Saúde, Azinhaga Santa Comba Polo 3, Coimbra, 3000-548 Portugal

ARTICLE INFO

ABSTRACT

Over the past decade, topically applied drug products have experienced extraordinary price increases, due to the shortage of multisource generic drug products. This occurrence is mainly related to the underlying challenges evolved in topical bioequivalence documentation. Although there has been continuing regulatory efforts to present surrogate *in vitro* methods to clinical endpoint studies, there is still a continued need for cost- and time-efficient alternatives that account for product specificities. Hence, this work intended to expose bioequivalence assessment issues for complex topical formulations, and more specifically those related with product efficacy guidance. As a model drug and product, a bifonazole 10 mg/g cream formulation was selected and two different batches of the commercially available Reference Product (RP) were used: RP1 that displayed lower viscosity and RP4 which presented high, but not the highest, viscosity. *In vitro* human skin permeation testing (IVPT) was carried out and the results were evaluated by means of the traditional bioequivalence assessment approach proposed by the EMA, as well as by the Scaled Average Bioequivalence assessment approach proposed by the FDA. Based on previous experience, there was an expectation of a high level of variability in the results, thus alternative methods to evaluate local drug skin availability were developed. More specifically, an infected skin disease model, where *ex vivo* human skin was infected and ATP levels were used as a biological marker for monitoring antifungal activity after product application. The results showed that permeation equivalence could not be supported between the different RP batches. In contrast, this statistical difference between the formulation batches was not indicated in the disease model. Nevertheless, in pivotal IVPT studies, the lowest permeant formulation (RP4) evidenced a higher antifungal *in vitro* activity as reported by the lower levels of ATP. A critical appraisal of the results is likewise presented, focusing on an outlook of the real applicability of the regulatory guidances on this subject.

1. Introduction

Topically applied dosage forms, commonly developed to exert a local action, have been used throughout history for cosmetic and therapeutic purposes, being one of the oldest medicinal dosage forms known to human civilization (Benson and Watkinson, 2012; Ilić et al., 2021).

Semisolids represent a significant proportion of these pharmaceutical preparations. These can be further subdivided into single-phase systems, in which all vehicle ingredients are miscible and dissolve into each other, and into multiphase systems, in which distinct solubility profile components can be combined. Examples of the first group are gels and ointments, while creams and pastes are key examples of semisolids with a multiphasic structure (Brown and Williams, 2019; Surber and Knie, 2018). All exhibit enhanced spreadability and bioadhesion characteristics due to their rheological properties. These sensorial attributes stimulate patient compliance to treatment, which is important as the patient has to apply the formulation directly to where the disease is visible. From a drug delivery perspective, this is also beneficial due to the avoidance of the hepatic first-pass metabolism and a lower side effect profile (Fernández-Campos et al., 2017; Simões et al., 2020; Surber and Knie, 2018).

The constraints regarding the development of new chemical entities and patient need to acquire more affordable drug products have led to an expansion of the generic drug products market (Brown and Williams, 2019; Fernández-Campos et al., 2017). Even though there are straightforward and highly defined regulatory mechanisms concerning bioequivalence documentation for oral generic products, the operational and regulatory aspects involved in topical bioequivalence demonstration are still far from being as reachable (Fernández-Campos et al., 2017; Lu et al., 2016; Miranda et al., 2018). The associated challenges are mainly related with the complexity of topical formulations (other than topical solutions), as well as with the complex delivery route associated with these products (US-FDA, 2022a). In an attempt to facilitate the development of topical generic drug products, several guidelines have been issued by the EMA and FDA. The agencies advise striving for qualitative (Q1), quantitative (Q2), microstructure (Q3) and performance (Q4) sameness of the test product to the comparator medicinal product, herein referred as the Reference product (RP). Then, mostly for multiphasic formulations , equivalence pertaining to the local drug skin availability should be achieved and can be determined through *in vitro* permeation tests (IVPT), tape stripping or vasoconstrictor studies, the latter solely applicable to corticosteroid products (EMA, 2018; Food and Drug Administration, CDER, 2002; Miranda et al., 2023a; US-FDA, 2022a). The guidelines provide defined acceptance criteria regarding all stages of the bioequivalence assessment workflow. However, several studies in the literature have shown that implementation of some of these acceptance criteria in daily practice might be challenging (Ilić et al., 2021; Mangas-Sanjuán et al., 2019; Miranda et al., 2023b, 2022, 2020b; Xu et al., 2020).

A topical antifungal cream formulation, more specifically a 10 mg/g bifonazole cream, was selected for the present study (Miranda et al., 2023b, Miranda et al., 2020a, 2020b). Topical antifungals represent the main therapeutic approach to dermatophytosis, a condition increasing worldwide due to changes in the socioeconomic status, occupation, climate as well as to individual predisposing factors (Sahoo and Mahajan, 2016). Bifonazole belongs to the imidazole group, together with a variety of other compounds such as clotrimazole, econazole, fenticonazole, ketoconazole, isoconazole, miconazole, oxiconazole, sulconazole, terconazole, tioconazole, sertaconazole and others (Burns et al., 2008). These drugs display similar properties and act by inhibiting ergosterol synthesis (Burns et al., 2008). Imidazole compounds are fungistatic and active against a wide range of fungal organisms including *Candida* and *Pityrosporum* yeasts, as well as dermatophytes. Furthermore, they also display antibacterial properties and, at least *in vitro*, can suppress *S. Aureus* growth (Burns et al., 2008). Technologically, imidazole compounds can be formulated in a wide range of dosage forms, including creams, powders, sprays, suspensions and nail lacquers (Burns et al., 2008).

In a previous research work of our group, the rheological and *in vitro* release variability of a bifonazole 10 mg/g cream Reference Product (RP) was investigated. Five different batches were screened in this study and the comparative analysis focused on microstructure (rheology) and performance (IVRT). The attained results showed statistically significant differences for both attributes, being particularly pronounced for rheology (Miranda et al., 2023a). These experimental observations led to a fundamental question to be addressed in the present study: How do these differences in microstructure and performance affect drug delivery i.e. the permeation profile? Answering this question is critical for selecting the RP batch to be used for local drug skin availability studies. From a generic manufacturer perspective, addressing this issue is key, as generic companies with an R&D project aiming a topical generic development need to be clear on which is the reference product. However, selecting a suitable RP batch whilst mitigating the risk of failure in product local drug skin availability studies, taking into account microstructure and performance results, may prove to be extremely challenging.

To address this, in the present work, IVPT studies were carried out using two RP batches with different rheology characteristics, RP1 being less viscous than RP4 (Miranda et al., 2023a). Permeation kinetic equivalence was then evaluated by means of the traditional bioequivalence assessment approach proposed by EMA, as well as by the Scaled Average Bioequivalence (SABE) assessment approach proposed by the FDA (US-FDA, 2022a). Experience suggesting a high level of variability, the present research also sought alternative methods to evaluate local drug skin delivery. According to the Draft Guideline on Quality and Equivalence of Topical Products, the presentation of a product-specific approach to documenting efficacy equivalence is possible, if it is satisfactorily validated. Within this framework, *in vitro* skin infection and decolonisation equivalence studies can be used for antifungal formulations. There are several literature reports on *ex vivo* skin infection models, but they are mainly focused on evaluating adhesion mechanisms of fungal growth and detailed analysis of gene expression pathways during skin infection (Faway et al., 2017). The vast majority of these infection models use cultured skin equivalents, which are unable to recreate the heterogeneous nature of the skin, including cellular metabolism, skin appendages, and the *stratum corneum* barrier functions (Ilić et al., 2021). In the present work, these hurdles were tentatively surpassed through the application of a novel infected skin disease model developed by MedPharm Ltd, where *ex vivo* human skin was infected and ATP levels were used as a biological marker for monitoring antifungal activity after product application. The aim of this study was to assess and compare current recommendations on the use of *ex vivo* skin models to demonstrate product bioequivalence. As a case study, 10 mg/g bifonazole creams were compared using IVPT and an infected skin disease model. The former is a more accepted and commonly used approach, but the European traditional bioequivalence assessment approach, as well as other acceptance criteria, are much tighter than those of the FDA. Conversely, an infected skin disease model provides an alternative and complementary approach to IVPT, in an attempt to provide insight into the demonstration of topical antifungal bioequivalence.

2. Materials and methods

2.1. Materials

In the present work, two distinct methodologies were employed to evaluate the topical availability of bifonazole 10 mg/g cream formulations. In the first one, regarding IVPT studies, the permeation profile of two batches of the reference formulation were compared – Canespor 10 mg/g bifonazol cream (RP1) and Canesten Extra® créme mit 1 % Bifonazol (RP4), respectively acquired from the Portuguese and German market. As controls, a placebo formulation and a 5 mg/g bifonazole cream formulation were used.

The second approach involved a disease model, tailored for antifungal activity assessment. For this test, the two RP batches were compared. A range of controls were assessed using this disease model: (i) water; (ii) 50:50 v/v polyethylene glycol 400:water and; (iii) an alternative placebo. The alternative placebo was manufactured with any

excipients that had the potential to have an antifungal effect removed. These controls were chosen to assess whether the act of dosing the infected skin with any sort of formulation resulted in a reduced ATP measurement. Table 1 summarizes the formulations used in each test, the qualitative composition of the reference product and the alternative placebo is described in the supplementary material (Table S1).

Propylene glycol, polyethylene glycol 400 and Tween 80 were acquired from Merck (Darmstadt, Germany), and phosphate buffered saline (PBS), Ringer's solution and ATP disodium hydrate were purchased from Sigma (Darmstadt, Germany). BacTiter-Glo substrate and BacTiter-Glo buffer were purchased from Promega (Southampton, United Kingdom). The lysing agent was provided by MedPharm Ltd. Water was purified using a Millipore MILLI-Q reagent water system and filtered through a 0.22 μm nylon filter before use. For the disease model experiments, HPLC grade water was acquired from VWR (Lutterworth, United Kingdom) and sterile water for irrigation from Fresenius-Kabi (Chesire, United Kingdom). All other chemicals were of analytical grade or equivalent.

2.2. Methods

2.2.1. Local availability assessment

Two methods were employed in the present study: IVPT and the antifungal disease model. For both methods, skin retrieved from human donors was used. The following paragraphs entail all the procedures involved in skin preparation.

2.2.1.1. Skin preparation. Human surgical waste skin pieces used for IVPT experiments were obtained from two different sources: (i) Centro Hospitalar de Lisboa Central, where the experimental protocol was approved by the Bioethics Committee. Written informed consent forms have been obtained from the participants involved in this study (Process number 447/2017); (ii) Genoskin® (Toulouse, France). The tissue was obtained from plastic reduction surgeries. In both skin sources, after tissue excision, all specimens were transported in saline solution (normal saline) under refrigeration (for less than 24 h). After transport, the subcutaneous fat was removed and the outer layers of skin containing the *stratum corneum* (SC), viable epidermis and some dermis were frozen at −20 °C. The day before the IVPT studies, the epidermis was isolated through a thermal process (heat separated epidermis, HSE). For that, the tissue was placed in a water bath at 60 ± 2 °C for 60 s and allowed to rest for 30 s at room temperature. With the aid of tweezers, the epidermis was separated, cut into 0.7 cm^2 and transferred to glass flasks filled with distilled water with the aid of a membrane support disk, used to keep the skin stretched. Special care was taken in order to maintain the *stratum corneum* side facing upwards. The skin sheets were then left overnight at 4 °C to stabilize.

On the day of the experiments, the skin was transferred into the

Table 1
Formulations used in the present study.

Study	Formulations	Description
IVPT	Canespor 10 mg/g bifonazol cream – RP1	Test - Pilot/Pivotal
	Canesten Extra® créme mit 1 % ifonazol – RP4	Test - Pilot/Pivotal
	Placebo	Validation
	5 mg/g bifonazol cream formulation	Validation
Disease model	Canespor 10 mg/g bifonazol cream – RP1	Test - Pivotal
	Canesten Extra® créme mit 1 % bifonazol – RP4	Test - Pivotal
	Water	Control - Pivotal
	50:50 v/v Polyethylene glycol 400:water	Control - Pivotal
	Alternative placebo*	Control, Validation

*Alternative placebo stands for a placebo formulation without excipients responsible for the antifungal activity.

diffusion cells and the barrier integrity of each skin piece was checked by measuring transepidermal water loss (TEWL) using a vapometer (Delfin Technology, Kuopio, Finland). Any skin piece with obvious signs of physical damage, stretch marks or a TEWL reading higher than 20.0 g/m^2/h was excluded from the study (Nagelreiter et al., 2013; Shin et al., 2020; Vitorino et al., 2014). Table 2 summarizes the donor demographics (sex, skin type and age), as well as the skin anatomical region.

According to EMA draft guideline requirements, for IVPT studies the number of skin donors should be 12, with at least 2 replicates per donor. All formulations should be tested using the same donor. Nevertheless, if properly justified, the number of donors may be reduced.

For the disease model, human abdominal skin, acquired from abdominoplasties performed in the United States, was obtained from ZenBio (Durham, USA) and was supplied in a frozen state. Each sample was derived from a competent volunteer adult donor who has signed an Institutional Review Board (IRB) validated donor consent form that specifically lists both the intended uses for the donation for non-clinical research and conforms the procedures for processing the samples: Standard Operating Procedure managed GLP protocols in compliance with ethical regulations. Two donors were used for the disease model investigation, one for method development/validation work and the other for the full scale active investigation. Heat separation was employed to isolate the epidermis. The tissue was first thawed before being submerged in deionised water, heated to 60 ± 2 °C, for 60 s. The epidermis was then separated and mounted on filter paper, taking care to ensure the *stratum corneum* side was facing upwards. The prepared epidermal membrane was stored at −20 °C until required.

2.2.1.2. In vitro permeation studies. IVPT studies were performed in static vertical Franz diffusion cell equipment, with a diffusion area of 0.636 cm^2 and a receptor compartment of 5 mL (PermeGear, Inc., PA, USA). Diffusional cell system and laboratory qualification studies were carried out as described in (Miranda et al., 2019). To mimic the in-use setting and to follow regulatory recommendations, all studies were performed under occlusive finite dose settings (8–12 mg/cm^2) (EMA, 2018; FDA, 2022; Kamal et al., 2020). The stirring speed was set to 600 rpm, and the temperature was controlled to ensure a 32 ± 1 °C temperature at the skin surface. Considering the limited solubility of bifonazole, a PBS-PEG (60:40, v/v, pH = 7.4) solution was used as the permeation medium. A timeframe of 48 h was considered, with sampling at 0, 4, 6, 20, 22, 24, 26, 28, 30, 44, 46 and 48 h. After each collection (300 μL), an equal volume of fresh-temperature-equilibrated permeation medium was added to the receptor chamber.

The cumulative amount of bifonazole permeated as a function of time (AMT, (μg/cm^2) was calculated using the following expression Eq. (1):

$$AMT = \frac{C_n \times V_0 + \sum_{i=1}^{n-1} C_i \times V_i}{A} \tag{1}$$

Where C_n (μg/mL) corresponds to the drug concentration of the receptor medium at each sampling time, C_i to the drug concentration of the ith sample, A to the effective diffusion area (cm^2), and V_0 and V_i to the volumes of the receptor compartment and the collected sample, respectively.

Bifonazole flux was calculated according to the FDA In Vitro Permeation Test Studies draft guidance, as well as in the acyclovir FDA draft guidance (FDA, 2022; US-FDA, 2022a). Briefly, the flux was calculated based upon: the receptor sample concentration at each time point; the measured volume of that specific diffusion cell; the area of dose application; and the duration for which the receptor volume was accepting the drug.

IVPT experiments firstly included a pilot study with 3 donors which aimed to check the suitability of the proposed conditions. Afterwards, a pivotal study, with a larger pool of donors was conducted to infer on the

Table 2
Human skin donors characteristics.

Study	Donor number	Gender	Skin type	Age	Anatomical region	Preparation method
Pilot study	1	Female	Type 3	43	Abdomen	HSE
	2	Female	Type 2	41	Abdomen	HSE
	3	Female	Type 2	29	Abdomen	HSE
Pivotal study	4	Female	Type 3	57	Arm	HSE
	5	Female	Type 3	55	Glute	HSE
	6	Female	Type 2	29	Abdomen	HSE
	7	Female	Type 2	29	Abdomen	HSE
	8	Female	Type 3	35	Thigh	HSE
	9	Female	Type 2	41	Abdomen	HSE
	10	Female	Type 3	37	Abdomen	HSE
	11	Female	Type 3	39	Thigh	HSE
Disease model	M1	Female	Unknown	26	Abdomen	HSE
	M2	Female	Type 2	55	Abdomen	HSE

Key: HSE – Heat Separated Epidermis.

permeation profile equivalence of the two RP batches.

According to the literature, as well as regulatory guidances, two endpoints should be retrieved from IVPT studies (EMA, 2018; US-FDA, 2022a). These regard the AMT and the J_{MAX}. The first endpoint, J_{MAX}, corresponds to the maximal rate of absorption and it is analogous to the C_{MAX} endpoint attained in traditional plasma pharmacokinetics. The second endpoint regards AMT, which is derived from equation (1) and can be compared to the area under the curve (AUC) of the incremental bifonazole permeation profile (Leal et al., 2017).

To assess the risk of interference from the biological matrix or dosage form, Franz cells containing non-dosed skin and a placebo formulation were likewise considered.

2.2.1.3. Mass balance studies. After IVPT experiments, mass balance studies were conducted to assess the amount of drug remaining on the donor compartment, on the skin and delivered into the skin (EMA, 2018; Hossain et al., 2019; OECD, 2010). At the end of the IVPT runs, the donor compartments were washed with 1 mL of acetonitrile and the respective washing solutions were collected. Afterward, the skin was transferred into eppendorf® tubes and the remaining bifonazole was likewise extracted with acetonitrile. All samples were sonicated (10 min), centrifuged at 11 740 x *g* for 10 min in a Minispin®(Eppendorf Ibérica S.L., Madrid, Spain), filtered by a 0.45 μm nylon membrane and transferred to HPLC vials for analysis.

The total recovery of drug at the end of the IVPT experiment was calculated by considering the formulation mass initially applied in the donor chamber ($m_{applied}$), and the sum of the final cumulative amount of bifonazole that permeates the biological membrane into the receptor chamber ($m_{permeated}$), the drug extracted from the formulation remaining in the donor chamber (m_{donor}) and the drug extracted from the biological membrane (m_{skin}) at the end of the experiments. The mass balance was then calculated according to the equation (2):

$$mass\ balance = \frac{m\ donor + m\ permeated\ + m\ skin}{m\ applied} \times 100 \quad (2)$$

Procedure reliability was confirmed by the total active ingredient recovery (%), which should be ideally within the range of 100 ± 10 % (EMA, 2018).

2.2.2. HPLC analysis and method validation

A Shimadzu LC-10AD apparatus, equipped with a quaternary pump (LC-10AD), an autosampler unit (SIL-10ADVP), a CTO-10AVP oven, and a CBM-20 A detector was used. All analyses were conducted in isocratic mode with a 40 °C temperature. A XBridge™ C18 5 μm (2.1 x 150 mm) column was used. The mobile phase consisted of a buffer solution (900 mL of a sodium dihydrogen phosphate solution (29 mM) with 100 mL of an orthophosphoric acid solution (25 mM), adjusted to pH 3.2 using trimethylamine) and acetonitrile (60:40, v/v). The flow rate was set to 0.35 mL/min and the run time was 8 min. The detection wavelength was 210 nm.

Analytical method validation was performed according to the International Council on Harmonization guidelines (CPMP/ICH/381/95, 2005). All experimental conditions are addressed in the supplementary material (Table S2-S6 and Figure S1).

2.2.3. Disease model investigation

2.2.3.1. Preparation of the trichophyton rubrum suspension. The disease model utilized a *T. rubrum* strain isolated from a patient suffering from onychomycosis. This strain was continuously subcultured to ensure a fresh stock plate was available for each experiment. Dermatophytes were transferred into Ringer's solution, spread onto a potato dextrose agar plate and incubated for 7 days at 25 °C to produce conidia.

Fresh *T. rubrum* organism suspension was prepared in sterile 0.1 % Tween 80 in Ringer's solution for each experimental set up, with the Tween 80 added to attempt to reduce conidial aggregation. The suspension was filtered through sterile gauze (Boots Sterile Gauze Swabs, 7.5 cm x 7.5 cm) to remove the mycelium. A 1:10 dilution of the suspension was then measured in the UV spectrophotometer (ThermoFisher Scientific Evolution 201) at 600 nm to obtain an optical density. The suspension was accepted if the optical density was between 0.7–0.9 Abs, if it fell outside this range it was appropriately diluted or more organisms added as required. The suspension was then centrifuged and the pellet resuspended in 0.1 % Tween 80 in Ringer's solution.

2.2.3.2. Infected skin studies. A novel infected skin disease model developed by MedPharm Ltd was employed. In this disease model, ATP levels were used as a biological marker for antimicrobial activity across human skin. Skin prepared as above was removed from the freezer and allowed to defrost at room temperature for 30 min. The skin was cut into sections with a surface area of approximately 0.5 cm^2, removed from the filter paper and mounted on sterile PTFE septa with the *stratum corneum* side facing upward. The skin was sterilised by application of 50 % ethanol and allowed to dry under the biological safety cabinet. The *stratum corneum* was inoculated with *T. Rubrum* (7.5 μL, using a positive displacement pipette) and allowed to dry prior to being mounted into the ChubTur® cells (*stratum corneum* facing down). A small amount of sterile Ringer's solution was added to the lower half of the ChubTur® cells and sterile water (20 μL) was applied to the epidermal side of the skin to aid with hydration. The mounted ChubTur® cells were incubated at 32 °C for 24 h before dosing the *stratum corneum* with the selected formulation (2.5 μL, using a positive displacement pipette). Following dosing, the mounted ChubTur® cells were incubated for an additional 24 h under the same prior conditions. The activity of the formulation was assessed using a previously validated bioluminescence ATP method.

2.2.3.3. ATP assay and sensitivity. The solution utilised in the ATP assay

was a mixture of the Promega BacTiter-glo kit and the MedPharm lysing agent, as used by Turner et al. (Turner et al., 2016). Wherever possible, the lysing solution was protected from light with foil, to prevent potential degradation from light.

ATP standards were prepared at known concentrations (1.1, 3.4375, 6.875, 11, 13.75, 27.5, 55, 110 and 220 ng/mL) by performing a range of dilutions of the stock ATP standard (1 mg/mL) in Ringer's solution. These standards were then plated in a 96 well white micro-titre plate (100 μL). The BioTek FLx800 fluorometer/luminometer was then used to dispense the lysing solution (100 μL) into the wells, and the light emitted from the well was read at 10 s and 5 min after dispensing. This data was used to produce a standard curve, demonstrating that the lysing solution was able to differentiate between ATP concentrations and thus was fit for purpose. An investigation into the potential quenching effect of the formulation was considered and no effect from the formulation was seen on the ATP assay.

In an attempt to validate the disease model, the sensitivity of the model was investigated. This was performed using the original Canesten Extra® formulation, a 50 % strength active formulation and a placebo. While performing this experiment, a significant placebo effect was observed which necessitated the inclusion of alternative controls.

2.2.4. Data analysis and statistics

2.2.4.1. IVPT data.
According to the FDA and EMA draft guidelines, to infer on efficacy equivalence, the 90 % confidence intervals (CI) for the ratio of means between the formulations being compared, should be determined for both IVPT endpoints – J_{MAX} and AMT (EMA, 2018; FDA, 2022). Furthermore, to document the discriminatory capacity of the IVPT method, the 90 % CI should also be presented for the target formulation and the control.

These intervals were determined following two approaches – EMA and FDA. Both approaches recommend a paired comparison. In the EMA approach, the data was natural log transformed. Then, the arithmetic mean of all individual T-R differences was calculated. Subsequently, the variability within subjects was calculated as the difference between each individual subject difference T-R and the previously determined mean. These squared differences were summed to obtain the sum of squares. The sum of squares was divided by n-1 degrees of freedom to obtain the variance of the differences. The standard error of the differences was obtained by dividing the variance by n and then calculating the square root. The CI was attained by the usual expression:

$$\overline{X} \pm t \times \frac{s}{\sqrt{n}} \qquad (3)$$

Wherein $\overline{X}$ is the previously calculated mean of all individual T-R differences, *t* is the t-value reporting to a 90 % CI with n-1 degrees of freedom, s regards the standard error calculated as previously described, and finally, n is the sample size.

To follow the FDA approach, the evaluation of BE was based on the natural log transformed total amount penetrated (AMT) and the maximum flux rate (J_{MAX}).

For the AMT and J_{MAX} CI calculations, the within-reference standard deviation (S_{WR}) was evaluated from the data as

$$S_{WR} = \sqrt{\frac{\sum_{j=I}^{n} \sum_{i=I}^{r} (R_{ij} - \overline{R_j})^2}{(r - I)n}} \qquad (4)$$

where n is the number of donors, r is the number of replicates, I is the donor average, R_{ij} is the observation from the ith replicate and the jth donor from the RP, $\overline{R_j}$ is the mean of the jth donor for the reference formulation.

The obtained value for S_{WR} was then used as a cut-off such that:

- For $S_{WR} \leq 0.294$, the test and reference formulations are declared bioequivalent if the (1-α) 100 % two-tailed confidence interval $\overline{X} \pm t_{(n-1),\,\alpha/2} \times \sqrt{\frac{s_L^2}{n}}$ is contained within the limits [1/m, m] (Pensado et al., 2019).
- When $S_{WR} > 0.294$, a scaled criterion is used. This is a similar approach to that used by the FDA for analysis of highly variable drugs, modified for the particular design. The hypotheses to be tested are:

$$H_0 : \frac{(\mu_T - \mu_R)^2}{\sigma_{WR}^2} > \theta.$$

$$H_a : \frac{(\mu_T - \mu_R)^2}{\sigma_{WR}^2} \leq \theta.$$

Where μ_T and μ_R are the population means of the test and reference formulations, respectively, θ corresponds to $\frac{(\ln(m))^2}{0.25^2}$ and m represents the choice of the bioequivalence limit. The two products are declared bioequivalent if the upper bound of the confidence interval for the quantity (SCI_{UB}), $(\mu_T - \mu_R)^2 - \theta \sigma_{WR}^2$, is less than or equal to zero. This criterion imposes the additional constraint that the point estimate must lie within the limits [1/m, m]. Rejection of the null hypothesis supports bioequivalence of the test and reference products (FDA, 2022).

Example calculations of the bioequivalence evaluations for AMT are provided in the supplementary material (Tables S7-S10).

Data analysis was conducted using Microsoft Office Excel®.

2.2.4.2. Disease model.
Initially, the data for each treatment group from the disease model investigation was tested for a normal distribution using a quantile–quantile plot (QQ plot) and the Shapiro-Wilk test. Two tests were chosen for this step to improve confidence in the data normality. The Shapiro-Wilk test was carried out to a 95 % confidence level. From these tests, it was found that the unprocessed data from the disease model was not normally distributed, which would be required for the proposed post-hoc tests. Therefore, the data was natural log transformed, and the normality tests repeated. It was found that the log transformed data appeared normally distributed on the QQ plot with the treatment groups having *p*-values > 0.05 in the Shapiro-Wilk test. These tests were performed using JMP®.

A Dixon's Q test was then executed on the log transformed data, grouped by experimental set up. This was performed on the log transformed data rather than the unprocessed data as the Dixon's Q test assumes that data is normally distributed. The Dixon's Q test was chosen for this analysis due to the small data set generated, as well as its ability to easily highlight individual data points as an outlier. As the number of data points between groups varied in size, multiple variations of the Dixon's formula were used; r_{10} for 10 or fewer replicates using a critical value of 0.560 for n = 6 and 0.412 for n = 10, r_{21} for 12 replicates using a critical value of 0.489 and r_{22} for 15 replicates using a critical value of 0.470. Each Dixon's Q test was carried out to a 90 % confidence level and the analysis showed that no individual data point was deemed an outlier and therefore, all data points were included in further tests. This test was performed using Microsoft Office Excel®.

Finally, Tukey's honestly significant difference (HSD) test was used to determine whether there was a statistically significant difference between treatment groups. Tukey's HSD was chosen as alternative one-way tests performed sequentially would accumulate errors due to the multiple pairings required. The Tukey-Kramer variation was used on the log transformed data pooled from all experimental set ups. This variation was used to address the treatment groups having differing sample sizes. The log transformed data was tested as the Tukey HSD test assumes a normal distribution. The analysis was performed at a 95 % confidence level. This test was performed using JMP®.

3. Results and discussion

3.1. Product efficacy profile – IVPT kinetic studies

The test guideline 428 by the Organization for Economic Cooperation and Development (OECD), the EMA draft guideline, the FDA In Vitro Permeation Test Studies for Topical Drug Products Submitted in ANDAs Guidance for Industry, the FDA product specific guidances for generic drug development, as well as the recently revised < 1724 > USP chapter on Semisolid Drug Products — Performance Tests, detail the methodologies that can be used to infer on the *in vivo* absorption of topical products, such as IVPT (EMA, 2018; FDA, 2022; Jin et al., 2022; OECD, 2010; USP, 2023). IVPT tests can be used to predict the *in vivo* absorption of topical products, as they quantify the drug flux over time using a suitable *ex vivo* human skin membrane (Jin et al., 2022). As the *stratum corneum* is the primary limiting barrier to dermal absorption, the determination of the IVPT profile using human skin closely resembles *in vivo* conditions. These reasons further ground the increasingly regulatory significance of this methodology (Abd et al., 2016; Franz, 1975; Leal et al., 2017).

Bifonazole cream formulations were selected as a case study for the present work due to previous works by our group, in which the variability of bifonazole 10 mg/g cream RP formulations was described. Briefly, an initial study involved the assessment of the microstructure characteristics (globule size, pH, and rheological attributes) of three batches from the same RP. Then, product performance, evaluated by means of IVRT, was likewise studied (Miranda et al., 2020b). The results showed that there were marked inter-batch differences between globule size and rheology profile, with the batch with lower globule size revealing a lower viscosity. Regarding the IVRT profile, even though the release profiles were similar between all batches, when applying the EMA draft guideline criteria, solely one batch-to-batch combination would yield regulatory compliant results (Miranda et al., 2020b). These experimental observations laid the ground for a second study (Miranda et al., 2023a). The primary objective of the latter was to establish a framework to comprehensively address the several variability scenarios that may occur in daily practice. In this study, the initial three RP batches were complemented with two extra RP batches, yielding a total of 5 batches. These were considered together with a Q1/Q2 equivalent formulation (TP), a Q1 equivalent formulation (CPA), and finally, a bifonazole cream formulation with Q1 and Q2 differences (CPB). Focusing on the RP characteristics, the results from this work continued to highlight the pronounced rheological variability between batches. Even when comparing the most similar RP batches in terms of viscosity, the statistical comparison continued to reveal uncompliant results according to the EMA draft guideline, as there was a difference of more than 10 % between the rheological endpoints attained with these RP batches, combined with a lack of compliance with the 90–111 % confidence interval (Miranda et al., 2023a). These rheological differences motivated the determination of a drug release profile for each RP batch, which again, revealed statistical differences for some batch-to-batch combinations. Even though the performance differences were not as distinct as the rheological ones, according to the EMA draft guideline 90–111 % equivalence interval acceptance criteria, overall equivalence between RP batches failed to be supported (Miranda et al., 2023a). These experimental observations led to a fundamental question: How do these differences in microstructure and performance affect the permeation profile? From a generic manufacturer point of view, addressing this question is essential to properly identify the RP batch to use in human permeation kinetic studies. In the present work, IVPT studies were carried out using two RP batches with different viscosity, with RP1 being less viscous than RP4 (Miranda et al., 2023a). The computation of the *in vitro* release rate (IVRR) 90 % CI for these batches yielded borderline, but still uncompliant results following EMA draft guideline directives (97.0–113.1 %).

When developing an IVPT method, there are several parameters that

need to be addressed: (i) human membrane characteristics, membrane preparation techniques, skin integrity evaluation methods, and respective acceptance criteria; (ii) Choice of receptor medium, which should comply with sink conditions. Although the FDA strongly discourages the use of organic solvents, according to the EMA draft guideline, their use may be justified, if the skin integrity is not compromised. In this work, due to limited solubility of bifonazole, a PBS-PEG (60:40, v/v, pH = 7.4) solution was used as a permeation medium. The solubility of bifonazole in this medium is 3.62 mg/mL, and the highest concentration of the API did not exceed 1/10 of this value; (iii) Selection of suitable sampling points regimen, capable of presenting a meaningful permeation profile; (iv) Selection and description of formulation dosing techniques. IVPT studies should be performed under finite dose conditions and a homogeneous spreading of the product over the skin should be ensured and finally; (v) Other parameters should also be verified such as the absence of contamination and/or interferences, randomization and blinding procedures following ICH E8 criterion, validation of suitable analytical procedures for drug quantification, documentation of API stability over the IVPT study timeframe, as well as mass balance studies.

Following IVPT method development studies, a pilot study should be performed to further confirm the suitability of method parameters. In this study, both RP batches were used, as well as a control formulation. This formulation regarded a 5 mg/g bifonazole cream. The inclusion of a different strength product aims to document the method sensitivity and discrimination.

The IVPT pilot study results are summarized in Fig. 1 and Table S11.

In this preliminary assessment, skin from 3 different donors was used. Two replicates per formulation were used in each donor, providing a total of 6 replicates per formulation. Due to the limited sample size, no statistical analysis of the pilot study was performed. Pooling all the data, IVPT pilot study results revealed that RP4 presented a superior permeation, followed by RP1 and the control formulation (5 mg/g bifonazole cream).

Pilot studies revealed that experimental procedures adequately described the cutaneous pharmacokinetics of bifonazole since the maximal rate of absorption is achieved, followed by a decrease. Furthermore, the developed method proved to be sensitive, being able to detect changes as a function of differences in drug delivery. Regarding the skin integrity results after IVPT experiments, all membranes were checked for leakage and none was observed. Although values of more than 20 g/m^2/h in TEWL were obtained in some diffusion cells, these results were borderline and did not correspond to higher AMT
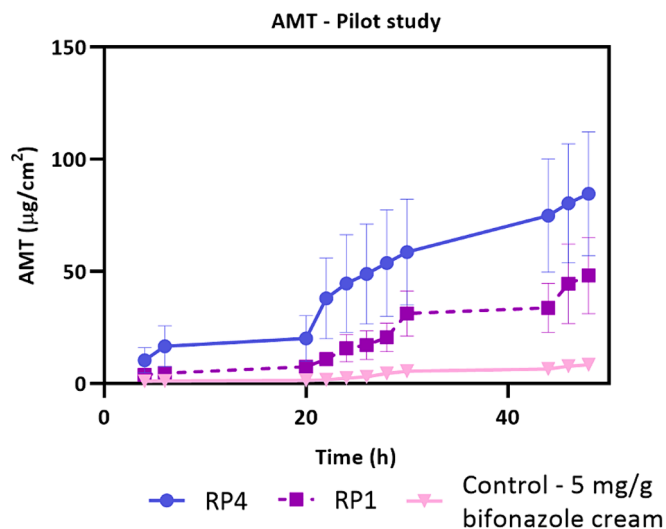

Fig. 1. Overall permeation profiles for bifonazole in pilot IVPT studies. All results report to mean ± SEM (n = 2, meaning 2 replicates per donor, 3 donors were considered in IVPT pilot studies).

(numerical data not shown). In this context, the usage of PEG as a cosolvent did not affect the integrity of skin barrier throughout the study timeframe. Mass balance results, presented in Table S11, were overall compliant with the established 90–110 % criteria with the exception of donor 1 (2nd replicate of RP1, and both replicates for RP4), as well as for donor 2 (a single replicate of RP1 and RP4).

The results produced from the pilot study demonstrated that the method was fit for the proposed purpose. Therefore, these conditions were adopted for the IVPT pivotal study.

The comparative pivotal permeation and flux profiles of the bifonazole cream formulations are presented in Fig. 2 and in Table 3.

As previously mentioned, a larger pool of skin donors was employed in pivotal studies to account for inter-donor variability. The final countdown for donors used in the present study was 11, with 3 donors employed in pilot studies and 8 in pivotal studies. Similarly to the pilot study design, 2 replicates per formulation per skin were always considered.

Fig. 2 shows that there are clear differences between both RP batches, with RP1 revealing a higher permeation when compared to RP4. This data is opposite to the one acquired in pilot studies. As can be retrieved from Table 3, donor 4 exhibited a 2.2-fold higher permeation of RP4, whilst donor 6 exhibited an even higher permeation discrepancy. For all other donors, there was generally a good agreement between both formulations, even though the majority of the donors exhibited a higher permeation of RP1. This can be ascribed to the reduced number of donors used in the pilot study, as well as to the fact that solely two replicates were considered per donor. Moreover, the inter-individual variability may have also contributed to these discrepancies.

Furthermore, the control formulation, that presented half of the

Table 3

Results from the bifonazole IVPT pivotal studies. The data (arithmetic mean $\pm$ SEM) were obtained from 8 donors and two replicates per donor were always considered.

Product	Skin (n)	AMT ($\mu g/cm^2$)	J_{MAX} ($\mu g/cm^2/h$)
RP1	Donor 4 (2)	21.87 ± 8.45	2.66 ± 0.59
	Donor 5 (2)	43.89 ± 20.59	5.00 ± 2.49
	Donor 6 (2)	97.49 ± 64.66	9.91 ± 7.22
	Donor 7 (2)	53.32 ± 1.77	6.15 ± 1.45
	Donor 8 (2)	82.96 ± 61.22	17.52 ± 15.52
	Donor 9 (2)	116.00 ± 50.82	7.66 ± 2.34
	Donor 10 (2)	50.33 ± 30.99	7.36 ± 4.33
	Donor 11 (2)	53.39 ± 30.47	5.69 ± 3.66
	Combined (16)	**64.91 ± 12.80**	**7.75 ± 2.00**
RP4	Donor 4 (2)	48.25 ± 24.20	8.62 ± 2.39
	Donor 5 (2)	2.53 ± 0.81	0.36 ± 0.18
	Donor 6 (2)	5.00 ± 2.23	0.50 ± 0.27
	Donor 7 (2)	35.88 ± 15.67	5.31 ± 0.61
	Donor 8 (2)	58.82 ± 28.58	8.81 ± 2.84
	Donor 9 (2)	68.27 ± 52.38	8.89 ± 5.95
	Donor 10 (2)	44.72 ± 1.51	5.61 ± 0.42
	Donor 11 (2)	30.68 ± 22.12	5.84 ± 5.25
	Combined (16)	**36.77 ± 8.55**	**5.49 ± 1.16**

bifonazole strength, exhibited a decrease in drug delivery. According to the *in vitro* permeation test studies for topical products submitted in ANDAs – FDA draft guidance, the modulation of topical product strength to support IVPT method sensitivity may not be a suitable strategy for some formulations, as it may not consistently reproduce the expected increase/decrease in drug delivery (US-FDA, 2022a). However, as displayed in Fig. 2, in the present case study, the lower strength product promoted a decrease in bifonazole cutaneous pharmacokinetics, relative
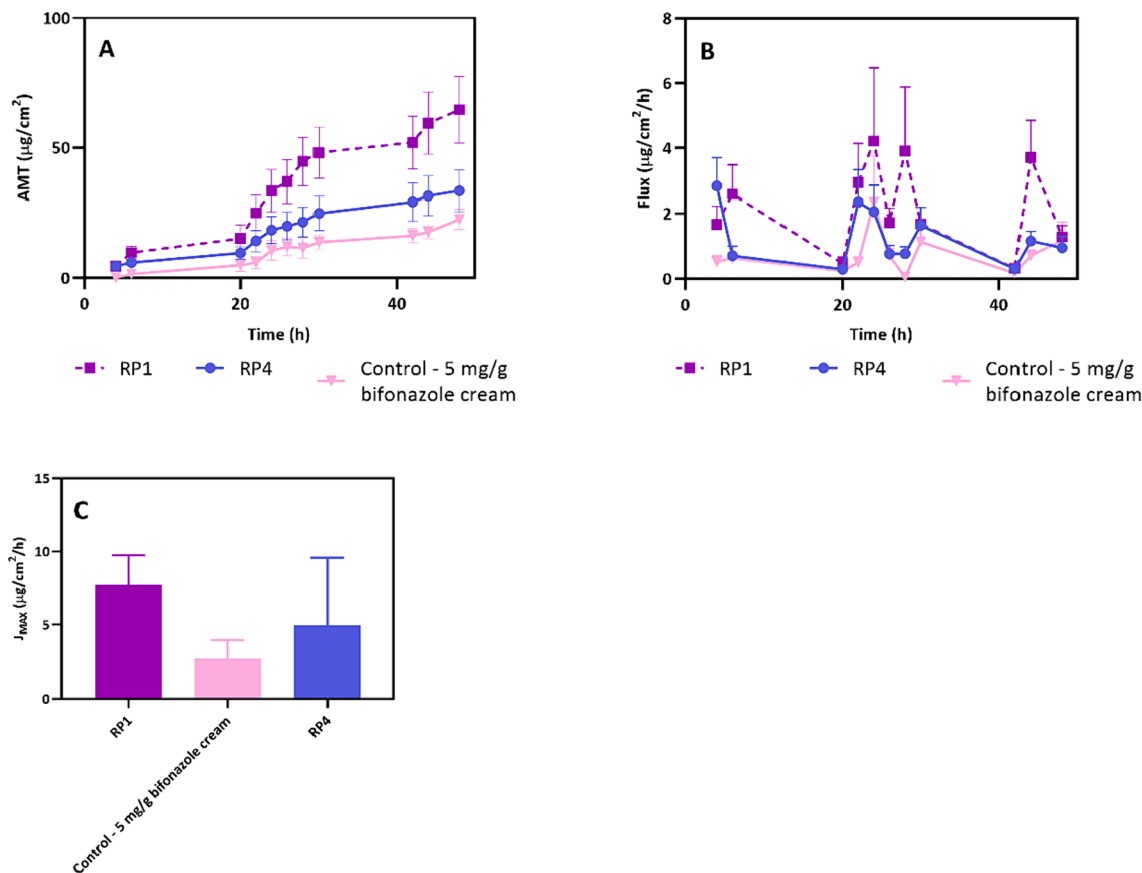

Fig. 2. A – Permeation profiles for all tested formulations in pivotal IVPT studies. B – Flux profiles attained during IVPT pivotal studies. C – Maximum flux attained during IVPT pivotal studies. Results report to the mean $\pm$ SEM calculated from duplicate sites from the same donor. RP1 = 8 donors; RP4 = 8 donors; Control formulation = 4 donors. Two replicates per donor were always considered.

to that obtained from the nominal strength product (US-FDA, 2022a). Taking this into account, the IVPT method selectivity was inferred, as the methodology was selective for differences in drug delivery (US-FDA, 2022a).

A thorough analysis of the pivotal data suggests that a potential correlation between the permeation behaviour and the rheological and performance characteristics of the products may be established (please see Fig. 3).

The rheological behaviour is a critical quality attribute of a semi-solid, as its biopharmaceutical characteristics, such as drug release and permeation, are reliant on the rheological profile (Simões et al., 2020; Sivaraman et al., 2017; Soriano-Ruiz et al., 2019).

When comparing both bifonazole RP batches, in the higher viscosity system (RP4), drug permeation was hampered, which in turn affected the skin bioavailability of the product. On the other hand, RP1 revealed an opposite trend, with higher permeation and lower viscosity. Note, however, that IVRT performance results did not reflect these differences. For example, a work by Binder *et al.* described the correlation between viscosity and drug penetration. The authors produced sulphadiazine sodium hydrogels with different rheological profiles. Then, *in vitro* skin penetration was monitored by dermatopharmacokinetic methods, combined with non-invasive confocal Raman spectroscopy (Binder et al., 2019). The authors performed these experiments in full-thickness porcine ear skin. Even though the results showed that the drug penetration was largely unaffected by hydrogel viscosity, the authors observed that drug penetration depth slightly decreased with an increase in viscosity, suggesting a slower drug permeation due to the increasingly dense gel networks (Binder et al., 2019). A study by Tanja Ilic *et al.* performed a comparative assessment of selected CQAs and *in vitro/in vivo* product performances, with distinct aceclofenac formulations (Ili and Daniels, 2017). The main rationale of this work was to document the ability of "ready-to-use" topical vehicles based on alkyl polyglucoside-mixed emulsifier (with/without co-solvent modifications), instead of pharmacopoeial bases, such as the non-ionic hydrophilic cream. The authors observed that the more viscous formulation showed the lowest aceclofenac permeation (Ili and Daniels, 2017).

According to OECD and EMA draft guidelines, mass balance studies should be conducted after the IVPT experiments to assess the amount of drug remaining on the donor compartment, on the skin and delivered into the skin (EMA, 2018; Hossain et al., 2019; OECD, 2010). The total recovery of drug at the end of the IVPT experiment should be calculated by considering the mass of formulation initially applied to the donor chamber ($m_{applied}$) and the sum of the final cumulative amount of drug that permeates the biological membrane into the receptor chamber (AMT), the drug extracted from the formulation remaining in the donor chamber (m_{donor}) and the drug extracted from the biological membrane (m_{skin}) at the end of the experiments. Procedure reliability should be confirmed by the total drug recovery (%), which, according to EMA draft guideline requirements should be within the 100 ± 10 % range (EMA, 2018).

Mass balance results, depicted in Table 4, were generally compliant with the established criteria. However, for some skin pieces, the extraction procedures did not meet the prescribed values. The evaluation of drug concentration in heat-isolated epidermis samples is challenging, due to several reasons: (i) interference coming from the biological matrix; (ii) difficulty in manipulating the skin sheets. The thermal separation of the skin leads to the attainment of a very thin layer, which may prove to be challenging to manipulate, in contrast to dermatomed or full thickness skin. Nevertheless, when using heat isolating epidermis separation technique, the *stratum corneum* is kept integral, and by using a thin skin layer, the chances of occurring an artificial delay of drug permeation are lessened (EMA, 2018); (iii) need for a very sensitive analytical method and; (iv) need to perform IVPT studies under finite dose conditions. These reasons, combined with the
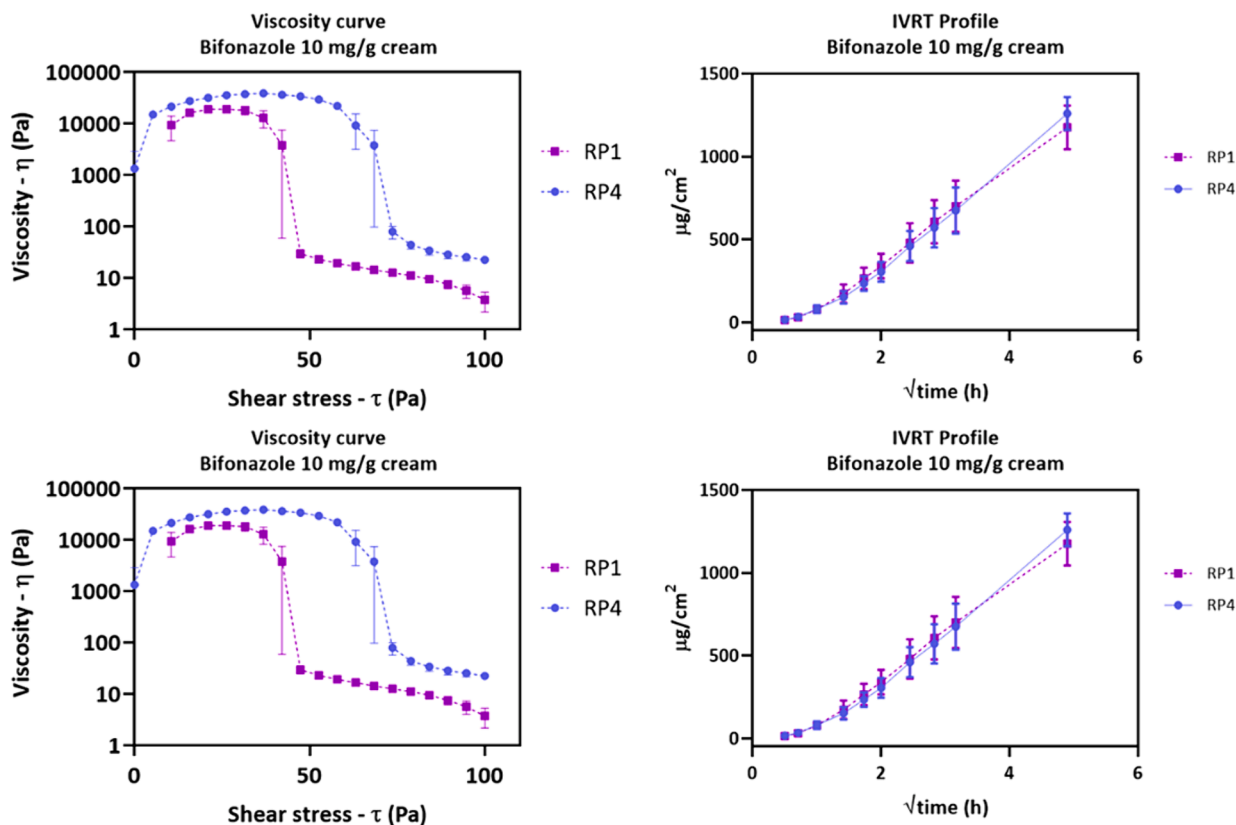

Fig. 3. Viscosity curve of bifonazole cream formulations. Results report to n = 3 (mean ± SEM). IVRT profiles of bifonazole cream formulations. Results report to n = 12 (mean ± SEM). Data retrieved with permission from (Miranda et al., 2023a).

Table 4
Skin integrity and mass balance results for bifonazole 10 mg/g cream pivotal study results.

Donors	4	4	5	5	6	6	7	7	8	8	9	9	10	10	11	11
RP1																
Initial TEWL (g/m^2/h)	0.00	6.70	0.00	6.90	2.50	6.90	9.90	4.20	0.00	0.00	0.00	0.80	2.90	2.30	0.00	6.50
Final TEWL (g/m^2/h)	10.4	13.4	15.2	12.0	15.8	6.60	21.0	14.3	15.9	14.3	4.30	5.60	8.60	15.2	21.0	14.3
Donor compartment (µg)	31.7	30.5	23.7	8.6	0.94	76.7	15.0	55.3	77.9	2.25	15.5	0.24	22.3	32.7	12.6	62.6
Skin (µg)	30.7	45.7	44.8	7.9	0.52	7.75	19.1	7.71	9.46	5.59	35.7	0.38	25.7	47.0	3.36	25.3
Mass Balance (%)	103	103	87.3	62.9	113	111	93.5	106	103	105	100	111	111	93.4	83.9	108
RP4																
Initial TEWL (g/m^2/h)	7.00	7.00	8.6	17.4	8.90	8.60	9.90	2.60	2.50	14.2	4.40	5.10	2.60	9.20	4.20	4.70
Final TEWL (g/m^2/h)	12.9	11.5	0.00	9.00	11.8	15.6	21	20.6	13.6	24.6	4.10	8.60	4.30	5.00	0.00	9.30
Donor compartment (µg)	41.15	63.77	16.9	12.8	12.4	18.3	15.0	64.2	7.72	23.0	2.13	61.8	6.38	23.0	23.6	48.4
Skin (µg)	5.81	12.33	53.3	67.0	51.2	31.0	19.1	11.6	11.8	41.3	7.31	16.4	5.08	16.4	33.5	13.1
Mass Balance (%)	112	110	90.0	90.2	91.7	68.7	93.5	110	111	103	96.2	95.6	52.7	102	73.3	110
Control formulation – 5 mg/g bifonazole cream																
Initial TEWL (g/m^2/h)	14.5	12.1	0.00	18.4	27.3	22.1	22.4	20.4	10.1	9.20	4.40	5.10	6.60	9.20	4.20	4.70
Final TEWL (g/m^2/h)	23.8	20.5	9.90	6.70	8.00	0.00	17.8	26.9	14.2	10.4	26.2	19	12.3	40.2	8.60	7.90
Donor compartment (µg)	N.D	N.D	N.D	N.D	N.D	N.D	N.D	N.D	N.D	N.D	N.D	N.D	11.9	N.D	10.1	9.42
Skin (µg)	N.D	N.D	N.D	N.D	N.D	N.D	N.D	N.D	N.D	N.D	N.D	N.D	14.8	N.D	18.7	19.3
Mass Balance (%)	N.D	N.D	N.D	N.D	N.D	N.D	N.D	N.D	N.D	N.D	N.D	N.D	112	N.D	94.4	99.7

Key: Green label – Compliant results; Red label – Non-compliant results. N.D. – Not determined. To interpret the color references in the table caption, the reader is referred to the web version of this article.

strict 90–110 % recovery criteria by EMA draft guideline, contributed to this occurrence (Demurtas et al., 2020; EMA, 2018).

Overall, mass balance results were in accordance with the IVPT pivotal data, as a superior bifonazole skin retention was observed in RP4, the most viscous batch (24.76 µg), when compared to RP1 (19.79 µg), the less viscous formulation. This behavior may render to a preferential surface therapeutic effect rather than that prompted by in-depth skin penetration.

Regarding TEWL results, also presented in Table 4, these were compliant with the 20 g/m^2/h threshold. Nevertheless, some borderline results were registered at the end of the IVPT experiment in some skin pieces. These, however, were not correlated with an enhanced permeation rate. Moreover, no leaks were observed in the excised human membranes.

The statistical analysis of the IVPT pivotal test is summarized in Table 5. As IVPT data does not follow a normal distribution, it should be log transformed prior to any calculation. Two statistical approaches were considered to analyze the permeation results: the EMA and the FDA. Both approaches advise a paired comparison, in which the differences between TP and RP, in permeation endpoints, should be individually calculated for each donor.

In the European approach, the variability within subjects is calculated as the difference between each individual T-R and arithmetic mean of the two replicates per donor, per formulation. On the other hand, the FDA follows a scaled average bioequivalence approach (SABE). A SABE analysis attempts to standardize the difference due to the observed variability in the reference product. To be applicable, the within-subject standard deviation (S_{WR}), calculated for each IVPT endpoint attained with the RP formulation, should be higher than 0.294 (FDA, 2022; Pensado et al., 2019). According to this approach, bioequivalence can then be inferred if the geometric mean ratio (GMR) falls within the range [0.8–1.25] for the selected bioequivalence margin and if the upper bound of the 90 % confidence interval (SClUB) for the quantity, (μ_T – μ_R)^2 – σ^2_{WR} (ln(1.25)/0.25)^2, is less than or equal to zero. μ_T and μ_R regard the population means of the test and reference products, respectively, and σ^2_{WR} refers to the reference population variance (Pensado

Table 5

90 % confidence interval calculated for J_{MAX} (µg/cm^2/h) and AMT (µg/cm^2) at the end of the permeation experiment (48 h) for bifonazole cream formulations following EMA and FDA approaches.

Approach used	Pairwise comparison	J_{MAX} (µg/cm^2/h)	Status	AMT (µg/cm^2)	Status	Acceptance criteria
EMA	RP1 vs RP4	38.43 – 73.57	NC	31.07 – 57.25	NC	90% CI = [80-125%]. A wider 90% confidence interval limit to a maximum of [69.84 – 143.19] may be accepted (EMA, 2018).
FDA	RP1 vs RP4	S_{WR} = 1.078 SCI_{UB} = 1.9888 GMR = 0.4512	NC	S_{WR} = 0.925 SCI_{UB} = 2.5677 GMR = 0.3919	NC	SCI_{UB} < 0 GMR ∈ [0.8 — 1.25]

Key: J_{MAX} – Maximal flux; AMT – Cumulative drug amount permeated at the end of the IVPT study. RP1 vs RP4 = 8 donors. In the EMA approach, the 90 % CI were calculated based on the geometric mean of the duplicate values obtained per donor. In the FDA approach: SCIUB – upper bound of the 90 % confidence interval; AMT and JMAX are reported as the anti-logarithm of the arithmetic mean (lower–upper 90 % confidence interval) of the natural log-transformed values; Green label – Compliant results; Red label – non-compliant (NC) results. To interpret the color references in the table caption, the reader is referred to the web version of this article.

et al., 2019). Example calculations are provided in the supplementary material (Tables S7-S10).

According to EMA draft guideline, a wider 90 % CI, up to a maximum of 69.84 – 143.19, may be accepted when high variability is observed with low strength and limited diffusion drug products. Nevertheless, even when considering this broader acceptance criteria, the RP product comparisons continue to fail to document product permeation equivalence. This scenario is also observed when addressing the FDA approach. A limitation of this study regards the sample size. A paper by Tothfalusi et al. demonstrated that for highly variable drugs in pharmacokinetic studies, the sample size to use should be established based on the within-subject variability. Since larger absolute differences between the two logarithmic means are expected to occur, it is recommended that a maximum 10 % deviation between the means (e.g. GMR = 1.10) should be considered (Tothfalusi and Endrenyi, 2011). Considering these premises, in order to achieve a 90 % power, and based on the variability encountered for the RP permeation data, the authors advice on a sample size of 85 donors, in order to comply with EMA. On the other hand, to fulfil with FDA requirements, the donor sample size should have been 68 (Tothfalusi and Endrenyi, 2011). The conduct of such a trial would be impractical, if not impossible. However, another strategy that could have been used during the IVPT study design, was the inclusion of more skin replicates per donor, instead of using a larger pool of donors, as detailed in the recent in vitro permeation test studies for topical products submitted in ANDAs – FDA draft guidance (US-FDA, 2022a). Furthermore, according to the same guidance, the usage of skin retrieved from the same anatomical region, would have provided a better IVPT study control.

In fact, the anatomical regions display distinctive permeation characteristics. A study by Bormann and Maibach acknowledges the regional variation in percutaneous absorption (Bormann and Maibach, 2020). The authors detail that in general the head, neck, and genital regions appear more absorptive than other body regions. On the other hand, the abdomen, chest, back and thighs proved to be less absorptive. These differences may be explained by several parameters such as skin anatomy, stratum corneum thickness, sebum production, proximity of vascular blood supply to the cutaneous surface, as well as appendage density (Bormann and Maibach, 2020). This last parameter, should be closely considered, as the storage reservoir capacity of the hair follicle is 10-fold longer than that of stratum corneum, and that the follicular route provides a pathway for topical drug absorption, in addition to the intercellular and intracellular routes (Akomeah et al., 2007; Feschuk et al., 2022). Also, when considering the anatomical variation in terms drug permeation, a study by Schmitt et al. (2010) tested the permeability of several compounds present in rose oil, in human skin samples. The authors aimed to compare the permeability differences of such

compounds, when applied to the upper arm, breast, and abdomen. These anatomical regions were also considered in the present study, as depicted in Table 2. The authors were able to denote permeation differences, however these appeared to be dependent upon the molecule – i.e. the permeability of abdominal skin was greater than that of upper arm, for linalool and eugenol, whilst in the upper arm, the skin appeared to be more permeable for β-myrcene, trans-rose oxide and cis-rose oxide (Schmitt et al., 2010). In the present work, we were not able to denote a trend of increased/decreased permeability according to the anatomical region. For RP1 the lowest permeability was attained with donor 4, which was skin from the arm. However, in RP4 the scenario was different, with donors 5 and 6 (respectively glute and abdomen) displaying the lowest permeation of bifonazole. This lack of correlation with the previously referred study may be ascribed to the fact that a complex formulation – a multiphasic cream – is being evaluated, rather than an isolated compound. Therefore, the permeant displays a more complex physicochemical profile that is likely to affect permeation, more than the anatomical variation of the skin. Furthermore, the overall conclusion also is likely to have been affected by the small but more practically relevant number, of donors studied - solely 8 in the pivotal studies, from which only skin from 4 different body sites was considered.

The age of a person also affects topical drug permeation. As the individual grows older there are several physiological constraints that contribute to a reduction of skin permeability. Firstly, there is an average increase of the size of stratum corneum corneocytes throughout their lifecycle. The dehydration of the outer layers of the stratum corneum, decreased epidermal turnover and decreased microvascular clearance are also contributing factors (Benson and Watkinson, 2012). These events should also be linked with the physicochemical characteristics of the drug being studied – bifonazole. This is a lipophilic drug with a log P of 4.8. In such circumstances, the increased lipid content and reduced hydration status of the stratum corneum in older skin donors may improve the permeation through this superficial layer, however, the passage into and through the remaining layers of the epidermis becomes rate limiting, therefore the permeation is lower (Poet and McDougal, 2002).

Another physiological alteration in the skin of older individuals regards the reduction in the skin blood flow. This event directly impacts the drug clearance rate from the skin, thereby affecting the drug penetration and permeation rate. These physiological alterations are extremely important in in vivo assessments, however when dealing with excised human skin, the literature points out that differences in age, even though important, are not the main drivers for IVPT variability (Benson and Watkinson, 2012). According to our results, when considering both formulations, the youngest skin donor in the pivotal studies (donor 6) was not the one with higher permeation, and the oldest skin

donor (donor 4) was not the one with lowest permeation, or vice versa. Although this study was not designed to examine this phenomenon, differences in donor age should be closely considered when correlating *in vitro* with *in vivo* data (Benson and Watkinson, 2012).

Despite these limitations, it is important to denote that for these batches, rheology combined with IVPT methods, clearly showed an interconnection between viscosity attributes and permeation behavior, for different batches of the same RP product. Conversely, even though equivalence as per EMA draft guideline requirements was not observed, the release differences between both batches were not as expressive (Miranda et al., 2023a).

3.2. Disease model

Aiming at the development of a surrogate method to support *in vivo* local bioavailability, a disease model mimicking an antifungal skin infection condition was established. To this end, the sensitivity of the proposed methodology was inspected based on the monitoring of ATP levels as outcome (Please see Figure S2). In high-throughput screening investigations, ATP assays are a common way to quantify the number of viable microbial cells as it is relatively easy and cheap to perform. After cell death has occurred the microbial cells lose the ability to produce new ATP molecules and ATPases present in the cell rapidly degrade any remaining ATP. This assay thus only quantifies cells that are metabolically active (Riss et al., 2016).

Figure S2 shows a representative calibration curve which presents the ATP concentration against luminescence units (LU) for measured ATP standards between 1.1 and 220.0 ng/mL. The sample calibration curve demonstrates that the amount of luminescence measured was directly proportional to the amount of ATP present (as evidenced by the R^2 value of 0.9999 when regression analysis was performed). Therefore, the ATP assay method was determined to be fit for purpose. Fig. 4 displays the application of the model developed to evaluate the antifungal activity of the formulations involved in the study and the respective control.

Using the Shapiro-Wilk test at a 95 % confidence level, it was determined that the raw data was not normally distributed, therefore the data was natural logged before further statistical tests were performed. The Tukey-Kramer HSD test was used to compare the mean natural logged luminescence between formulations and the infected control. As shown in Fig. 4, both formulations are not significantly different from each other, however RP4 is statistically different from the infected control.

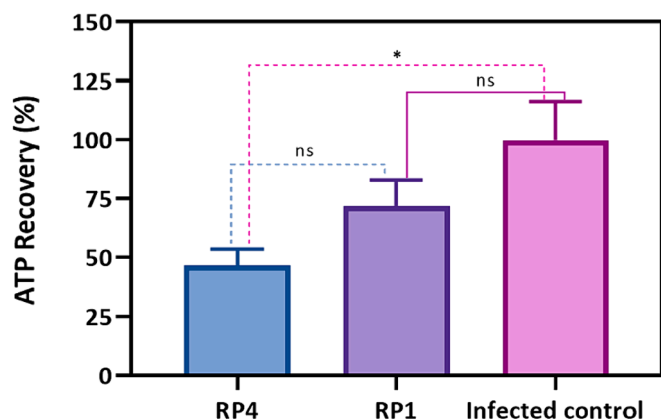As previously stated, during the model sensitivity investigation a large placebo effect was observed. This led to further investigation to find a suitable control, using the method described in section 2.2.3.2. A range of controls were trialed, including a placebo cream and an alternative placebo with potential antifungal excipients removed. This investigation found that the alternative placebo provided a lesser placebo effect than the original placebo formulation. The original placebo formulation had the largest effect on ATP recovery, showing a 20 % ATP recovery compared to the infected control. It was also found that the controls assigned a smaller effect on ATP recovery than both placebo formulations, with water as a control showing an 80 % ATP recovery compared to the infected control. It is worth noting that this placebo effect can also be seen in clinical trial data for bifonazole products in which mycological cure rate in comparison to the placebo was not always statistically significant (Bayer, 2012). In addition, a vehicle effect can often be seen with topical formulations in general; for example, a study by Epstein and Stein Gold (2013) showed up to a 45 % reduction in inflammatory lesions in patients that received a vehicle foam for treatment of acne vulgaris (Epstein and Gold, 2013). This demonstrates that the action of applying a topical product can result in an objective beneficial patient outcome. Although this disease model failed to show sensitivity, it could be argued that the level of vehicle effect and efficacy of the API against the target organism may influence the level of validation that can be achieved. For example, previous studies using more efficacious drugs (such as efinaconazole) against *T. Rubrum* (Turner et al., 2016) have had greater success in validating the disease model used in this study, resulting in a higher level of discrimination between formulations tested. In this regard alternate organisms which bifonazole has been shown to be efficacious against, such as *Candida albicans* (Lalošević et al., 2009), may have been more appropriate.

Despite the fact that the disease model was unable to statistically differentiate between the formulation batches where the IVPT model was discriminatory, the lowest permeant formulation (RP4) evidenced a higher antifungal *in vitro* activity as reported by the lower levels of ATP. This behavior could be ascribed to the reduced permeation kinetics observed for this formulation, leading to the accumulation of the drug in the upper layers of the skin where the disease is prominent, a pattern compatible with the pharmacodynamic profile of the drug. Note that this trend is also consistent with the rheological behavior, where a lower zero-shear viscosity and yield point is reported for RP1, the higher permeant formulation.

3.3. Weight of evidence

A critical appraisal of the IVPT data and disease model results is necessary, in order to draw the most reasonable conclusions. This evaluation should closely consider the variability inherent to both methods. As highlighted in the literature, managing variability is one of the main hurdles associated with topical bioequivalence assessment (Benson and Watkinson, 2012; Brown et al., 2023; Brown and Williams, 2019; Ili and Daniels, 2017; Ilić et al., 2021; Mangas-Sanjuán et al., 2019; Xu et al., 2020). As mentioned in the introductory section, besides documenting the qualitative and quantitative sameness of the TP to the RP, generic manufacturers are expected to prove microstructure equivalence. Within this scope, there are a myriad of physicochemical and structural characterization tests that should be employed in order to promptly describe, and consequently compare, the type, amount and arrangement of matter in the dosage form (Raney et al., 2023). According to the FDA Guidance for Industry - Physicochemical and Structural (Q3) Characterization of Topical Drug Products Submitted in ANDAs, as well as the recently published stimuli article on the USP expert panel, these tests include:

- Organoleptic characterization;
- Microscopic analysis aiming the evaluation of the phase states, as well as the structural organization of matter;
- Polymorphs characterization;
- Rheology characterization;

Fig. 4. ATP recovery from each formulation compared to the infected control, presented as a percentage. Results report the mean ± SEM calculated from pooled values across experimental set ups. RP1 n = 36, RP4 n = 36 and Infected Control n = 31.

- Solvent activity;
- pH;
- Oleaginous components characterization, if applicable;
- Specific gravity.

In previous studies, it was found that there were large variability differences in the rheology profile of the two batches. Depending on the selected endpoint, RSD values would range from 0 to 45 % (3 replicates). Inversely, when considering IVRT data, used to support performance equivalence, the registered variability was less (< 10.0 %). However, in the present work, the encountered variability in IVPT data was again much higher. This occurrence can be attributed to the use of human skin, but also to the variability inherent to IVPT methodology. The results attained with the disease model also exhibited high variability, as a consequence of the use of biological membranes, combined with the use of live microorganisms and inherent methodological variability.

It is important to correlate such *in vitro* variability with that observed *in vivo*. High variability is registered in clinical endpoint studies (CES) - the gold standard method for evaluating topical bioequivalence. As skin permeation is affected by several factors and some reference products possess a modest therapeutic efficacy, in order to manage variability, CES require a complex structure, with the enrolment of many subjects (n > 500) (Boix-Montanes, 2011; Chang et al., 2013; Harris, 2015; Narkar, 2010). Furthermore, the method itself displays many critical parameters, including the difficulty in standardizing patient dosing. A paper by Dina Vind-Kezunovic and colleagues states that the application of topical products by individuals is inherently variable and can account for a 13.6-fold difference alone (Vind-Kezunovic and Serup, 2016). The difficulty of ensuring compliance with treatments is also problematic, along with the differences between the skin of the individuals entering the study and their respective disease state, i.e. severity based on organism and damage perspective (Ariyanayagam et al., 1985).

On a practical note: according to regulatory requirements, the number of required replicates to validate the differences observed in IVRT is 6–12. This number obviously dramatically increases when dealing with product efficacy equivalence studies, such as IVPT or disease model methods, where there is going to be a greater variability in the data. For instance, the EMA draft guideline for IVPT suggests that the pilot study could be useful in estimating sample size but then prescribes that the number of donors should be not less than 12 and the number of replicates for each donor should be at least 2. Such a lack of clarity concerning the distribution of these numbers across the pilot and pivotal studies is confusing. Nevertheless, as shown in this study, these numbers may not suffice and the study could have been improved by using the pilot study to allow for a better estimation of the sample size to power the pivotal study, in much the same way as that proposed by the US-FDA in 2022. In fact, the FDA In Vitro Permeation Test Studies draft guidance takes this even further by advising on a more practical inclusion of 4–6 skin donors but with a minimum of four replicate skin sections per donor per treatment group. Like the EMA, the FDA recommends testing on the same donors and the normalization of the number of replicates per donor, in order to perform a balanced statistical analysis (EMA, 2018; US-FDA, 2022a). Other limitations of the EMA draft guideline includes the tighter acceptance criteria, as well as in the absence of a recommended statistical analysis protocol. The lack of detail in such critical aspects poses as a significant challenge to any company involved in such studies and thus may act as a deterrent towards topical generic development. Within this scope, despite all the regulatory efforts on providing alternative frameworks for topical bioequivalence documentation, there is still a real need for the regulatory agencies to consider a more bespoke specific product approach in which a biowaiver from product efficacy studies could be applicable.

Fig. 5 intends to provide a schematic flow of the several steps involved in the documentation of topical bioequivalence and at the same time to shed light on the most problematic features that can arise from this framework.

The general framework for documenting topical bioequivalence is reliant on the technological features of the formulation. In this context, according to EMA draft guideline requirements, for monophasic formulations, equivalence should be sustained for Q1-Q3 attributes, as well as for performance (Q4). Single phase systems may be defined as "a single-phase base in which the active substance is in solution or suspension e.g. cutaneous solutions, single phase gels and ointments; cutaneous suspensions" (EMA, 2018). However, from a regulatory perspective, applying this biowaiver from product efficacy equivalence studies could also be useful for multiphasic formulations, in which the time since approval for the reference product exceeds 10 years. This resembles the requirement for biographic applications. Note, however, that this abridged process is commonly adopted for other dosage forms.

Documenting Q1 and Q2 equivalence is one of the cornerstones of topical generic products development, as every batch of a pharmaceutical product is designed to have the same Q1 and Q2 (Raney et al., 2023). However, formulators may need to overcome several challenges (Chang et al., 2013):

- Patent protection of the RP;
- Undesirable RP characteristics;
- Difficulty in determining the excipient grade of the RP;
- Excipients not present in the market.

In what concerns Q2 equivalence, according to EMA draft guideline, a strength biowaiver can be applied, for different strength products. These products solely differ in the amount of API, being the basic formulation, equipment and manufacturing process the same for all strengths. Therefore, if there is a proper documentation of extended pharmaceutical equivalence parameters, a biowaiver from product efficacy studies could be granted. This is a positive feature of the European guideline, as it facilitates the work and costs burden for the generic manufacturer.

When it is not possible to assure Q1 and Q2 equivalence, it is important to reflect upon the framework provided by Vinod Shad and collaborators, concerning the Topical Drug Classification system (Shah et al., 2015). The authors stated that for products that are not Q1 and Q2 equivalent (which are collectively referred to as class 3 topical products), a biowaiver could also be considered. For instance, API with specific physicochemical properties that strongly influence the formulation main critical attributes, i.e. extreme lipophilicity, the excipients impact on the permeation kinetics is not significant. Additionally, if product performance, evaluated through IVRT, is not altered by Q1/Q2 changes, there is no scientific reason to believe that the product will not be therapeutically equivalent (Shah et al., 2015).

As previously detailed, Q3 characterization involves a range of specific tests aiming to best describe the critical quality attributes of the product. In this context, and as stated by Sam Raney and colleagues, the information gathered by each method should be collectively analysed to infer on Q3 (Raney et al., 2023). There are several identified challenges in this task: (i) identifying the tests that provide factual information on product CQA, as these are product specific; (ii) Managing RP variability, as referred in this study as well as in other literature reports, the selection of the RP batch may not be irrelevant (Ili and Daniels, 2017; Miranda et al., 2023b, 2020b; Xu et al., 2020); (iii) Defining suitable acceptance criteria for Q3 equivalence. According to the FDA Guidance for Industry - Physicochemical and Structural (Q3) Characterization of Topical Drug Products Submitted in ANDAs, Q3 sameness may only be supported if the product is Q1/Q2. Furthermore, each relevant Q3 attribute of the TP must be within the range characterized for the RP, potentially characterized in multiple batches. The guideline also states that Q3 sameness may be declared by the Agency, if there is an acceptable variability of the TP/RP microstructure parameters (US-FDA, 2022b). However, this view differs from the EMA draft guideline where it is requested that for all quantitative quality characteristics, the 90 % confidence interval for the difference of means of the test and
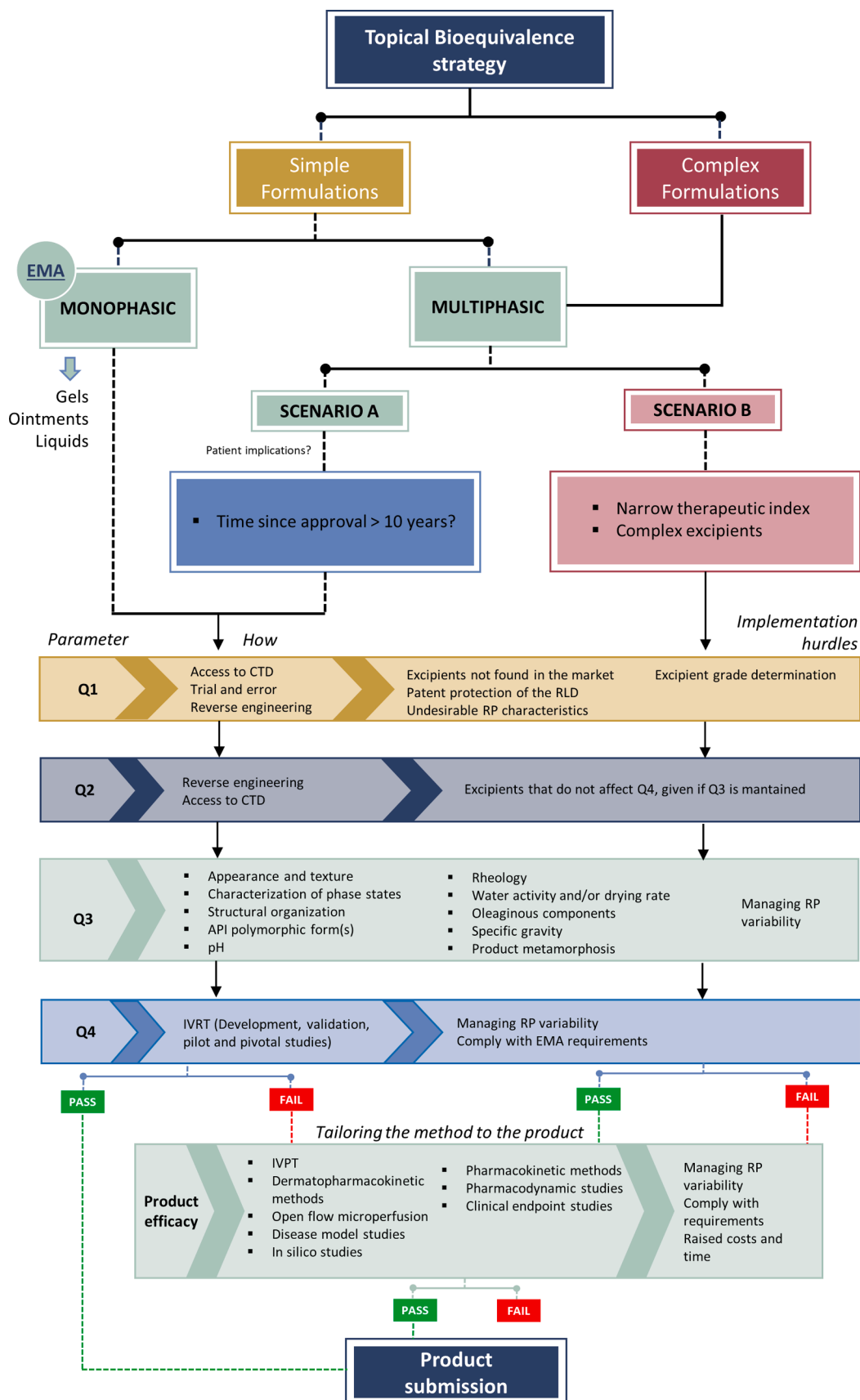
Fig. 5. Implementation hurdles to topical bioequivalence documentation.

comparator products should be contained within the acceptance criteria of +/- 10 % of the comparator product mean, assuming normal distribution of data. This criteria, taking into account the complexity of semisolid microstructure, is in many cases too strict (Ili and Daniels, 2017; Miranda et al., 2023b, 2020b; Xu et al., 2020).

In what concerns the documentation of product performance equivalence, clear information is provided in the guidelines, ranging from the workflow to be applied to IVRT studies (method development, analytical method validation, IVRT validation, pilot and pivotal studies), to the corresponding acceptance criteria. However, a significant challenge in this step relies on the restrictive EMA draft guideline criteria for statistical equivalence (90–111 %), which is not shared by the FDA (EMA, 2018; US-FDA, 2022c). A recent study by Wellington *et al.* used a validated IVRT method to conduct comparative IVRT runs on five generic products marketed in South Africa and one Canadian generic. These products were compared with a relevant RP. When the authors compared the RP product batches, despite the release profile being extremely close, performance equivalence according to EMA draft guideline failed to be supported, as the confidence interval was 84.03–102.86 % (Wellington et al., 2023).

The next step regards product efficacy documentation. As these studies are far more complex than the previous ones, their presentation would yield meaningful results for multiphasic formulations with added complexity, or in cases in which the other equivalence parameters failed to be supported. The experimental approach to follow is highly reliant on the product features/clinical target. Depending on the product action site, as well as pharmacotherapeutic profile, IVPT studies, tape stripping, microdialysis, open flow microperfusion, disease model studies, may be viable options. As previously reviewed, managing method variability may pose as a significant challenge in this step. Both IVPT and disease model methods displayed a level of variability consistent with clinical therapeutic equivalence studies, the gold standard, which is an extremely costly and time consuming method for evaluating topical bioequivalence. The skin disease model studies were unable to statistically differentiate between the formulation batches where the IVPT model was discriminatory; however, it should be pointed out that differences in the efficacy profile of two batches of the same product are not expected to occur *in vivo*. Furthermore, the results of the disease model are in agreement with those attained in IVPT (lowest permeant formulation had a higher antifungal *in vitro* activity), as well as rheology (lowest viscosity formulation had a higher antifungal *in vitro* activity). However, there were practical hurdles during the development of the present model, namely in what concerned the documentation of the method sensitivity. In the manuscript, the use of *Candida albicans,* instead of *T. Rubrum* is appointed as a strategy to improve this aspect.

Previous studies using more efficacious drugs (such as efinaconazole) against *T. Rubrum* (Turner et al., 2016) have had greater success in demonstrating sensitivity in the disease model used in this investigation, resulting in a higher level of discrimination between formulations tested. Therefore, we may have had more success using a different organism (although perhaps less relevant) against which bifonazole has greater efficacy, such as *Candida albicans* as shown by Lalosević et al. (Lalosević et al., 2009); (Lalosević et al., 2009). Alternatively, it is possible that more pre-established microbiological test methods, for example zone of inhibition investigations (perhaps with a keratin supplemented media to more closely resemble *in vivo* skin) or a minimum inhibitory concentration experiment could have been used, although obviously the skin itself is not present. Furthermore, when addressing the conduct of IVPT studies, the regulatory agencies, more specifically, the FDA, mainly focus on the topical acyclovir cream PSG. Even though this product is highly complex, the therapeutic perspective is very poor, and for that reason not comparable to other products available in the market. The presentation of other case studies would be beneficial for generic manufacturers.

As referenced in the regulatory guidances, clinical studies, including vasoconstriction, pharmacokinetic and clinical endpoint studies, are of course available, but are extremely time consuming and expensive and their need should be solely required for the worst cases. Nevertheless, no matter what developments and changes are made in the processes to demonstrate acceptable bioequivalence for a topical generic product, the return on investment will be what determines whether any topical generic will be commercialised.

4. Concluding remarks

In this work, we have investigated both the experimental procedures and regulatory mechanisms underlying the efficacy profile equivalence assessment of bifonazole 10 mg/g cream formulations.

According to European and American regulatory agencies, semisolid dosage forms that exhibit a complex microstructure, such as creams, should present comparative local availability studies to document product efficacy equivalence. IVPT studies were carried out using two RP batches that displayed distinct rheological profiles. The IVPT results were then analysed according to two statistical approaches – EMA and FDA. Equivalence between RP1 *vs*. RP4 failed to be supported by both approaches. This highlights that the selection of RP batches for this specific case study is a critical step in documenting bioequivalence of topical generic products.

Overall, skin permeation kinetics methodology has an intrinsically high variability that could hampered result interpretation. The proposal of wider acceptance criteria could be highlighted as a measure to circumvent this issue. In contrast, the disease model herein presented could provide an alternative and more straightforward product comparison approach and more reliable strategy to infer on topical antifungal bioavailability, as it is not expected that different batches would yield a different therapeutic response. Note that this disease model has already shown promise for evaluating the efficacy of formulations containing other APIs, for example when evaluating the formulation Lamisil®.

CRediT authorship contribution statement

Margarida Miranda: Writing – original draft, Formal analysis, Data curation, Conceptualization. **Zoe Volmer:** Investigation, Formal analysis, Data curation. **Alicia Cornick:** Investigation, Formal analysis, Data curation. **Aidan Goody:** Investigation, Formal analysis, Data curation. **Catarina Cardoso:** Supervision, Resources. **Alberto A.C.C. Pais:** Writing – review & editing, Validation, Formal analysis, Data curation. **Marc Brown:** Writing – review & editing, Resources, Methodology, Formal analysis, Data curation, Conceptualization. **Carla Vitorino:** Writing – review & editing, Supervision, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijpharm.2024.124012.

References

Abd, E., Yousef, S.A., Pastore, M.N., Telaprolu, K., Mohammed, Y.H., Namjoshi, S., Grice, J.E., Roberts, M.S., 2016. Skin models for the testing of transdermal drugs. Clin. Pharmacol. Adv. Appl. 8, 163–176. https://doi.org/10.2147/CPAA.S64788.

Akomeah, F.K., Martin, G.P., Brown, M.B., 2007. Variability in human skin permeability in vitro: comparing penetrants with different physicochemical properties. J. Pharm. Sci. 96, 824–834.

Ariyanayagam, M., Barlow, T.J.G., Graham, P., Hall-Smith, S.P., Harris, J.M., 1985. Topical sodium cromoglycate in the management of atopic eczema—a controlled trial. Br. J. Dermatol. 112, 343–348. https://doi.org/10.1111/j.1365-2133.1985.tb04863.x.

Bayer, 2012. An Investigator-blind, Randomized, Multicenter, 5-arm, Placebo- and Active Controlled Parallel Group Pilot Trial to Explore the Efficacy and Tolerability of Topical Bifonazole Liquid Spray in Patients With Athlete's Foot [WWW Document]. Clin. NCT01013909. URL https://www.clinicaltrialsregister.eu/ctr-search/trial/2008-005654-21/results (accessed 2.20.24).

Benson, H.A.E., Watkinson, A.C., 2012. Topical and transdermal drug delivery: principles and practice. John Wiley & Sons.

Binder, L., Mazál, J., Petz, R., Klang, V., Valenta, C., 2019. The role of viscosity on skin penetration from cellulose ether-based hydrogels. Ski. Res. Technol. 1–10 https://doi.org/10.1111/srt.12709.

Boix-Montanes, A., 2011. Relevance of equivalence assessment of topical products based on the dermatopharmacokinetics approach. Eur. J. Pharm. Sci. 42, 173–179. https://doi.org/10.1016/j.ejps.2010.11.003.

Bormann, J.L., Maibach, H.I., 2020. Effects of anatomical location on in vivo percutaneous penetration in man. Cutan. Ocul. Toxicol. 213–222. https://doi.org/10.1080/15569527.2020.1787434.

Brown, M.B., Williams, A.C., 2019. The art and science of dermal formulation development. CRC Press.

Burns, T., Breathnach, S.M., Cox, N., Griffiths, C., 2008. Rook's textbook of dermatology. John Wiley & Sons.

Chang, R.-K., Raw, A., Lionberger, R., Yu, L.X., 2013. Generic development of topical dermatologic products: formulation development, process development, and testing of topical dermatologic products. AAPS J. 15, 41–52. https://doi.org/10.1208/s12248-012-9411-0.

CPMP/ICH/381/95, 2005. Validation of analytical procedures: Text and Methodology Q2(R1).

Demurtas, A., Pescina, S., Nicoli, S., Santi, P., De, D.R., Padula, C., 2020. Validation of a HPLC-UV method for the quantification of budesonide in skin layers. J. Chromatogr. B 122512. https://doi.org/10.1016/j.jchromb.2020.122512.

EMA, 2018. Draft guideline on quality and equivalence of topical products. CHMP/QWP/708282/2018.

Epstein, E.L., Gold, L.S., 2013. Safety and efficacy of tazarotene foam for the treatment of acne vulgaris. Clin. Cosmet. Investig. Dermatol. 6, 123–125. https://doi.org/10.2147/CCID.S34054.

Faway, É., Cambier, L., Mignon, B., Poumay, Y., De Rouvroit, C.L., 2017. Modeling dermatophytosis in reconstructed human epidermis: a new tool to study infection mechanisms and to test antifungal agents. Med. Mycol. 55, 485–494. https://doi.org/10.1093/mmy/myw111.

FDA, 2022. Draft Guidance on Acyclovir [WWW Document]. URL https://www.accessdata.fda.gov/drugsatfda_docs/psg/Acyclovir_topical cream_RLD 21478_RV12-16.pdf (accessed 12.24.21).

Fernández-Campos, F., Obach, M., Moreno, M.C., García, A., González, J., 2017. Pharmaceutical development of a generic corticoid semisolid formulation. J. Drug Deliv. Sci. Technol. 42, 227–236. https://doi.org/10.1016/j.jddst.2017.03.016.

Feschuk, A.M., Kashetsky, N., Chiang, C., Burli, A., Burdick, H., Maibach, H.I., 2022. Regional variation in percutaneous absorption in in vitro human models: a systematic review. J. Toxicol. Environ. Heal. - Part B Crit. Rev. 25, 97–112. https://doi.org/10.1080/10937404.2022.2032517.

Food and Drug Administration, CDER, 2002. Draft Guidance for Industry: In Vitro Release Test Studies for Topical Drug Products Submitted in ANDAs FDA-2022-D.

Franz, T.J., 1975. Percutaneous absorption. on the relevance of in vitro data. J. Invest. Dermatol. 64, 190–195. https://doi.org/10.1111/1523-1747.ep12533356.

Harris, R., 2015. Demonstrating therapeutic equivalence for generic topical products. Pharm. Technol. 1–4.

Hossain, A.S.M.M.A., Sil, B.C., Iliopoulos, F., Lever, R., Hadgraft, J., Lane, M.E., 2019. Preparation, Characterisation, and topical delivery of terbinafine. Pharmaceutics 11, 548. https://doi.org/10.3390/pharmaceutics11100548.

Ili, T., Daniels, R., 2017. Critical quality attributes, in vitro release and correlated in vitro skin permeation — in vivo tape stripping collective data for demonstrating therapeutic (non) equivalence of topical semisolids. A Case Study of 528, 253–267. https://doi.org/10.1016/j.ijpharm.2017.06.018.

Ilić, T., Pantelić, I., Savić, S., 2021. The implications of regulatory framework for topical Semisolid drug products: from critical quality and performance attributes towards establishing bioequivalence. Pharmaceutics 13, 710. https://doi.org/10.3390/pharmaceutics13050710.

Jin, X., Imran, M., Mohammed, Y., 2022. Topical Semisolid products—Understanding the impact of metamorphosis on skin penetration and physicochemical properties. Pharmaceutics 14, 2487. https://doi.org/10.3390/pharmaceutics14112487.

Kamal, N.S., Krishnaiah, Y.S.R., Xu, X., Zidan, A.S., Raney, S., Cruz, C.N., Ashraf, M., 2020. Identification of critical formulation parameters affecting the in vitro release, permeation, and rheological properties of the acyclovir topical cream. Int. J. Pharm. 590, 119914 https://doi.org/10.1016/j.ijpharm.2020.119914.

Lalošević, J., Rojas, R., Astorga, E., Gip, L., 2009. Bifonazole cream in the treatment of superficial Candidosis: a double-blind Comparative study. Dermatologica 169, 99–106. https://doi.org/10.1159/000249647.

Leal, L.B., Cordery, S.F., Delgado-Charro, M.B., Bunge, A.L., Guy, R.H., 2017. Bioequivalence methodologies for topical drug products. in vitro and ex vivo studies with a corticosteroid and an anti-fungal drug. Pharm. Res. 34, 730–737. https://doi.org/10.1007/s11095-017-2099-1.

Lu, M., Xing, H., Chen, X., Xian, L., Jiang, J., Yang, T., Ding, P., 2016. Advance in bioequivalence assessment of topical dermatological products. Asian J. Pharm. Sci. 11, 700–707. https://doi.org/10.1016/j.ajps.2016.04.008.

Mangas-Sanjuán, V., Pleguezuelos-Villa, M., Merino-Sanjuán, M., Hernández, M.J., Nácher, A., García-Arieta, A., Peris, D., Hidalgo, I., Soler, L., Sallan, M., Merino, V., 2019. Assessment of the inter-batch Variability of microstructure Parameters in topical Semisolids and impact on the demonstration of equivalence. Pharmaceutics. https://doi.org/10.3390/pharmaceutics11100503.

Miranda, M., Cova, T., Augusto, C., Pais, A.A.C.C., Cardoso, C., Vitorino, C., 2020. Diving into batch-to-batch Variability of topical products-a regulatory bottleneck. Pharm. Res. 37, 218. https://doi.org/10.1007/s11095-020-02911-y.

Miranda, M., Cardoso, C., Pais, A.A.C.C., Brown, M., Vitorino, C., 2023. Drilling down the bioequivalence assessment of topical antifungal products: microstructure and release. Eur. J. Pharm. Biopharm. 185, 94–106. https://doi.org/10.1016/j.ejpb.2023.02.006.

Miranda, M., Cardoso, C., Vitorino, C., 2020a. Fast Screening Methods for the Analysis of Topical Drug Products. Process. 2020, Vol. 8, Page 397 8, 397. https://doi.org/10.3390/PR8040397.

Miranda, M., Sousa, J.J., Veiga, F., Cardoso, C., Vitorino, C., 2018. Bioequivalence of topical generic products. Part 1: where are we now? Eur. J. Pharm. Sci. 123, 260–267. https://doi.org/10.1016/j.ejps.2018.07.050.

Miranda, M., Pais, A.A.C.C., Cardoso, C., Vitorino, C., 2019. aQbD as a platform for IVRT method development–a regulatory oriented approach. Int. J. Pharm. 118695 https://doi.org/10.1016/j.ijpharm.2019.118695.

Miranda, M., Veloso, C., Brown, M., Pais, A.A.C.C., Cardoso, C., Vitorino, C., 2022. Topical bioequivalence: Experimental and regulatory considerations following formulation complexity. Int. J. Pharm. 121705 https://doi.org/10.1016/j.ijpharm.2022.121705.

Nagelreiter, C., Raffeiner, S., Geyerhofer, C., Klang, V., Valenta, C., 2013. Influence of drug content, type of semi-solid vehicle and rheological properties on the skin penetration of the model drug fludrocortisone acetate. Int. J. Pharm. 448, 305–312. https://doi.org/10.1016/j.ijpharm.2013.03.042.

Narkar, Y., 2010. Bioequivalence for topical products-an update. Pharm. Res. 27, 2590–2601. https://doi.org/10.1007/s11095-010-0250-3.

OECD, 2010. OECD Guidance Notes on Dermal Absorption Draft 22 october 2010 1–53.

Pensado, A., Chiu, W.S., Cordery, S.F., Rantou, E., Bunge, A.L., Delgado-Charro, M.B., Guy, R.H., 2019. Stratum corneum sampling to assess bioequivalence between topical Acyclovir products. Pharm. Res. 36, 180. https://doi.org/10.1007/s11095-019-2707-3.

Poet, T.S., McDougal, J.N., 2002. Skin absorption and human risk assessment. Chem. Biol. Interact. 140, 19–34. https://doi.org/10.1016/S0009-2797(02)00013-3.

S.G. Raney S. Senemar M. Burke C. Lee J.R. Shah K.S. Li O. Anand K.S. Warner S.P. Scientist Advances in Product Quality and Performance Tests for Topical and Transdermal Products-View of the USP Expert Panel 2023 6 11 .

Riss, T.L., Moravec, R.A., Niles, A.L., Duellman, S., Benink, H.A., Worzella, T.J., Minor, L., 2016. Cell viability assays. Assay Guid. Man. [Internet].

Sahoo, A., Mahajan, R., 2016. Management of tinea corporis, tinea cruris, and tinea pedis: a comprehensive review. Indian Dermatol. Online J. 7, 77. https://doi.org/10.4103/2229-5178.178099.

Schmitt, S., Schäfer, U.F., Döbler, L., Reichling, J., 2010. Variation of in vitro human skin permeation of rose oil between different application sites. Complement. Med. Res. 17, 126–131.

Shah, V.P., Yacobi, A., Flavian, Ş., Miron, D.S., Majella, E., 2015. A science based approach to topical drug classification system (TCS). Int. J. Pharm. 1–15 https://doi.org/10.1016/j.ijpharm.2015.06.011.

Shin, S.H., Rantou, E., Raney, S.G., Ghosh, P., Hassan, H., Stinchcomb, A., 2020. Cutaneous Pharmacokinetics of Acyclovir cream 5% products: evaluating bioequivalence with an in vitro permeation test and an adaptation of scaled average bioequivalence. Pharm. Res. 37 https://doi.org/10.1007/s11095-020-02821-z.

Simões, A., Miranda, M., Cardoso, C., Veiga, F., Vitorino, C., 2020. Rheology by design: a regulatory tutorial for analytical method validation. Pharmaceutics 12, 820. https://doi.org/10.3390/pharmaceutics12090820.

Sivaraman, A., Ganti, S.S., Nguyen, H.X., Birk, G., Wieber, A., Lubda, D., Banga, A.K., 2017. Development and evaluation of a polyvinyl alcohol based topical gel. J. Drug Deliv. Sci. Technol. 39, 210–216. https://doi.org/10.1016/j.jddst.2017.03.021.

Soriano-Ruiz, J.L., Calpena-Capmany, A.C., Cañadas-Enrich, C., Febrer, N.B., Suñer-Carbó, J., Souto, E.B., Clares-Naveros, B., 2019. Biopharmaceutical profile of a clotrimazole nanoemulsion: evaluation on skin and mucosae as anticandidal agent. Int. J. Pharm. 554, 105–115. https://doi.org/10.1016/J.IJPHARM.2018.11.002.

Surber, C., Knie, U., 2018. Metamorphosis of vehicles: mechanisms and opportunities. Curr. Probl. Dermatology 54, 152–165. https://doi.org/10.1159/000489529.

Tothfalusi, L., Endrenyi, L., 2011. Sample sizes for designing bioequivalence studies for highly variable drugs. J. Pharm. Pharm. Sci. 15, 73–84. https://doi.org/10.18433/j3z88f.

Turner, R., Weaver, S., Caserta, F., Brown, M.B., 2016. A novel vehicle for enhanced drug delivery across the human nail for the treatment of onychomycosis. Int. J. Pharm. Compd. 20, 71–80.

Us-fda,, 2022a. In vitro permeation test studies for topical drug products submitted in ANDAs guidance for industry. Guid, Ind.

Us-fda,, 2022b. Guidance for industry - physicochemical and structural (Q3) Characterization of topical drug products submitted in ANDAs. Guid, Ind.

Us-fda,, 2022c. In vitro release test studies for topical drug products submitted in ANDAs guidance for industry. Guid, Ind.

USP, 2023. Semisolid Drug Products - Performance Tests. Pharmacopeial forum 1225.

Vind-Kezunovic, D., Serup, J.V., 2016. Variation of topical application to skin under good clinical practice (GCP): a "best performance" scenario. J. Dermatolog. Treat. 27, 422–424. https://doi.org/10.3109/09546634.2015.1124973.

Vitorino, C., Almeida, A., Sousa, J., Lamarche, I., Gobin, P., Marchand, S., Couet, W., Olivier, J.C., Pais, A., 2014. Passive and active strategies for transdermal delivery using co-encapsulating nanostructured lipid carriers: in vitro vs. in vivo studies. Eur. J. Pharm. Biopharm. 86, 133–144. https://doi.org/10.1016/j.ejpb.2013.12.004.

Wellington, H., Rath, S., Abboo, S., Kanfer, I., 2023. Utility of in vitro release testing (IVRT) to assess 'sameness' of 1 % clotrimazole creams for use as a biowaiver. AAPS Open. https://doi.org/10.1186/s41120-023-00087-4.

Xu, Z., Mangas-Sanjuán, V., Merino-Sanjuán, M., Merino, V., García-Arieta, A., 2020. Influence of inter-and intra-batch variability on the sample size required for demonstration of equivalent microstructure of semisolid dosage forms. Pharmaceutics 12, 1–15. https://doi.org/10.3390/pharmaceutics12121159.