Regulatory Toxicology and Pharmacology 132 (2022) 105184

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

An OECD TG 428 study ring trial with ^{14}C-Caffeine demonstrating repeatability and robustness of the dermal absorption *in vitro* method

Felix M. Kluxen [a,†], Styliani Totti [b,†,1], Wilfred Maas [c], Frank Toner [d], Leanne Page [d], Kathryn Webbley [e], Rajendra Nagane [f], Robert Mingoia [g], Christine Whitfield [g], John Kendrick [j], Claire Valentine [h], Jeanne Bernal Dorange [i], Camille Egron [l], Camille Imart [l], Jeanne Y. Domoradzki [j], Philip Fisher [k], Christine Lorez [l], Steve McEuen [m], Edgars Felkers [a], Tao Chen [b], Christiane Wiemann [n,*]

[a] *ADAMA Deutschland GmbH, Cologne, Germany*
[b] *University of Surrey, Guildford, United Kingdom*
[c] *Charles River Laboratories, Den Bosch, the Netherlands*
[d] *Charles River Laboratories, Tranent, United Kingdom*
[e] *Pharmaron, Rushden, United Kingdom*
[f] *Jai Research Foundation, Valvada, India*
[g] *Corteva Agriscience, Newark, United States*
[h] *Labcorp Drug Development, Harrogate, United Kingdom*
[i] *Eurofins Agroscience Services Chem SAS, Vergeze, France*
[j] *Corteva Agriscience, Indianapolis, United States*
[k] *Bayer SAS, Bayer Crop Science, Sophia, Antipolis, France*
[l] *Syngenta Crop Protection AG, Basel, Switzerland*
[m] *FMC Corporation, Philadelphia, PA, USA*
[n] *BASF Österreich GmbH, Vienna, Austria*

A R T I C L E I N F O

Handling Editor: Dr. Lesa Aylward

Keywords:
Dermal absorption
In vitro
Human skin preparations
Reference compound
OECD TG 428
Caffeine
Inter-laboratory repeatability
Intra-laboratory repeatability
Ring trial

A B S T R A C T

The dermal absorption potential of ^{14}C-Caffeine applied as a 4 mg/mL concentration (10 μL/cm^2 finite dose) was investigated in six laboratories under Good Laboratory Practice conditions using an OECD TG 428-compliant *in vitro* assay with flow-through cells and split-thickness human skin. Potential sources of variation were reduced by a standardized protocol, test item and skin source. Particularly, skin samples from same donors were distributed over two repeats and between labs in a non-random, stratified design. Very similar recovery was achieved in the various assay compartments between laboratories, repeats and donors, demonstrating that the assay can be robustly and reliably performed. The absorption in one laboratory was 5-fold higher than in the others. This did not clearly correlate with skin integrity parameters but might be associated with an accidental COVID-19 pandemic-related interruption in sample shipment. It is possible that other factors may affect dermal absorption variation not routinely assessed or considered in the current method. The mean receptor fluid recovery, potential absorption (recovery in receptor fluid and skin except tape strips 1 and 2) and mass balance of caffeine was 6.99%, 7.14% and 99.13%, respectively, across all and 3.87%, 3.96% and 99.00% in the subset of five laboratories.

Abbreviations: CLE, CropLife Europe; ECPA, European Crop Protection Association; ADME, Absorption, Distribution, Metabolism and Excretion; OECD TG, Organisation for Economic Co-operation and Development test guideline; HPLC-MS/MS, high-pressure liquid chromatography/mass spectrometry; LSC, liquid scintillation counting; PBS, physiological saline solution; GLP, Good Laboratory Practice; HIV, human immunosufficiency virus; HBV, hepatitis B virus; HCV, hepatitis C virus; SOPs, standard operation procedures; TWL, titriated water flux; TEER, transepidermal electrical resistance; TEWL, transepidermal water loss.

* Corresponding author.
E-mail address: christiane.wiemann@basf.com (C. Wiemann).
† These authors contributed equally to the publication.
[1] Now affiliated with Syngenta Syngenta Jealott's Hill International Research Centre, Bracknell, United Kingdom.

https://doi.org/10.1016/j.yrtph.2022.105184
Received 31 January 2022; Received in revised form 19 April 2022; Accepted 1 May 2022
Available online 13 May 2022

1. Introduction

Dermal absorption estimates are used in non-dietary risk assessment to convert dermal doses into systemic exposures. For this, dermal penetration studies are conducted to determine a suitable fraction, the relative dermal absorption value, which is the fraction that reaches the systemic compartment surrogate, based on the applied dose. Today, most regulatory studies for pesticide registration purposes are conducted according to the *in vitro* Organisation for Economic Co-operation and Development test guideline (OECD TG) 428 (OECD, 2004) using human skin and radioactive-labeled material.

The absorption value used for risk assessment is the sum of receptor fluid recovery and, based on the regulation or guidance, additionally recovery in different skin strata, *i.e.*, dermis, epidermis and *stratum corneum*. *Stratum corneum* recovery is determined by tape-stripping, which provides an approximate residue distribution in the *stratum corneum;* tape stripping does not result in exact or reproducible single skin layers. The layers' thickness may vary based on the tape strip material, applied pressure, residual material or moisture, etc. Assays are conducted usually considering 6–10 h (hr) exposure with a follow-up observation period resulting in a total assay time of 24 h to mimic the daily exposure of workers. As the penetration assay using radioactive-labeled material does not consider absorption, distribution, metabolism or excretion (ADME) processes, it has to be considered to be conservative *per se* (Kluxen et al., 2021), *i.e.*, it most likely overpredicts the absorbed fraction available to induce toxic effects.

While an internationally accepted test guideline (OECD, 2004) and further international (OECD, 2011) and regional guidance (EFSA, 2017) is available, the assay's robustness and reliability has not been formally validated.

The methodology has significantly evolved from its status at the time of OECD guideline development and is under continuous further development (Fabian et al., 2017; Hewitt et al., 2020; Heylings et al., 2018; Hopf et al., 2020; SCCS, 2010; Sullivan et al., 2017). Further, due to improvements in detections limits and specificity of high-pressure liquid chromatography/mass spectrometry (HPLC-MS/MS) these analytical methods play a more and more important role in these kind of studies (Gerstel et al., 2016; Wargniez et al., 2017) as an alternative to the former "gold standard" of radiolabel analytics by liquid scintillation counting (LSC)(Gerstel et al., 2016; Wargniez et al., 2017). To assess the methodology some limited inter-laboratory ring trials have been published but comparing different methodologies (Gerstel et al., 2016; Liu et al., 2018; Wargniez et al., 2017) and not necessarily conducting state-of-the-art fully OECD guideline compliant studies (Schäfer-Korting et al., 2008; van de Sandt et al., 2004) or having another focus like e.g. to assess skin sub-compartment distribution (Liu et al., 2018) or establishing alternatives to human skin as a suitable surrogate (Liu et al., 2018; Schäfer-Korting et al., 2006, 2008). Further, several laboratories have reported individual study results with the reference compounds mentioned in the OECD TG usually when assessing particular aspects of the methodology (Abd et al., 2019; Dreher et al., 2002; Guth et al., 2015; Hewitt et al., 2020; Heylings et al., 2018; Hui and Maibach, 2020; Im et al., 2021; Wilkinson et al., 2006).

Between 2016 and 2017, a multicenter ring trial (1st ring trial) was organized by CropLife Europe (CLE, previously the European Crop Protection Association, ECPA) and has been performed to test the *in vitro* dermal absorption of testosterone across 12 laboratories according to OECD TG 428 using human split-thickness skin. The testosterone application dose was 1 mg/mL in 40% aqueous ethanol vehicle and physiological saline solution (PBS) with 5% bovine serum albumin was used as the receptor fluid. The intra-laboratory repeatability was not intended to be evaluated in the 1st ring trial therefore, the participating laboratories performed a single experimental run. For the 1st ring trial many factors varied across the participant laboratories including, the diffusion cell type and size, reagent sources, the skin tissue suppliers, the skin origin and the skin membrane preparation. The outcome from the 1st ring trial showed unexpectedly high variability, for which no clear reason could be identified due to multifactorial variants in study design. However, it was hypothesized that the lipophilic properties of the test material in the context of the selected vehicle could have biased the results. It was discussed whether the evaporative properties of the solvent (containing ethanol), in relation to the size and dimension of the diffusion cells donor compartment could have influenced the speed of evaporation thus leading to variance of skin surface distributed solubilized testosterone available for penetration during the course of the study. Another potentially relevant variant debated, was the efficiency of the washing procedure. The results of the 1st ring trial are summarized in supplementary material 1.

Hence, there was a need to design and conduct a more controlled multicenter study in order to assess the robustness of the methodology and identify the main factors that could contribute to variability when conducting *in vitro* dermal absorption studies. Donor variability is one factor that is considered to have a relevant impact on the study outcome and thus was acknowledged by recent study conduct recommendations (EFSA et al., 2017; EFSA, 2012; SCCS, 2010).

Hence, CLE in collaboration with six laboratories, which regularly conduct dermal absorption studies, initiated a 2nd ring trial managed by the University of Surrey (United Kingdom). The studies were conducted under Good Laboratory Practice (GLP) conditions using an OECD TG 428-compliant *in vitro* assay with flow-through cells and split-thickness human skin. Potential sources of variation were reduced by a common protocol, test item and skin source. Particularly, skin samples from the same donors were distributed over two repeats and between labs in a non-random, stratified design to assess potential impact of the donor.

The methods and results are extensively described in a dedicated report available as supplementary material 2. The current manuscript summarizes the ring trial report.

2. Methods

The 2nd ring trail was conducted according to a common protocol, test item and skin source.

Please refer to supplementary material 2 for detailed methods and materials.

2.1. Test substance preparation and application

1-Methyl -^{14}C-Caffeine (PerkinElmer) was used as the testing compound. The batch number of ^{14}C-Caffeine has been standardised across the laboratories, to avoid any batch-to batch variability. ^{14}C- Caffeine was dissolved in Phosphate Buffered Saline (PBS, Merck) at a dose concentration of 4 mg/mL. The radioactivity applied to the skin was *ca.* 10 kBq/cm^2. This corresponds to a desired specific activity of the testing compound preparation of *ca.* 1 MBq/mL. ^{14}C-Caffeine was applied to the skin membranes uniformly with a pipette at a dose of 10 μL/cm^2.

2.2. Preparation of skin membranes and donor distribution across the laboratories

16 mm frozen abdominal skin membranes (Biopredic) donated post-surgery were used in the ring trial caffeine study. The skin membranes were dermatomed, with 300–400 μm thickness, without stretch marks, hair or marks and had been tested negative for human immunodeficiency virus (HIV), hepatitis C virus (HCV) and hepatitis B virus (HBV). The skin samples were delivered to the laboratories on dry ice and were stored at −20 °C until the conduct of the experimental runs. Fig. 1 shows the shipping times in 2020. Due to the Covid-19 pandemic there was a significant shipping delay for the samples shipped to laboratory "a", where the samples were stored in a storage center of a 3rd country. Laboratory f received the samples in two shipments (indicated as f and f rest) (see Fig. 1).

On the day of dosing or the day before, according to laboratories'

standard operation procedures (SOPs), skin membranes were removed from the freezer and brought at room temperature. The skin membranes were hydrated in physiological saline for about 10 min before mounting on the diffusion cells. The laboratories that removed the skin membranes from the freezer the day before dosing, followed the same procedure keeping the skin membranes mounted on the diffusion cells at approximately 32 °C. Skin integrity was assessed both visually and by different experimental methods (Fig. 2) depending on the laboratory default methods: Tritiated Water Flux (TWF, cut-off $>1.5 \times 10^{-3}$ cm/h, for laboratories a, d, f) or Trans Epidermal Electrical Resistance (TEER, cut off >17 kΩ, for laboratory c-cut-off >7.7 kΩ for laboratory e) or Trans Epidermal Water Loss (TEWL, acceptable range >0.5 and < 13 g $\times$ m^{-2} $\times$ h^{-1} i.e. cut-off < 13 g $\times$ m^{-2} $\times$ h^{-1} for laboratory b) tests. Although the respective integrity measures were determined, membranes not fulfilling the acceptance criteria were not excluded from the below shown comparative data-analysis. Instead, potentially impact of impaired skin integrity was addressed in the results interpretation.

A semi-automatic design of experiments for the assessment of inter-laboratory and intra-laboratory variability has been followed for the skin donor distribution across the laboratories. The study design was developed to fulfil the specific study requirements and facilitate the statistical data analysis. In particular, the design was suitable for analysis of the intra- and inter-laboratory variabilities whilst meeting the constraints of the number of experiments that can be conducted at each lab, and the number of skin samples that can be obtained from a single donor. However, while it was planned to distribute the samples between runs, this was not achieved by all laboratories, see Table 1.

Due to the scarcity of samples, *i.e.*, donated human skin, the study design is unbalanced (see Table 1). However, it was intended that each laboratory performed two experimental runs, utilizing 4 donors per experiment, each donor in duplicate. For the experimental repeat, the planned design considered that 2 out of the 4 donors were the same with the donors used in the first experimental run. Further, the donor distribution was aimed in a way that 3 labs tested the same donor in duplicate at least once. However, the intended design was not fully applied by all participating laboratories leading to some deviations as illustrated in Table 1.

2.3. Diffusion cells and receptor fluid

Flow-through diffusion cells with 0.64 cm^2 exposure area were used for the study. Laboratory b conducted the caffeine dermal absorption study with 1 cm^2 diffusion cells. The flow rate was maintained at 1.5 mL/h and the skin surface temperature was maintained at 32 ± 1 °C. PBS solution with 0.01% wt. sodium azide (NaN_3, Merck) was used as a receptor fluid.

2.4. Exposure time, skin washings and tape stripping

The exposure time was 8 h, during which the donor compartment remained un-occluded. The receptor fluid samples were collected at various time intervals (hourly intervals from 0 to 8 h after application and 2 h intervals until 24 h). Following completion of the 7–8 h sample of receptor fluid, the skin membrane surface was washed thrice with 100 μL of a 3% w/v soap solution (Estesol® Hair and Body, SC Johnson Professional GmbH, Krefeld Germany) by pipetting up and down, followed by twice gentle cleaning with liquid soap soaked Q-Tips. The washing solution was removed by skin membrane rinse using 100 μL tap water then dried with a further Q-tip. Following completion of 24 h, the skin membrane surface was washed as described before. Tape stripping was performed with D-Squame sampling discs (CuDerm Corporation). Each tape was pressed onto the skin surface with a pressure device (D-Squame Pressure Device) to achieve a standardized pressure of approximately 225 g/m^2 for approximately 5 s. All tape strips have been analyzed separately for radioactivity.

2.5. Compartment analysis

The amount of radioactivity was determined in the samples taken from the receptor fluid for the different time intervals, samples taken from the remaining receptor fluid after 24 h (the receptor fluid which is still left in the system at the time of study termination), extraction fluid from pipette tips and Q-tips obtained during washings and the washing fluid from skin washes, extracting fluid from tape strips, extraction fluid from the remaining skin preparations, extraction fluid from the receptor chamber, extracting fluid from the donor chamber.

2.6. Definitions

Depending on the regulation or guidance document, dermal absorption values can be differently defined. The following describes the definitions used in the manuscript.

- Receptor fluid is the amount recovered in receptor fluid, i.e. the receptor part of the dermal absorption cell, collected during the study conduct and at study termination after 24 h, including receptor cell wash.
- Potentially absorbed is the receptor fluid and amount recovered in dermis, epidermis and *stratum corneum* excluding tape-strips 1 and 2.

2.7. Manuscript software

The free statistical software R (R Core Team, 2020) was used for calculations and plots (Wickham, 2016) in this manuscript.

3. Results

Mean absorption results of the participating laboratories are sumarized in Table 2. Selected results from the ring trial are graphically assessed in the following. A graphical assessment allows an abstract evaluation of the overall response differences and between the participating labs. A more refined statistical assessment of the ring trial data is available in supplementary material 2.

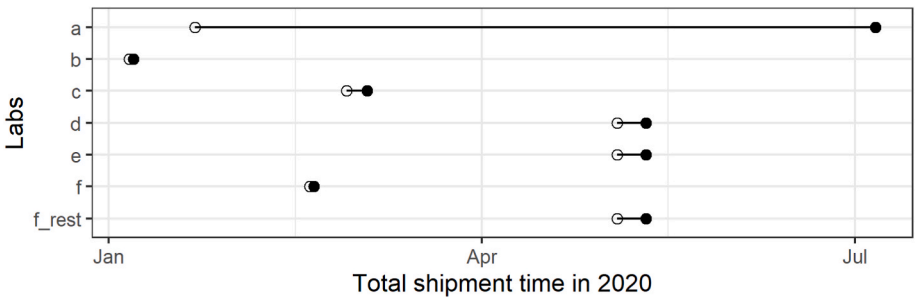

Fig. 1. Total shipment time in 2020 from Biopredic to the laboratories.

3.1. Skin integrity

Results of the skin integrity tests of the supplied skin samples according to internal laboratory validity criteria are shown in Fig. 2. While several samples failed the individual labs' quality assessment, the samples were used for the ring trial, as it was hypothesized that the different quality criteria could have influenced assay outcome. The difference of acceptability criteria is most pronounced in run 2 of lab c and e, which all show similar TEER values. For example, lab c would have discarded most of the received samples, while lab e received mostly acceptable samples, according to their internal TEER thresholds. Further, interestingly, about half of the samples for run 1 in lab a were below the quality criterion while those of run 2 were mostly acceptable. Adherence to acceptability criteria of individual donors over laboratories is discussed below.

3.2. Kinetics

The cumulative absorption into the receptor fluid over 24 h assay duration is shown in Fig. 3. Receptor cell rinse or recoveries in other compartments is only assessed after the final 24 h time point. The cumulative absorption pattern shows the kinetics of the absorption process. It also allows the identification of outliers, e.g. due to insufficient sealing of the donor compartment, which may allow rapid penetration in the receptor fluid without penetrating the skin membrane.

Notable is the difference in the kinetic pattern between laboratory a and the other laboratories. A CLE current project on the interpretation of dermal absorption studies discusses options to detect and handle outliers (Kluxen et al., unpublished). In short, outliers are values that deviate substancially in response from other replicates in the same experimental treatment group. This deviation may occur due to an experimental error or a "normal" but extreme response, which however only occurs with a low probability in random sampling. Either way, outliers skew the general response pattern and thus obscure general underlying effects.

All laboratories and runs contain outlier responses that substancially deviate from the others.

Obvious outliers are.

- cell 5 in the first experimental study (b1) (ID b1.5),
- cell 7 in the repeat study (b2) (ID b2.7).

These two outliers probably occur due a mistake during washing, as the penetration spikes at 8 h (ID numbers are generated by considering lab, run number and cell number).

Also, the high response in run 2 of lab c (ID c2.4) is noticeably different from the rest of the values in the lab, which occurs directly after study initiation and probably indicates insufficient sealing. Similarly, the two high responses in laboratory f (IDs f2.1, f2.5), already show substantial receptor fluid penetration at study initiation. Further, the two high responses from run 1 in laboratory d (IDsd1.4, d1.3) show a very different absorption pattern, as compared to the rest of the responses.

While all noted responses may affect individual bioassay evaluation, and would presumably be excluded from individual assay evaluations, only the outliers in lab b were excluded from calculating summary statistics for receptor fluid recovery and potential absorption in this ring trial. The reason for this is that the high responses observed in the studies of the other labs are within range of the responses in the other labs.

The outliers are excluded from calculating the overall summary statistics but were included in the plots of the following sections.

Fig. 4 shows the maximum flux calculated based on the slope of the linear portion of the cumulative caffeine mass in the receptor fluid, excluding the lag time and the plateau, as recommended by OECD TG 428. The mean maximum flux for lab a was 2.66 ± 0.79 μg $cm^{-2}{\cdot}h^{-1}$ for the 1st experimental run (R1) and 1.78 ± 0.79 μg $cm^{-2}{\cdot}h^{-1}$ for the second experimental run (R2). The maximum mean fluxes of laboratory a in both experiments are higher than maximum mean fluxes reported by the rest of the laboratories, which is 0.34 μg $cm^{-2}{\cdot}h^{-1}$.

3.3. Overview of compartment recovery distribution

Fig. 5 shows the overall residue distribution in the various compartments that are assessed in a dermal absorption study by individual cell. The figure shows a comparable distribution pattern between all laboratories except laboratory a, which shows a higher amount in receptor fluid and skin associated material. Individual extreme values can be identified in several runs and can be related to those identified earlier.

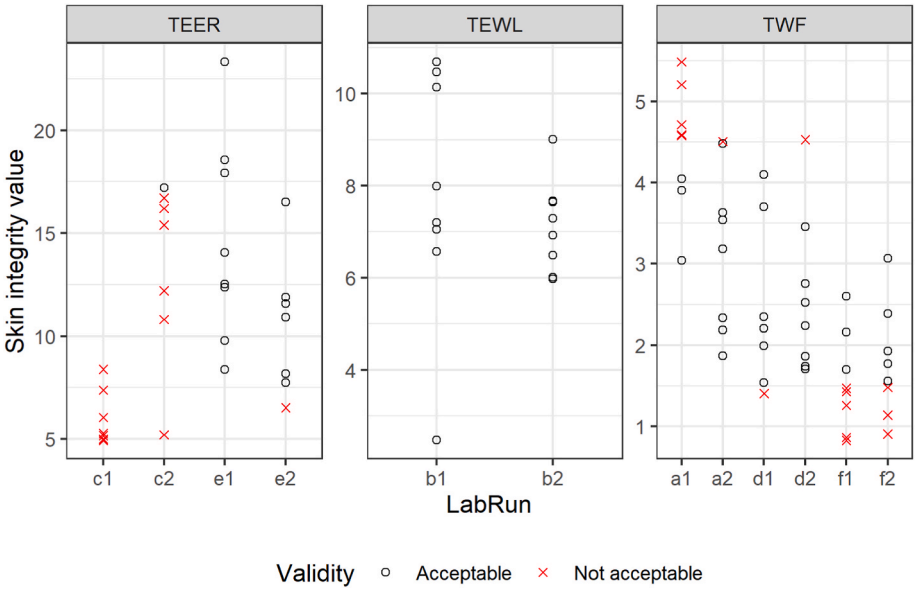

Fig. 2. Results of the skin integrity tests of the supplied skin samples according to internal laboratory validity criteria. While several samples failed the quality assessment, the samples were used for the ring trial. TWF = Tritiated Water Flux; TEER = Trans Epidermal Electrical Resistance; TEWL = Trans Epidermal Water Loss.

Table 1
Skin sample allocation between laboratories.

Lab	Run	Donor											
		1	2	3	4	5	6	7	8	9	10	11	12
a	R1	××××	××××										
	R2			××	××	××	××						
b	R1			××	××			××	××				
	R2					××	××	××	××				
c	R1			××	××			××	××				
	R2			××	××				××	××			
d	R1	×	×								××	×××	×
	R2	×	×								××	×	×××
e	R1							××	×		×××	×	×
	R2								×	×	××××	×	×
f	R1	××				××	××			××			
	R2		××			××	××			××			

Table 2
Absorption results of the participating laboratories.

Laboratory	Experimental Run	Mean Receptor fluid (%) after 24 h (SD)	Mean Receptor fluid + skin + tape strips 3-x (%) after 24 h (SD)
a[a]	R1	22.09 (5.38)	22.77 (5.44)
	R2	21.68 (7.17)	22.13 (7.25)
	Mean (R1, R2)	21.88 (6.13)	22.44 (6.20)
	Historical data 1 run with 8 replicates	5.08% (1.30)	5.84% (1.53)
b	R1	2.80 (2.48)	2.92 (2.56)
	R2	2.05 (0.60)	2.17 (0.70)
	Mean (R1, R2)	2.42 (1.78)	2.55 (1.85)
c	R1	3.40 (1.28)	3.46 (1.31)
	R2	3.58 (2.46)	3.65 (2.46)
	Mean (R1, R2)	3.48 (1.90)	3.56 (1.91)
d	R1	4.72 (3.06)	4.78 (3.13)
	R2	3.41 (1.20)	3.41 (1.20)
	Mean (R1, R2)	4.07 (2.35)	4.09 (2.39)
e	R1	5.54 (5.67)	5.75 (6.00)
	R2	6.22 (3.35)	6.23 (3.35)
	Mean (R1, R2)	5.86 (4.58)	5.98 (4.79)
f	R1	3.06 (0.76)	3.06 (0.76)
	R2	4.07 (2.87)	4.08 (2.88)
	Mean (R1, R2)	3.56 (2.09)	3.57 (2.10)

[a] The systematically deviating results obtained for laboratory a are attributed to a COVID-19 associated impaired skin sample shipment. The shown lab internal data historical data are comparable to the data of the other laboratories obtained in this ring-trial.

3.4. Mass balance

The total mass balance is shown in Fig. 6. Mass balance is a quality criterion used in dermal absorption studies, which must achieve certain limits for the study being considered acceptable. For studies using radioactive material, OECD TG 428 requires a mass balance between 90 and 110% and EFSA considered 95–105% necessary for studies investigating pesticides.

Before application, the labelled material is assessed by scintillation counting and an appropriate dose calculated based on the scintillation results, which is then considered to be 100%. However, since less material can be applied than pulled into a pipette, a certain inherent loss by pipetting maybe relevant at low doses where study analytics in such studies are close to the border of detection. Often less material can be extracted from the used materials like biological membranes, swabs, test-system surfaces. Therefore, recovery should be on average $< 100\%$. Further, experimental variation affects mass balance. When pipetting the intended dose, the experimenter aims to fully apply and evenly spread the dose over the entire skin surface within the diffusion cell, but this is subject to individual variation. Also, the skin washing procedure albeit aimed to be harmonized as much as possible is dependent on the experimenter's behavior.

All laboratories achieve mean mass balances $>95\%$, while the overall mass balance is below 100%, which is expected if insufficient dosing is assumed (Kluxen et al., 2019). The mass balance variation, when determined by the 25th to 75th interquartile size (box size in the plot), was notably higher in run 1 of laboratory f, when compared to the other results. Also, the variation in laboratory a was slightly higher than in the other laboratories, but on average very close to 100%. Overall, this indicates that the laboratories have the experimental procedures well in control.

3.5. Absorption

While the outliers identified in the kinetic penetration profiles of run b1 and b2 were included in the figures investigating receptor fluid recovery (Fig. 7) and potentially absorbed dose (Fig. 8), they are clearly distinct from most of the responses within the run and between groups. Similarly, the curiously shaped penetration profiles of c2 and d1 (in previous Figs. 3 and 4) are identified as outliers in the plots.

The figures also show the huge difference of laboratory's a response

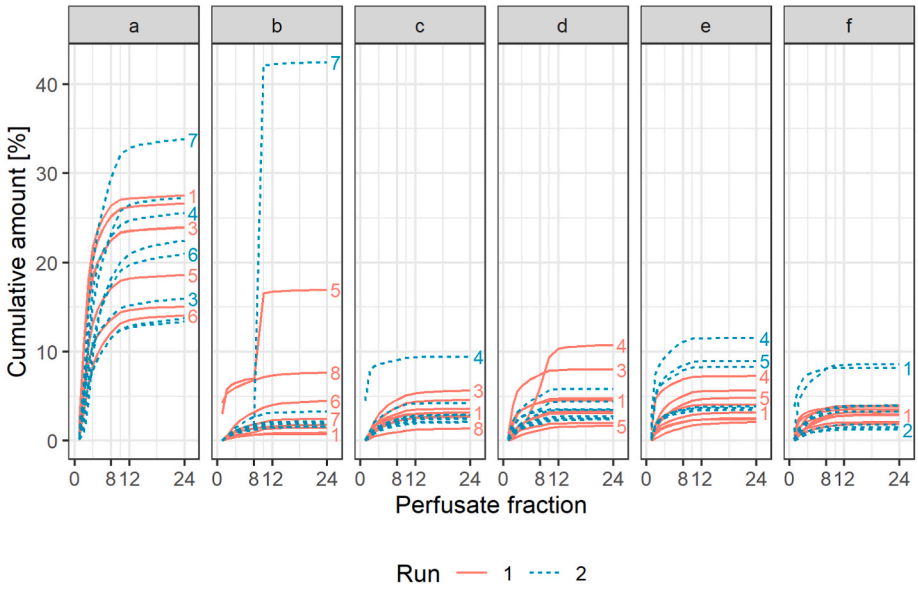

Fig. 3. Cumulative amount penetrated in receptor fluid [in % applied] over time. Overall, the runs were very consistent between labs. Outlier responses can be clearly identified. The response from laboratory a is very different from the profiles observed in the other labs. Cell numbers are added after the final observation time point (not all numbers can be shown due to overplotting).

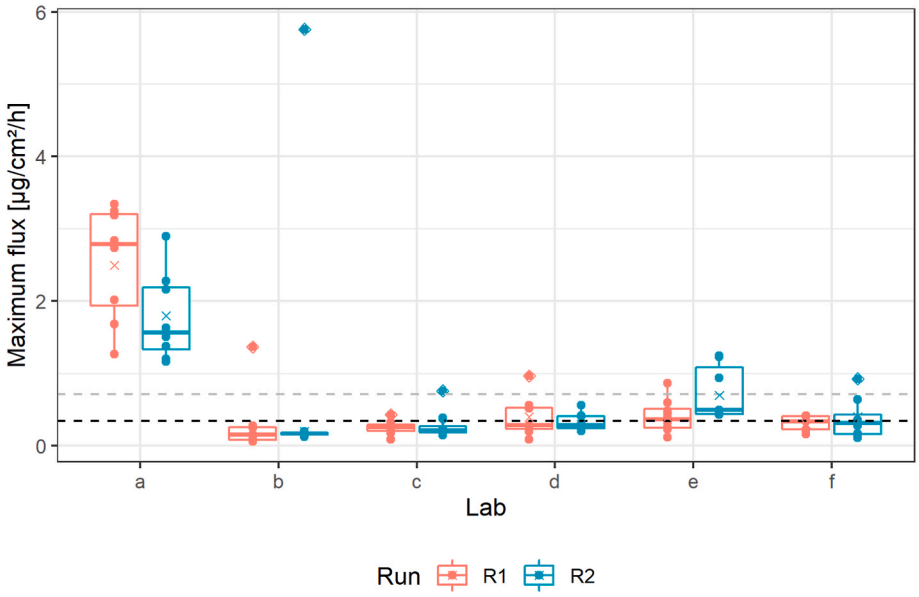

Fig. 4. Comparison of maximum flux [in µg/cm^2/h] between laboratories and runs (red/blue), as individual values (dots), boxplots and means (multiplication sign). Boxplot outliers are highlighted with a diamond shape. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

towards the responses from the other laboratories.

Curiously the high response in one replicate of e1 (ID e1.4 see Figs. 5 and 7) does not occur in its corresponding kinetic profile after 24 h (Figs. 3 and 4). It thus stems from the post sampling receptor cell wash step. Its relevance for a reliable dermal absorption estimate may thus be scrutinized and one may consider this replicate to be an outlier, as well. Figs. 7 and 8 are very similar as the contribution of the skin fraction is overall negligible <1%.

3.6. Donor effect

To understand why the results in laboratory a are different to those of the others, the potential absorption was plotted by donor and colored by laboratory, see Fig. 9. A red rectangle was added to the plot to highlight that the results of laboratory a are consistently higher than those of the other laboratories, independent of the donor.

The donor effect is further investigated in Fig. 10 by faceting laboratories over donors and vice-versa. It shows that the laboratories achieve very consistent results independent of the donor used. However, all donors show individual outliers, when compared to the rest of the data distributed over laboratories. Interestingly, those values are not different to the other cells by the respective laboratory's acceptance criteria. Only donor 3 has consistently higher absorption values than the other donors independent of laboratory. This may indicate that individual donors have on average different absorption properties, however, also this donor shows an outlier. Here, the very high response for donor

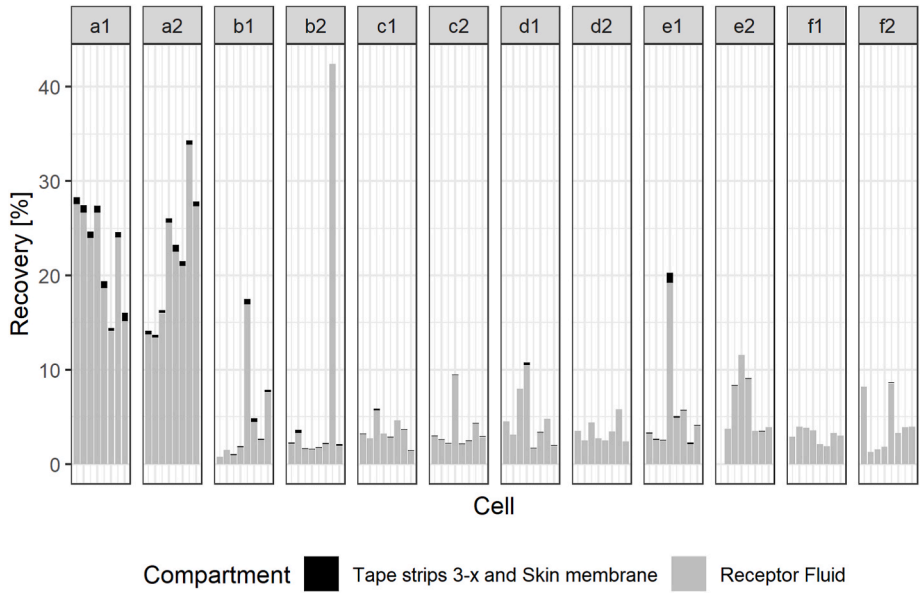

Fig. 5. Distribution profile of recovery [in % applied] in Tape strips 3-x (black), skin membrane (black) and receptor fluid (grey). Tape strips and skin membrane are not further distinguished because of the very low amounts recovered. The profile of lab a is noticeably different to the other profiles in that the amount recovered in receptor fluid and skin membrane is much higher. Individual extreme values can be clearly identified.

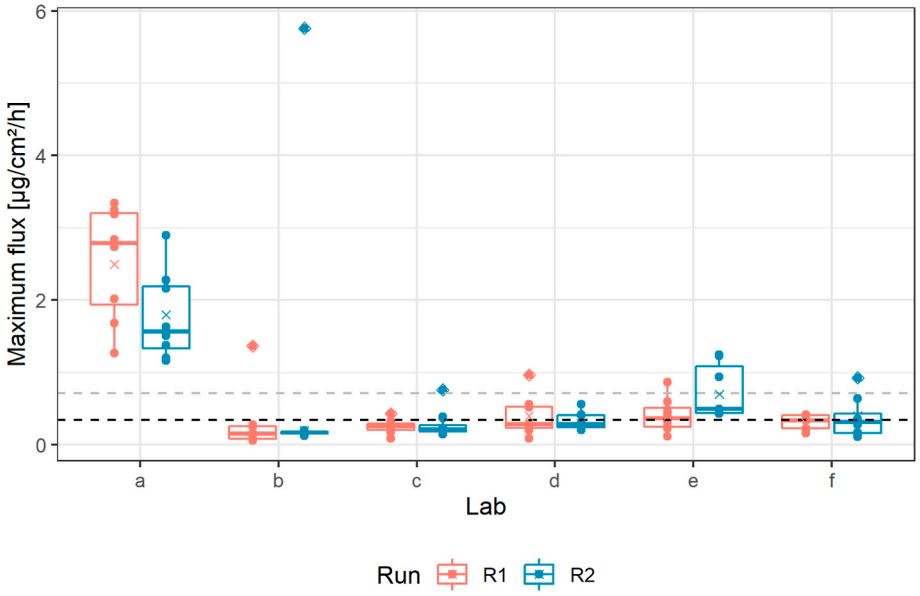

Fig. 6. Mass balance [% applied] between laboratories and runs (red/blue), as individual values (dots), boxplots and means (multiplication sign). Boxplot outliers are highlighted with a diamond shape. Very high mass balance was achieved by all laboratories. Only run 1 in laboratory f was notably different, however, still within test guideline requirements. Mass balance of laboratory a is more variable between the cells than in the other laboratories, otherwise the response is not noticeable different. Dashed horizontal grey line: overall mass balance; black line laboratory a excluded. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3 occurs to a high receptor fluid wash recovery and not due to amount penetrated into the receptor fluid within 24 h and might thus occur due to an experimental error.

Further, interesting are the differences between donor 7 and 9, where the laboratories respective acceptability criteria would result in very different dermal absorption estimates. Further, donor 12 was not acceptable to either laboratory but resulted in similar absorption values of the other laboratories.

3.7. Compatibility of responses between laboratories

To formally investigate whether the results between laboratories and runs are comparable with a grand mean, an analysis of means (Ott, 1983) was performed on a linear mixed effects model, using the R packages lmer (Bates et al., 2015), multcomp (Hothorn et al., 2008) and ANOM (Pallmann and Hothorn, 2016). It is investigated whether the means of the potentially absorbed dose in the different laboratories is compatible to the grand mean over all laboratories, i.e., whether the laboratories report similar dermal absorption results for the penetration of ^{14}C- caffeine.

For this analysis the values from laboratory "a" were excluded, and similarly values that are indicated to be outliers throughout the manuscript, i.e., values with the IDs b1.5, b2.7, c2.4, d1.3, d1.4, e1.4, f2.1, f2.5 (see Fig. 11A for the resulting dataset). The mixed-effects model considers "Runs" as the random effect, i.e. a correlated slope and intercept with the formula potentially. absorbed ~ Lab + (Lab |Run). The obvious weakness of the model is that the random factor has only two levels and is strongly correlated. However, the model's residuals appear to be normal distributed when investigated by graphical analysis (plot not shown).

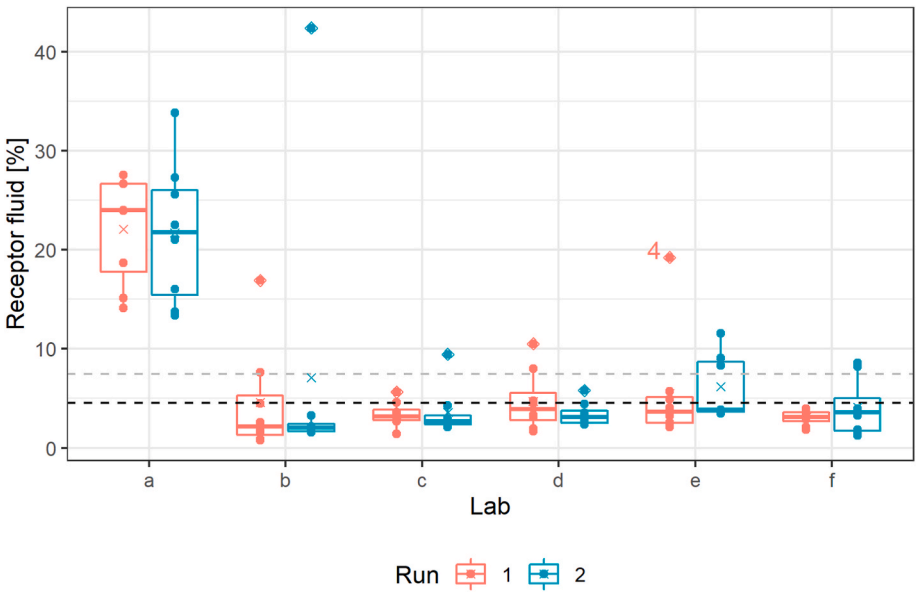

Fig. 7. Receptor fluid recovery [% applied] between laboratories and runs (red/blue), as individual values (dots), boxplots and means (multiplication sign). Boxplot outliers are highlighted with a diamond shape. Very similar receptor fluid values are observed between labs and runs, with the exception of the results from lab a. Dashed horizontal grey line: overall receptor fluid value; black line laboratory a excluded. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

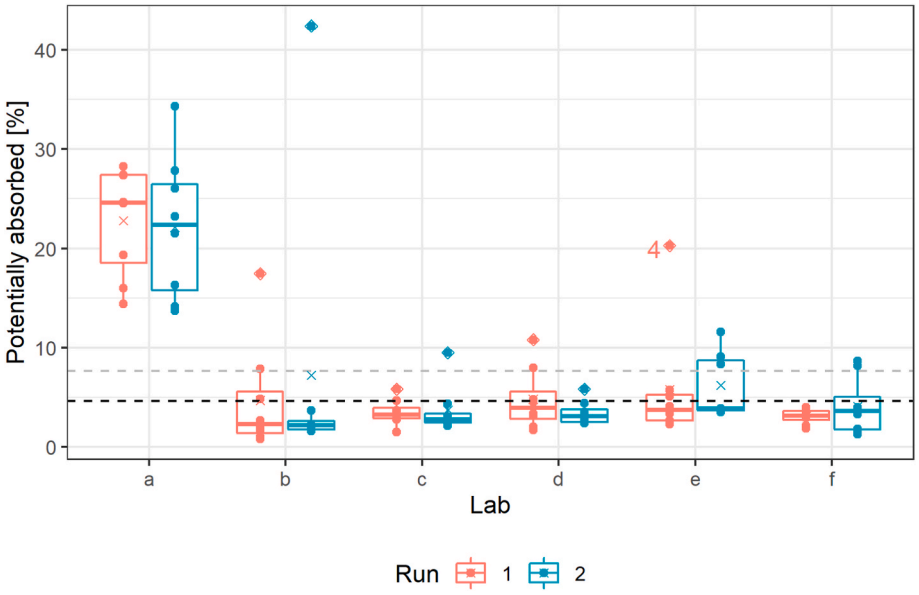

Fig. 8. Recovery [% applied] in receptor fluid, skin without tape strips 1 and 2, i.e. potentially absorbed dose, between laboratories and runs (red/blue), as individual values (dots), boxplots and means (multiplication sign). Boxplot outliers are highlighted with a diamond shape. Since skin residue ≪1% the values are almost identical to the receptor fluid values. Dashed horizontal grey line: overall potentially absorbed value; black line laboratory a excluded. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

The analysis is presented in Fig. 11B. It shows that the results of all laboratories are comparable to the grand mean of 3.37%, when runs are considered as a random effect.

Further, the laboratories have compatible means when compared to each other with a Tukey test (Tukey, 1949) on the same mixed-effect model (data not shown).

4. Discussion

The 2nd ring trial demonstrates that the OECD TG 428-compliant *in vitro* assay can be robustly and reliably performed in different laboratories. This applies for both the intra-laboratory as well as the inter-laboratory reproducibility (Table 2). Within all the participating laboratories the outcome of the first and second study repeat was very consistent and without evidence for significant donor driven variability. Also, when comparing the inter-lab repeatability, a very good concordance was evident except for one laboratory.

Due to a COVID-19 pandemic-associated delay in shipment and thus induced systematic difference in laboratory "a", the experiment further indicates that skin associated factors other than internal integrity testing may affect absorption and variation. This is also evident when individual donor absorption is compared over laboratories. While one out of 12 donors shows a notably higher absorption than the others, all donors showed outlier responses independent of the respective laboratory's acceptability criteria. These skin integrity determinations are routinely lab specific and differ with measurement device, cell size and measurement technique as detailed in the labs standard operation procedure applied. This indicates again that the current acceptability criteria may have to be reviewed. This observation is in line with comparative investigations on different skin integrity tests reported by Guth et al. (2015). Cells that are acceptable in one laboratory may be deemed unacceptable in another, while the absorption results are comparable. Hence, it seems unclear how the current acceptability criteria relate to the biological relevance of the observed absorption values.

A factor that is generally considered major contributor to high variability during dermal absorption studies is the skin tissue (Finnin

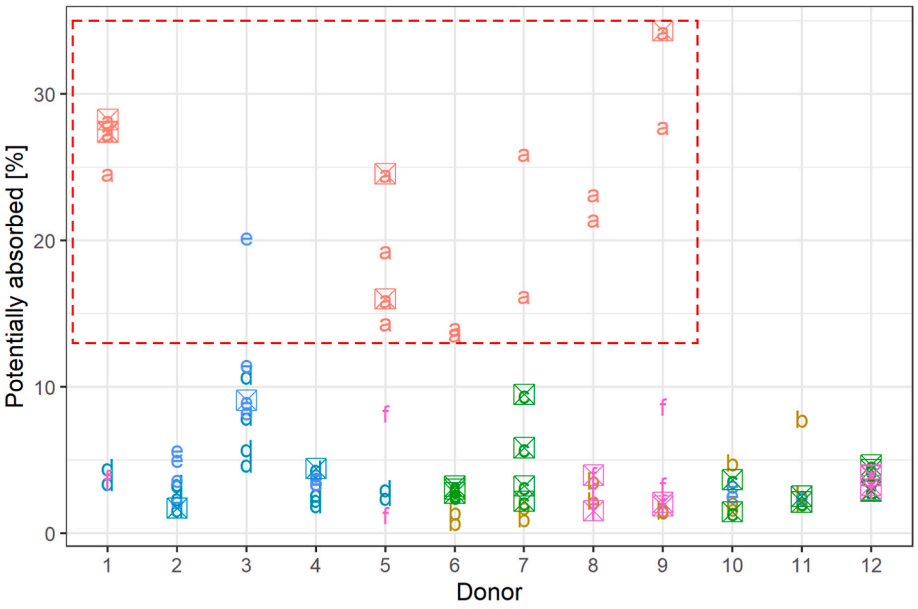

Fig. 9. Potentially absorbed material [% applied] between donors and colored by labs. The red rectangle with dashed lines highlights responses from lab a, which are consistently higher than those of the other labs. Responses from cells that did not achieve internal validity criteria are depicted overplotted with a crossed-out box. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

et al., 2012; van de Sandt et al., 2004). For the 2nd ring trial, abdominal skin tissues from 12 donors sourced from a single supplier have been utilized across 6 laboratories, with skin samples from the same donor distributed across 2 or 3 laboratories. It should be stated that the skin characteristics (abdominal origin, split-thickness skin prepared by the skin supplier, Caucasian female donors) were standardized for the study. Statistical evaluation of the donor effect on the caffeine dermal absorption showed that when the outliers are excluded, skin donor significantly contributed to the result variability.

While some donors may show generally higher absorption patterns (in the ring trial only 1 of 12 donors, i.e., donor 3), extreme absorption values are observed in almost all laboratory runs and independently of donors. Hence, isolated high responses need to be carefully investigated before including them into a dermal absorption estimate. The ring trial shows consistent responses when the isolated extreme values are excluded from an overall absorption value.

Laboratory a also participated the previous 1st ring trial (data provided in the Supplement), where it was designated as laboratory "i". In the 1st ring trial, the laboratory produced the lowest absorption values of all participating laboratories. Further, the laboratory's internal data on ^{14}C-Caffeine are compatible with the results of the 2nd ring trial for laboratories b-f (see Fig. 5 in the supplementary material, historical internal data is indicated as a circle-cross symbol). This again indicates that skin quality may be the driving factor for the results of laboratory a in the current ring trial and that its performance is robust. Fig. 5 in the supplementary material shows further that the variation in the 2nd ring trial was substantially decreased compared to the 1st ring trial and that the variation within the 1st ring trial is grouped in a higher (labs a, c, d, e) and a lower response group (labs b, f - l). The higher response group in the 1st ring trial is compatible with the high response group in the 2nd ring trial, where an effect on skin quality can be inferred from the systematic difference in shipment.

There is currently no robust methodology in place to compare the skin quality between laboratories except for standard skin integrity testing (Guth et al., 2015; Schäfer-Korting et al., 2008). Hence, it is unclear whether the high response group in the 1st ring trial is potentially also associated with skin sample quality, like the high responses in the 2nd ring trial.

There are currently no high quality intra- and interlaboratory studies on dermal absorption publicly available that can be directly compared to our results. The *in vitro* dermal absorption method was not formally validated, when developed and implemented as an OECD test guideline in 2004 (OECD, 2004). In the multicenter comparison study of van de Sandt et al., (2004), many variants were not harmonized. All participating laboratories did the study according to their own procedures. Differences lie for example with the skin origin because both post-mortem and surgical waste material was tested. Further, skin thickness ranged from 300 to 1800 μM as both split thickness and full thickness skin were investigated. Flow-through and static diffusion cells systems were used with exposed skin areas from 0.3 to 3.14 cm^2. Further, the number of replicates per test-run and/or lab and the number of donors was not standardized, and different detection systems were used (LSC and HPLC-UV). Moreover, the exposure duration in the van de Sandt study was 24 h, i.e. different to the meanwhile recommended duration for pesticidal products testing, and did not include a skin wash after 6–10 h. Also, the applied volume was 25 μL/cm^2 instead of the recommended 10 μL/cm^2. Due to the 24-hr exposure and the multiple differences the reported caffeine results differ significantly from what is reported here.

The later reported interlaboratory study (Schäfer-Korting et al., 2006, 2008) did not aim to investigate methodological aspects relevant for GLP studies for pesticidal products. It investigated the applicability of reconstructed skin models for dermal absorption testing in comparison to human or pig split thickness skin and the study design focused on permeation assessment only. Both finite and infinite exposure were assessed, but no skin wash was implemented. Receptor fluid was compared at 6- and 24-hrs collection timepoints only. No mass-balance was conducted. Thus, the study does not allow a comparison to this ring trial.

The interlaboratory comparative studies using human and pig skin (Gerstel et al., 2016; Liu et al., 2018; Wargniez et al., 2017) focused on the species comparison and a high qualitative analytical method by HPLC-MS to be consistently applied. It is however noteworthy that the conducted in-depth analytical method validation and the subsequent assessment of study techniques allows some very valuable insights into methodological aspects of application protocol, washing protocol and tape-stripping procedure (Wargniez et al., 2017).

Overall, since the initial OECD guideline development, criteria on

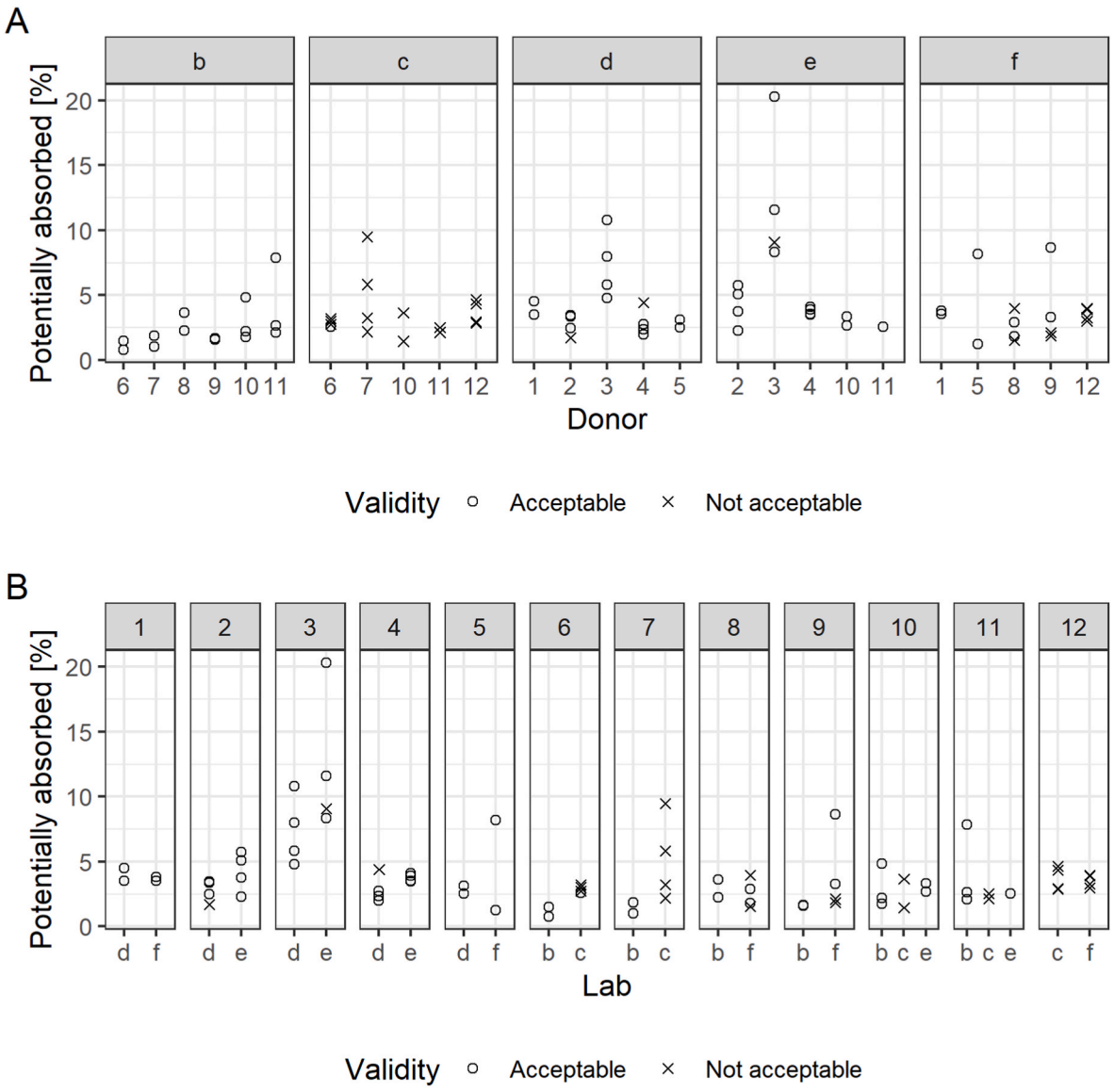

Fig. 10. Similar to Fig. 9, but A) laboratories are shown in facets b-f or B) donors are shown in facets 1–12. The high response for donor 3 in laboratory e (ID68) occurs due to a high receptor fluid wash value.

study conduct and interpretation have evolved (Fabian et al., 2017; Wargniez et al., 2017). The routinely generated registration studies today, which correspond to the studies generated in the 2nd ring trial, likely differ in many details from the procedure at the time the guideline was developed and released (Aggarwal et al., 2019; Fabian et al., 2017; Hewitt et al., 2020; Heylings et al., 2018; Wargniez et al., 2017). This is also reflected in recent reviews expressing the need of a guideline revision (So et al., 2014; Sullivan et al., 2017; Hopf et al., 2020).

Several laboratories have reported individual study results with the reference compounds mentioned in the OECD test guideline. Their intend was usually to assess particular aspects of the methodology or assessing the impact of vehicle components and thus not necessarily aiming to apply an across laboratories routinely standardized methodology (Abd et al., 2019; Dreher et al., 2002; Guth et al., 2015; Hewitt et al., 2020; Heylings et al., 2018; Hui and Maibach, 2020; Im et al., 2021; Wilkinson et al., 2006). Therefore, it is not surprising that reported caffeine absorption values in % of the applied dose vary generally in the literature.

Even the recently reported caffeine testing in the cosmetic component database (Hewitt et al., 2020), where a standardized routine methodology was applied, is not directly comparable to the above reported results. While most of the methodological details are well in line, the major difference is that no 8-hr skin wash was included in the study design and thus the here reported results refer to a 24-hr exposure period. In addition, a different concentration of 0.1 mg/L instead of the tested 4 mg/L was applied. Consequently, it is not surprising that the absorbed dose with about 40% of the applied dose is much higher than what was determined in our study. Nonetheless, the results of this recent Cosmetics Europe project also demonstrate good reproducibility of dermal absorption estimates with OECD TG 428 for 56 cosmetic-related chemicals.

Caffeine is a small hydrophilic and thus the understanding obtained from this ring-trial with caffeine cannot easily be transferred without further investigation to more hydrophobic compounds like testosterone, bigger molecules or even complex mixtures. There is likely further demand to conduct such or similar comparative studies to support the conclusions drawn in here, to elaborate the impact of potential methodological aspects on variability of study outcome and to improve the method as such.

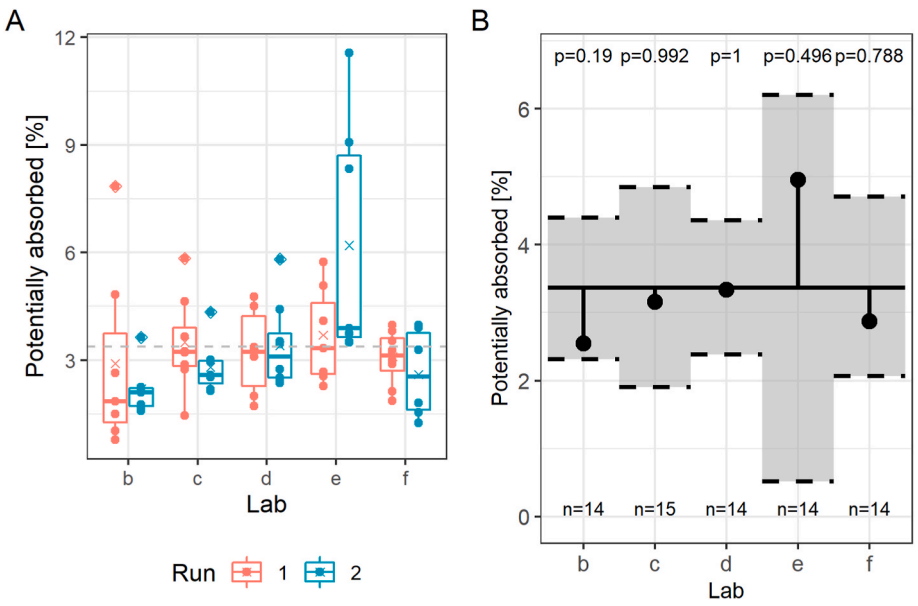

Fig. 11. Analysis of means on potentially absorbed dose of a refined dataset, excluding additional outliers as highlighted in the manuscript. A) Boxplot similar to Fig. 1, B) analysis of means on a mixed effect model with considering run as a random factor. The grey area shows the region of compatibility with the grand mean.

5. Conclusion

The ring trial demonstrates that the OECD TG 428 assay can be robustly and reliably performed by different laboratories. However, the experiment indicates that other skin associated factors than internal integrity testing may affect absorption and variation. Skin quality criteria not routinely assessed or considered in the current method may explain dermal absorption variations observed in practice regulatory testing, i.e., outliers within a dataset or differences between datasets. Hence, it is suggested to explore further quality control measures for the used skin samples.

Funding body information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. CLE funded the project management by the University of Surrey. The laboratories participating conducted the studies on their own expense. The skin membrane provision and distribution was however funded by CLE to allow a supervised and controlled distribution across the participating laboratories. The research was conducted as scientific expert contribution of the authors to the above-mentioned industry associations CLE or of the participating laboratories as part of their work to their respective affiliations.

CRediT authorship contribution statement

Felix M. Kluxen: Conceptualization, Visualization, Formal analysis, Writing – original draft, Writing – review & editing. **Styliani Totti:** Conceptualization, Visualization, Data curation, Formal analysis, Investigation, Project administration, Writing – original draft, Writing – review & editing. **Wilfred Maas:** Investigation, Resources. **Frank Toner:** Investigation, Resources. **Leanne Page:** Investigation, Resources. **Kathryn Webbley:** Investigation, Resources. **Rajendra Nagane:** Investigation, Resources. **Robert Mingoia:** Investigation, Resources. **Christine Whitfield:** Investigation, Resources. **John Kendrick:** Investigation, Resources. **Claire Valentine:** Investigation, Resources. **Jeanne Bernal Dorange:** Investigation, Resources. **Camille Egron:** Investigation, Resources. **Camille Imart:** Investigation, Resources. **Jeanne Y. Domoradzki:** Conceptualization, Writing – review & editing, Bob Parr-Dobrzanski: Conceptualization, Supervision. **Philip Fisher:** Conceptualization, Writing – review & editing. **Christine Lorez:** Writing – review & editing. **Steve McEuen:** Conceptualization, Writing – review & editing. **Edgars Felkers:** Conceptualization, Writing – review & editing. **Tao Chen:** Conceptualization, Supervision, Writing – review & editing. **Christiane Wiemann:** Conceptualization, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Christiane Wiemann, Felix M. Kluxen, Jeanne Y. Domoradzki, Edgars Felkers, Philip Fisher, Christine Lorez, Steven McEuen and Bob Parr-Dobrzanski are employees of agrochemical companies that conduct and evaluate dermal absorption studies in vitro through human skin for regulatory purposes in the context of authorization and marketing of their companie's products. Wilfred Maas, Frank Toner, Leanne Page, Kathryn Webbley, Rajendra Nagane, Robert Mignoia, Christine Whitfield, John Kendrick, Clair Valentine, Jeanne Bernal Dorange, Camille Egron and Camille Imart are employees from the laboratories that offer study services for this study type either within the industrial companies in charge or as contract research providers.

They contribute as scientific experts to the project in terms of study planning, study conduct, data evaluation and data presentation with the aim of improving the study methodology.

Acknowledgements

The authors would like to thank the participating laboratories of the 1st ring trial, BASF SE Biokinetic Laboratory of Experimental Toxicology and Ecology, Charles River Laboratories Den Bosch and Charles River Laboratories Edinburgh, Labcorp Drug Development, CXR Biosciences, Dermal Technology Laboratory, Envigo CRS, Eurofins Agroscience Services Chem SAS, Innovative Environmental Services, Jai Research Foundation, Pharmaron UK and Triskelion B.V.

Felix M. Kluxen, Jeanne Domoradzki, Philip Fisher, Christine Lorez, Edgars Felkers, Bob Parr-Dobrzanski, Steven McEuen and Christiane Wiemann (chair) are members of the CLE Dermal absorption technical subgroup. Wilfred Maas, Frank Toner, Leanne Page, Kathryn Webbley,

Rajendra Nagane, Robert Mignoia, Christine Whitfield, John Kendrick, Clair Valentine, Jeanne Bernal Dorange, Camile Egron and Camile Imart are study directors and scientists from the laboratories that conducted the studies. Laboratories were provided with skin membranes purchased by CLE to allow a supervised stratified distribution about the participating laboratories. Study conduct in term of test system and personal contribution was done on the individual laboratories own expense. Styliani Totti and Tao Chen managed the project, monitored the dermal absorption studies, collected the data, and conducted the results analysis and were reimbursed via a CLE project fund. This research did otherwise not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.yrtph.2022.105184.

References

Abd, E., et al., 2019. Permeation mechanism of caffeine and naproxen through in vitro human epidermis: effect of vehicles and penetration enhancers. Skin Pharmacol. Physiol. 32, 132–141.
Aggarwal, M., et al., 2019. Assessing in vitro dermal absorption of dry residues of agrochemical sprays using human SKIN within OECD TG 428. Regul. Toxicol. Pharmacol. 106, 55–67.
Bates, D., et al., 2015. Fitting linear mixed-effects models using *lme4*. J. Stat. Software 67, 1–48.
Dreher, F., et al., 2002. Improvement of the experimental setup to assess cutaneous bioavailability on human skin models: dynamic protocol. Skin Pharmacol. Appl. Skin Physiol. 15 (Suppl. 1), 31–39.
EFSA, 2017. Guidance on dermal absorption. EFSA J. 15, 4873.
EFSA, et al., 2017. Guidance on dermal absorption. EFSA J. 15, e04873.
EFSA, 2012. SCIENTIFIC OPINION guidance on dermal absorption EFSA panel on plant protection products and their residues (PPR). EFSA J. 10.
Fabian, E., et al., 2017. A protocol to determine dermal absorption of xenobiotica through human skin in vitro. Arch. Toxicol. 91, 1497–1511.
Finnin, B., et al., 2012. In vitro skin permeation methodology. In: Benson, A.E., Watkinson, A.C. (Eds.), Transdermal and Topical Drug Delivery: Principles and Practice. John Wiley & Sons, pp. 85–108.
Gerstel, D., et al., 2016. Comparison of protocols for measuring cosmetic ingredient distribution in human and pig skin. Toxicol. Vitro 34, 153–160.
Guth, K., et al., 2015. Suitability of skin integrity tests for dermal absorption studies in vitro. Toxicol. Vitro 29, 113–123.
Hewitt, N.J., et al., 2020. Measurement of the penetration of 56 cosmetic relevant chemicals into and through human skin using a standardized protocol. J. Appl. Toxicol. 40, 403–415.
Heylings, J.R., et al., 2018. Dermal absorption of testosterone in human and pig skin in vitro. Toxicol. Vitro 48, 71–77.
Hopf, N.B., et al., 2020. Reflections on the OECD guidelines for in vitro skin absorption studies. Regul. Toxicol. Pharmacol. 117.
Hothorn, T., et al., 2008. Simultaneous inference in general parametric models. Biom. J. 50, 346–363.
Hui, X., Maibach, H., 2020. In vitro human skin percutaneous penetration: does a second topical application effect flux of first application? J. Dermatol. Treat. 1–6.
Im, J.E., et al., 2021. Effect of application amounts on in vitro dermal absorption test using caffeine and testosterone. Pharmaceutics 13.
Kluxen, F.M., et al., 2021. Compounded conservatism in European re-entry worker risk assessment of pesticides. Regul. Toxicol. Pharmacol. 121, 104864.
Kluxen, F. M., et al., unpublished. Interpretation of in Vitro Dermal Absorption Studies.
Kluxen, F.M., et al., 2019. Dermal absorption study OECD TG 428 mass balance recommendations based on the EFSA database. Regul. Toxicol. Pharmacol. 108, 104475.
Liu, Y., et al., 2018. Inter-laboratory study of the skin distribution of 4-n-butyl resorcinol in ex vivo pig and human skin. J. Chromatogr. B: Analytical Technologies in the Biomedical and Life Sciences 1093–1094, 77–79.
OECD, 2004. Test No. 428: Skin Absorption: in Vitro Method. OECD Publishing, Paris.
OECD, 2011. Guidance notes on dermal absorption No. 156. OECD Publishing, Paris.
Ott, E.R., 1983. Analysis of means—a graphical procedure. J. Qual. Technol. 15, 10–18.
Pallmann, P., Hothorn, L.A., 2016. Analysis of means: a generalized approach using R. J. Appl. Stat. 43, 1541–1560.
R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.
SCCS, 2010. Basic Criteria for the in Vitro Assessment of Dermal Absorption of Cosmetic Ingredients. Vol. SCCS/1358/10.
Schäfer-Korting, et al., 2006. Reconstructed human epidermis for skin absorption testing: results of the German prevalidation study. ATLA 34, 283–294.
Schäfer-Korting, M., et al., 2008. The use of reconstructed human epidermis for skin absorption testing: results of the validation study. Atla-Alternat. Lab. Anim 36, 161–187.
So, J., et al., 2014. Comparison of international guidelines of dermal absorption tests used in pesticides exposure assessment for operators 30 (4), 251–260. http://dx.doi.org/10.5487/TR.2014.30.4.251.
Sullivan, K.M., et al., 2017. Dermal absorption for pesticide health risk assessment: harmonization of study design and data reporting for North American Regulatory submissions. Regul. Toxicol. Pharmacol. 90, 197–205.
Tukey, J.W., 1949. Comparing individual means in the analysis of variance. Biometrics 5, 99–114.
van de Sandt, J.J., et al., 2004. In vitro predictions of skin absorption of caffeine, testosterone, and benzoic acid: a multi-centre comparison study. Regul. Toxicol. Pharmacol. 39, 271–281.
Wargniez, W., et al., 2017. Inter-laboratory skin distribution study of 4-n-butyl resorcinol: the importance of liquid chromatography/mass spectrometry (HPLC–MS/MS) bioanalytical validation. J. Chromatogr. B: Analytical Technologies in the Biomedical and Life Sciences 1060, 416–423.
Wickham, H., 2016. ggplot2 - Elegant Graphics for Data Analysis. Springer International Publishing.
Wilkinson, S.C., et al., 2006. Interactions of skin thickness and physicochemical properties of test compounds in percutaneous penetration studies. Int. Arch. Occup. Environ. Health 79, 405–413.

Regulatory Toxicology and Pharmacology 133 (2022) 105213

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

Corrigendum

Corrigendum to "An OECD TG 428 study ring trial with 14C-Caffeine demonstrating repeatability and robustness of the dermal absorption in vitro method" [Regul. Toxicol. Pharmacol. 132 (2022) 105184]

Felix M. Kluxen [a], Styliani Totti [b], Wilfred Maas [c], Frank Toner [d], Leanne Page [d], Kathryn Webbley [e], Rajendra Nagane [f], Robert Mingoia [g], Christine Whitfield [g], John Kendrick [h], Claire Valentine [h], Jeanne Bernal Dorange [i], Camille Egron [i], Camille Imart [i], Jeanne Y. Domoradzki [j], Philip Fisher [k], Christine Lorez [l], Steve McEuen [m], Edgars Felkers [a], Tao Chen [b], Christiane Wiemann [n,*]

[a] *ADAMA Deutschland GmbH, Cologne, Germany*
[b] *University of Surrey, Guildford, United Kingdom*
[c] *Charles River Laboratories, Den Bosch, the Netherlands*
[d] *Charles River Laboratories, Tranent, United Kingdom*
[e] *Pharmaron, Rushden, United Kingdom*
[f] *Jai Research Foundation, Valvada, India*
[g] *Corteva Agriscience, Newark, United States*
[h] *Labcorp Drug Development, Harrogate, United Kingdom*
[i] *Eurofins Agroscience Services Chem SAS, Vergeze, France*
[j] *Corteva Agriscience, Indianapolis, United States*
[k] *Bayer SAS, Bayer Crop Science, Sophia, Antipolis, France*
[l] *Syngenta Crop Protection AG, Basel, Switzerland*
[m] *FMC Corporation, Philadelphia, PA, USA*
[n] *BASF Oesterreich GmbH, Vienna, Austria*

The authors regret that a wrong figure is shown as Fig. 6. In the published manuscript Fig. 4 is shown twice, i.e., as Fig. 4 and as Fig. 6; the error occurred during re-formatting the manuscript after peer-review. Thus, the correct figure was available to the reviewers.

The correct Fig. 6 is given below:

In addition, in the legend of Fig. 10 a change in the outlier assignment adopted during the peer-review is not correctly reflected. The legend should thus read:

Fig. 10. Similar to Fig. 9, but A) laboratories are shown in facets b-f or B) donors are shown in facets 1–12. The high response for donor 3 in laboratory e (ID e1.4) occurs due to a high receptor fluid wash value.

Further the authors noted that in the published version a mismatch of some affiliations of authors has happened:

John Kendrick is affiliated with Labcorp Drug Development, Harrogate, United Kingdom.

Camille Egron and Camille Imart are affiliated with Eurofins Agroscience Services Chem SAS, Vergeze, France.

Attached to this corrigendum is also the Supplement 2 which was missed to be uploaded along with the corrected manuscript after the peer review but has been available for the reviewers.

The authors would like to apologise for any inconvenience caused.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.yrtph.2022.105184.

DOI of original article: https://doi.org/10.1016/j.yrtph.2022.105184.
* Corresponding author.
E-mail address: christiane.wiemann@basf.com (C. Wiemann).

https://doi.org/10.1016/j.yrtph.2022.105213

Available online 8 July 2022

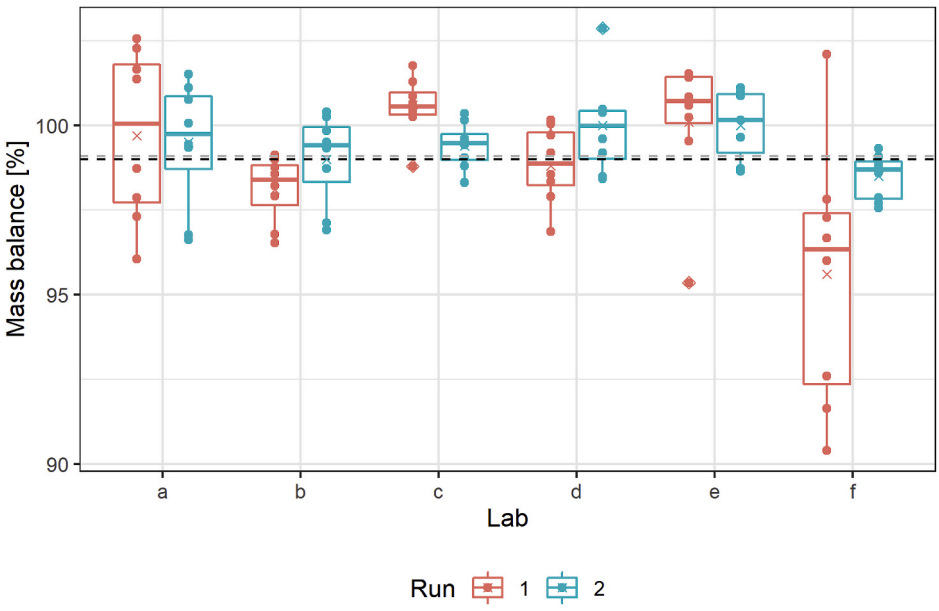

Fig. 6. Mass balance [% applied] between laboratories and runs (red/blue), as individual values (dots), boxplots and means (multiplication sign). Boxplot outliers are highlighted with a diamond shape. Very high mass balance was achieved by all laboratories. Only run 1 in laboratory f was notably different, however, still within test guideline requirements. Mass balance of laboratory a is more variable between the cells than in the other laboratories, other-wise the response is not noticeable different. Dashed horizontal grey line: overall mass balance; black line laboratory a excluded (note the lines are very close). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)