Batch-to-Batch Pharmacokinetic Variability Confounds Current Bioequivalence Regulations: A Dry Powder Inhaler Randomized Clinical Trial

E Burmeister Getz[1], KJ Carroll[2], B Jones[3] and LZ Benet[4]

Current pharmacokinetic (PK) bioequivalence guidelines do not account for batch-to-batch variability in study design or analysis. Here we evaluate the magnitude of batch-to-batch PK variability for Advair Diskus 100/50. Single doses of fluticasone propionate and salmeterol combinations were administered by oral inhalation to healthy subjects in a randomized clinical crossover study comparing three different batches purchased from the market, with one batch replicated across two treatment periods. All pairwise comparisons between different batches failed the PK bioequivalence statistical test, demonstrating substantial PK differences between batches that were large enough to demonstrate bio-*in*equivalence in some cases. In contrast, between-replicate PK bioequivalence was demonstrated for the replicated batch. Between-batch variance was ~40–70% of the estimated residual error. This large additional source of variability necessitates re-evaluation of bioequivalence assessment criteria to yield a result that is both generalizable and consistent with the principles of type I and type II error rate control.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ Current pharmacokinetic bioequivalence guidelines neither consider nor account for batch-to-batch variability in study design or analysis. Data demonstrating batch-to-batch variability are not well described in the literature, making it difficult for regulatory and scientific communities to openly discuss the magnitude and consequences of this variability component.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ The current study characterized the magnitude of Advair Diskus 100/50 within-subject between-batch pharmacokinetic variability separately from within-subject residual error when administered as a single dose to healthy subjects per the FDA's product-specific draft guidance.

WHAT THIS STUDY ADDS TO OUR KNOWLEDGE

☑ The current study demonstrates that Advair Diskus 100/50 batch-to-batch pharmacokinetic variability is a reproducible phenomenon and a substantial component of total variability. Current FDA regulatory requirements are confounded by this batch-to-batch PK variability.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY AND THERAPEUTICS

☑ Substantial batch-to-batch variability necessitates re-evaluation of bioequivalence assessment criteria to ensure results that are generalizable and consistent with the FDA's standards of type I and type II error rate control in bioequivalence testing.

In the US, marketing approval of a new generic drug (the "Test" product) generally requires a demonstration of pharmacokinetic (PK) bioequivalence to a reference listed (innovator) drug (the "Reference" product). This requirement also applies to locally acting orally inhaled products. The standard criterion for statistical bioequivalence applied by the US Food and Drug Administration (FDA) requires that the 90% confidence interval around the geometric mean Test/Reference ratio (GMR) for both the maximum observed plasma concentration (C_{max}) and the area under the concentration-vs.-time curve (AUC) be entirely contained within 80% to 125%.[1] Effectively, the PK bioequivalence test places a requirement on both the center of the GMR confidence interval (the Test/Reference point estimate) and the GMR confidence interval width, i.e., on the mean and variance of the underlying (log-scale) treatment difference distribution. These two requirements ensure that differences between a generic and its Reference Listed Drug comparator are adequately small, and that the data supporting this claim allow sufficient certainty in the conclusion.

In PK bioequivalence testing, it is regularly assumed that the Test and Reference products can each be adequately represented by a single manufacturing batch, i.e., it is assumed that between-batch differences within a product are small enough to ignore. Internal data led us to consider the validity of this underlying assumption. Prior to the current study, a set of five bioequivalence studies comparing a Test product to Advair Diskus 100/50

[1]Oriel Therapeutics, Inc., Berkeley, California, USA; [2]KJC Statistics, Stockport, Cheshire, UK; [3]Novartis Pharma, Basel, Switzerland; [4]University of California, San Francisco, California, USA. Correspondence: E Burmeister Getz (EGetz@orieltherapeutics.com)

Received 26 January 2016; accepted 27 March 2016; advance online publication 2 April 2016. doi:10.1002/cpt.373

Table 1 Randomization schedule for the four-treatment, four-period Williams crossover trial design

Number of subjects randomized	Number of subjects completed	Sequence no.	Sequence	Period 1	2	3	4
7	7	1	A-B-C-D	A	B	C	D
7	7	2	B-D-A-C	B	D	A	C
8	7	3	C-A-D-B	C	A	D	B
8	8	4	D-C-B-A	D	C	B	A

Treatments A and B were replicates of a single manufacturing batch of Advair Diskus 100/50. In total, three different manufacturing batches were administered, one batch in treatments A and B, one in treatment C, and one in treatment D.

demonstrated widely varying individual Test/Reference PK point estimates. Common sources of between-study variability could not reasonably explain the differences—the only clear difference between the studies was the use of different manufacturing batches of the Test and Reference products. To pursue the hypothesis that between-batch differences were the cause of the observed between-study differences, we estimated variance components for the two studies in which several (three) batches of either Test or Reference were administered. While statistically significant between-batch variability of the Test product was not found, the estimated between-batch variance of Reference for FP C_{max}, as an example metric, was 1.4-fold larger than the estimated residual variance, and hence a highly significant contributor to total variability of the Reference product.

To confirm the observation that batch-to-batch PK variability of Advair Diskus 100/50 is a substantial component of total variability, we performed the study reported here. The current study was specifically designed to separately estimate within-subject between-batch variance and within-subject residual variance. The design of the current study is what has seemingly been cited as the FDA recommendation for this purpose,[2] in which comparison across batches is accompanied by batch replication. In the current study, one batch of Advair Diskus 100/50 was replicated across two treatment periods and also compared to two different batches of Advair Diskus 100/50 in a four-way randomized Williams crossover design. Further, the design compared the PK of Advair Diskus 100/50 batches at both different and similar stages of the shelf-life to investigate whether the PK differences were associated with batch age.

Batch-to-batch PK variability can significantly impact estimation of both components of the bioequivalence test, i.e., the mean and the variance of the treatment difference. In order for the Test/Reference ratio point estimate to be considered a sensible value it should be consistent (reproducible) and unbiased. Clearly, if manufacturing batches differ from one another with regard to PK, the conventional PK bioequivalence study in which a single batch of Reference product is compared, often in a two-way crossover design, to a single batch of Test product will not yield a consistent or reliable Test/Reference point estimate. Instead, the measured Test/Reference point estimate will be highly dependent on the batches that happened to be selected. Equally, the width of the GMR confidence interval should reflect all substantial sources of uncertainty in the point estimate—ignoring batch-to-batch variability, if substantial, leads to an artificially narrow confidence interval that overestimates the certainty in the GMR. Despite its potential significance, batch-to-batch variability is not currently considered or accounted for in the PK bioequivalence basis for generic drug approval.

Both US and European regulators are aware that batch-to-batch variability, among orally inhaled drug products (OIDPs),[2,3] in particular, poses additional challenges for follow-on drug development. However, no revised guidance for the design and analysis of PK bioequivalence studies has yet been established to accommodate this additional variability component. There has been little to no discussion in the scientific community regarding how the number of Reference (or Test) batches should be determined or how such a multiple-batch design should be analyzed in order to construct a meaningful confidence interval around the Test/Reference GMR. In part, perhaps, this is because data demonstrating batch-to-batch variability (for OIDPs or any other product) are not currently well described in the literature, making it difficult for both the regulatory and scientific communities to understand the magnitude and circumstances of this PK variability component. The aim of the current article is to share such data on the observed batch-to-batch PK variability of a US marketed drug product, here Advair Diskus 100/50, when administered as a single dose to healthy adult subjects as per the FDA's product-specific draft guidance.[4]

RESULTS

Twenty-eight subjects were allocated to seven copies of the Williams design given in **Table 1** and an additional two subjects were allocated to sequences 3 and 4. Of the 30 randomized subjects, 29

Table 2 Subject demographics

EudraCT number	2015-000068-32
Population	Healthy
FEV_1 (% predicted)	≥90%
Age (years)	35 ± 8.9 (18–49)
M/F	22/8
Weight (kg)	74.2 ± 12.3 (51.6–97.6)
Height (cm)	173 ± 9 (153–194)
BMI (kg/m^2)	24.7 ± 2.9 (19.5–29.7)

Data are mean ± standard deviation (minimum–maximum).

Table 3 Summary of pharmacokinetic parameters for fluticasone propionate, 100 μg (FP) and salmeterol, 50 μg (S) following administration to healthy subjects as Advair Diskus 100/50

	Batch 1 – Replicate A	Batch 1 – Replicate B	Batch 2	Batch 3
FP C_{max} (pg/mL)	44.7 [11.1–89.4]	45.4 [19.9–78.5]	69.2 [22.2–163]	58.9 [19.9–101]
FP T_{max} (min)	9 [4–60]	8 [4–60]	6 [3–30]	8 [3–60]
FP $AUC_{(0\text{-}t)}$ (h·pg/mL)	178 [80–328]	177 [87–377]	230 [102–392]	220 [83–411]
FP $AUC_{(0\text{-}inf)}$ (h·pg/mL)	210 [94–350]	192 [96–401]	256 [118–405]	236 [146–431]
FP $t_{1/2}$ (h)	10.3 [3.0–18.7]	9.1 [3.5–15.1]	11.4 [6.7–16.2]	12.5 [6.6–24.4]
S C_{max} (pg/mL)	81.6 [34.9–193]	85.8 [38.9–188]	132 [41.8–287]	104 [28.1–201]
S T_{max} (min)	4 [3–5]	4 [3–5]	4 [3–5]	4 [3–5]
S $AUC_{(0\text{-}t)}$ (h·pg/mL)	114 [46–437]	122 [44–431]	154 [84–509]	145 [61–528]
S $AUC_{(0\text{-}inf)}$ (h·pg/mL)	137 [64–526]	154 [89–481]	180 [93–606]	170 [73–605]
S $t_{1/2}$ (h)	12.5 [4.5–22.9]	12.9 [7.1–17.5]	14.5 [5.1–20.5]	13.3 [4.8–18.7]

Least squares geometric mean [range] except T_{max} for which the median [range] is reported.

subjects completed all four periods of the study. One subject (female, 40 years) was withdrawn from the study prior to treatment in period 4 (sequence 3) due to an adverse event of cough. The analysis population was therefore comprised of 30 subjects for three of the four treatments and 29 subjects for the fourth treatment.

Demographics of the clinical study participants are given in **Table 2**.

Individual batch pharmacokinetics

Least-square geometric mean values, along with minimum and maximum values across all subjects, for individual manufacturing batches of Advair Diskus 100/50 are given in **Table 3**. **Figure 1** illustrates the average blood concentration-vs.-time profile for FP (first 4 hours after inhalation) and salmeterol (first hour after inhalation) from each batch of product.

Fluticasone propionate (FP) PK, at a dose of 100 μg to healthy adult subjects, has not been previously published. Maximum plasma concentration (C_{max}) was typically reached ~10 minutes after dosing, although with a range in T_{max} values (time of maximum observed plasma concentration) between individual profiles of 3–60 minutes. T_{max} variability was observed within subjects across all treatments. The wide T_{max} range is due in part to the erratic absorption of FP, leading to multiple peaks in plasma concentration during the first hour after dosing. This behavior is presumably a consequence of FP's poor solubility, which results in dissolution being the rate-limiting step in the pulmonary absorption process. The erratic nature of the concentration profile during absorption is prominent in the individual subject data, although less evident in the average profiles. FP concentrations decline with an apparent terminal elimination half-life of ~11

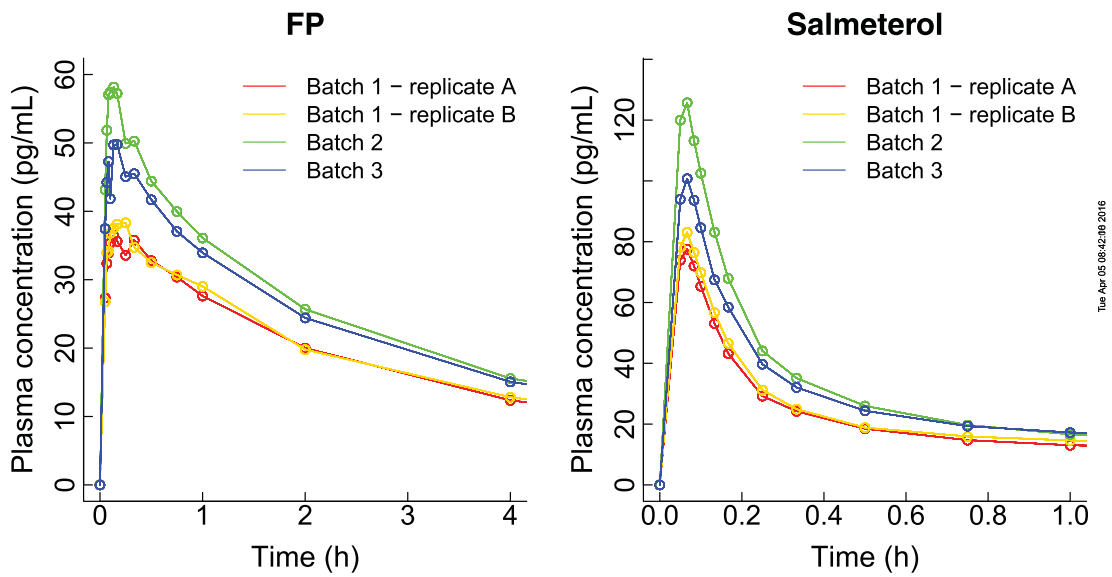

Figure 1 Plasma concentration-vs.-time profiles for fluticasone propionate (FP; 100 μg) and salmeterol (50 μg) following single-dose dry powder oral inhalation to healthy adult subjects as Advair Diskus 100/50.

Table 4 Bioequivalence assessment within and between manufacturing batches of Advair Diskus 100/50

	Geometric mean ratio (%)	
	Estimate	90% CI
Batch 1 (replicate A)– vs. –Batch 1 (replicate B)		
FP C_{max}	98.66	87.29–111.50
FP $AUC_{(0\text{-}t)}$	100.36	92.29–109.14
FP $AUC_{(0\text{-}tcommon)}$	100.68	92.87–109.15
FP $AUC_{(0\text{-}inf)}$	109.17	95.59–124.68
S C_{max}	95.15	82.75–109.42
S $AUC_{(0\text{-}t)}$	93.54	86.81–100.79
S $AUC_{(0\text{-}tcommon)}$	95.40	89.66–101.49
S $AUC_{(0\text{-}inf)}$	88.78	81.07–97.21
Batch 1– vs. –Batch 2		
FP C_{max}	65.05	58.56–72.26
FP $AUC_{(0\text{-}t)}$	77.02	71.67–82.77
FP $AUC_{(0\text{-}tcommon)}$	77.99	72.80–83.54
FP $AUC_{(0\text{-}inf)}$	78.39	69.66–88.21
S C_{max}	63.44	56.27–71.52
S $AUC_{(0\text{-}t)}$	76.73	71.97–81.81
S $AUC_{(0\text{-}tcommon)}$	79.64	75.55–83.95
S $AUC_{(0\text{-}inf)}$	80.72	74.68–87.25
Batch 1– vs. –Batch 3		
FP C_{max}	76.47	68.84–84.94
FP $AUC_{(0\text{-}t)}$	80.68	75.08–86.70
FP $AUC_{(0\text{-}tcommon)}$	81.37	76.06–87.06
FP $AUC_{(0\text{-}inf)}$	85.25	76.90–94.51
S C_{max}	80.24	71.17–90.46
S $AUC_{(0\text{-}t)}$	81.20	76.15–86.57
S $AUC_{(0\text{-}tcommon)}$	81.68	77.41–86.18
S $AUC_{(0\text{-}inf)}$	85.58	78.42–93.40
Batch 2– vs. –Batch 3		
FP C_{max}	117.55	104.16–132.65
FP $AUC_{(0\text{-}t)}$	104.75	96.43–113.79
FP $AUC_{(0\text{-}tcommon)}$	104.34	96.55–112.76
FP $AUC_{(0\text{-}inf)}$	108.75	95.63–123.68
S C_{max}	126.48	110.18–145.19
S $AUC_{(0\text{-}t)}$	105.82	98.30–113.91
S $AUC_{(0\text{-}tcommon)}$	102.56	96.48–109.02
S $AUC_{(0\text{-}inf)}$	106.02	96.04–117.04

Average bioequivalence methodology was used for the treatment comparisons based on ln-transformed data. Geometric mean ratios and confidence intervals (CI) were exponentiated to the original scale for display. FP, fluticasone propionate; S, salmeterol.

hours. Variability between subjects in FP PK was high, as judged by the wide range observed for each of the PK metrics, thus highlighting the importance of the crossover study design for bioequivalence testing where possible.

Maximum salmeterol plasma concentration was reached on average 4 minutes after dosing, with little variability between subjects or treatment periods; the range of individual salmeterol T_{max} values was 3–5 minutes. Salmeterol concentrations decline with an apparent terminal elimination half-life of ~13 hours.

Pharmacokinetic batch comparisons

Geometric mean ratios with 90% confidence intervals for the comparison of PK metrics between the treatments are summarized in **Table 4**, and presented graphically for the C_{max} metrics in **Figure 2**. The two replicates of Advair Diskus 100/50 "batch 1" were demonstrated to be bioequivalent with respect to C_{max}, $AUC_{0\text{-}t}$, and $AUC_{0\text{-}inf}$ for both FP and salmeterol, thus fully meeting the FDA's PK bioequivalence criteria.[1]

In contrast, all between-batch PK comparisons failed bioequivalence. The most extreme failure was "batch 1" vs. "batch 2," for which bio-*in*equivalence[5] (defined as a 90% geometric mean ratio confidence interval entirely excluded from the 80–125% bioequivalence zone) was demonstrated for both FP C_{max} and salmeterol C_{max}. The importance of identifying the result as bio-*in*equivalent, instead of simply failing bioequivalence, is to highlight that the failure allows a definitive regulatory conclusion and is not the consequence of inadequate study size. The FP C_{max} 90% confidence interval around the geometric mean batch ratio was 58.56–72.26%, fully outside the 80–125% bioequivalence window. Similarly, the corresponding salmeterol C_{max} 90% confidence interval was 56.27–71.52%, also outside 80–125% and thus demonstrating that Advair Diskus 100/50 "batch 1" is bio-*in*equivalent to "batch 2" with regard to maximum blood concentrations of both active ingredients. Batch 1–vs.–batch 2 AUC ratios also failed bioequivalence, with point estimates ranging from 76.73% to 80.72% and all lower confidence limits below 80%. With batch differences this large, no increase in study size would yield a passing result.

The batch 1-vs.-batch 3 comparison failed bioequivalence on every PK metric, with point estimates ranging from 76.47% to 85.58%. The batch 2-vs.-batch 3 comparison failed bioequivalence on both FP and salmeterol C_{max}. The FP C_{max} point estimate of 117.55%, with the observed within-subject coefficient of variation of 28.67%, would have required 262 subjects in a two-way crossover study to have an 80% probability of passing the bioequivalence requirement. Thus, batch differences this large are not consistent with the statistical bioequivalence test. The salmeterol C_{max} point estimate of 126.48% cannot pass the bioequivalence test for any size study because the point estimate itself is outside the 80–125% bioequivalence window.

As previously reported,[6] within-subject differences in the time of the last quantifiable drug concentration (t_{last}) between PK profiles that differ in magnitude (and therefore fall below assay limit of quantitation [LOQ] at different times) can lead to bias in the

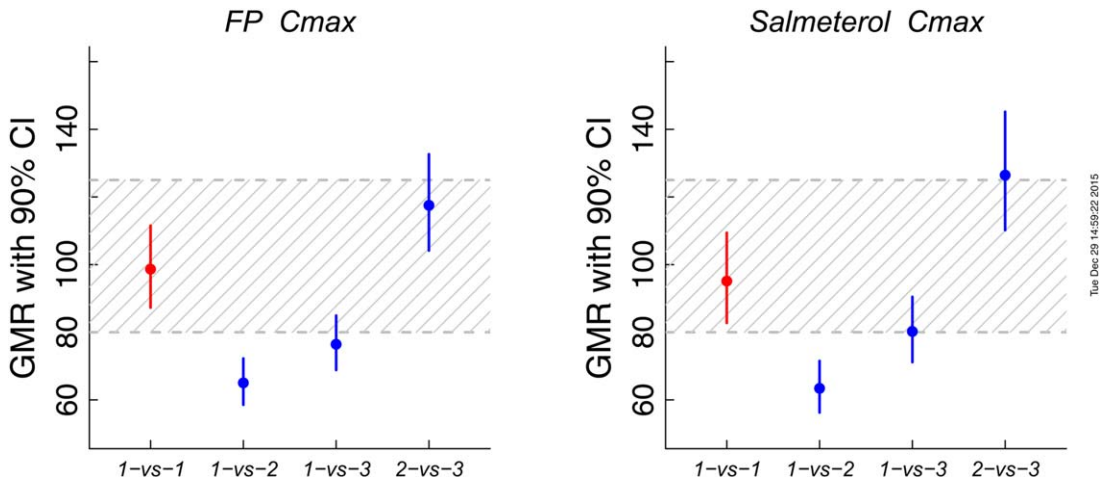

Figure 2 Pharmacokinetic bioequivalence assessment within and between manufacturing batches of Advair Diskus 100/50. The maximum plasma concentration (C_{max}) of fluticasone propionate (FP) and salmeterol was compared among three manufacturing batches. Batches are identified numerically as batch 1, batch 2, and batch 3. Batch 1 was replicated across two treatment periods; the comparison of the two replicates of batch 1 is indicated in red. Comparisons between different batches are indicated in blue. For between-batch comparisons, the younger batch is represented in the numerator of the GMR. The 80–125% bioequivalence region is shaded.

$AUC_{(0\text{-}t)}$ geometric mean ratio. This is simply a consequence of comparing $AUC_{(0\text{-}t)}$ values that have been calculated using different time windows, and is remedied by use of a common t_{last} for all profiles for a given subject and analyte (the revised PK parameter is referred to as $AUC_{(0\text{-}tcommon)}$). As expected, **Table 4** displays this phenomenon, with all $AUC_{(0\text{-}t)}$ geometric mean ratios differing from 100% more than the corresponding $AUC_{(0\text{-}tcommon)}$ ratios.

Estimation of variance components

The PK differences observed between batches of Advair Diskus 100/50 indicate that it is necessary to consider not only the within-subject residual variability, but also the within-subject batch-to-batch variability in bioequivalence study designs and corresponding power calculations. Estimation of the relative magnitudes of these two variance components from the current study data is provided in **Table 5**.

Table 5 Variance component estimation following administration of a single dose of Advair Diskus 100/50 from three different manufacturing batches to healthy subjects

Fluticasone propionate			Type 3 (Method of Moments)			REML	
Parameter	**Variance component**	**DF**	**Estimate**	$\hat{\sigma}^2_{bb}/\hat{\sigma}^2_{wb}$	***P*-value**	**Estimate**	$\hat{\sigma}^2_{bb}/\hat{\sigma}^2_{wb}$
C_{max}	σ^2_{bb}	2	0.0524	55%	0.0001	0.0455	62%
	σ^2_{wb}	26	0.0959	—	—	0.0731	—
$AUC_{(0\text{-}t)}$	σ^2_{bb}	2	0.0220	46%	0.0001	0.0189	54%
	σ^2_{wb}	26	0.0474	—	—	0.0348	—
$AUC_{(0\text{-}tcommon)}$	σ^2_{bb}	2	0.0199	46%	0.0001	0.0173	58%
	σ^2_{wb}	23	0.0429	—	—	0.0300	—
Salmeterol			**Type 3 (Method of Moments)**			**REML**	
Parameter	**Variance component**	**DF**	**Estimate**	$\hat{\sigma}^2_{bb}/\hat{\sigma}^2_{wb}$	***P*-value**	**Estimate**	$\hat{\sigma}^2_{bb}/\hat{\sigma}^2_{wb}$
C_{max}	σ^2_{bb}	2	0.0540	41%	0.0001	0.0493	51%
	σ^2_{wb}	26	0.1314	—	—	0.0972	—
$AUC_{(0\text{-}t)}$	σ^2_{bb}	2	0.0222	59%	0.0001	0.0192	67%
	σ^2_{wb}	26	0.0379	—	—	0.0286	—
$AUC_{(0\text{-}tcommon)}$	σ^2_{bb}	2	0.0184	70%	0.0001	0.0156	83%
	σ^2_{wb}	25	0.0262	—	—	0.0188	—

One batch was replicated to allow estimation of within-batch variance. σ^2_{bb}: within-subject, between-batch variance. σ^2_{wb}: within-subject residual error variance (within-subject, within-batch variance).

For all FP PK parameters when using the Method of Moments approach, the between-batch, within-subject variance component (σ^2_{bb}) was highly significant ($P = 0.0001$), with an estimated magnitude as a percentage of the between-replicates within-batch, within-subject (i.e., residual error) variance component (σ^2_{wb}) of 46–55%. Restricted maximum likelihood (REML) estimation yielded a higher estimation of the magnitude of σ^2_{bb} as a percentage of σ^2_{wb}, 54-62%.

Similar results were obtained for salmeterol. For all salmeterol PK parameters, when using the Method of Moments approach, the between-batch, within-subject variance component (σ^2_{bb}) was highly significant ($P = 0.0001$), with an estimated magnitude as a percentage of the between-replicates within-batch, within-subject (i.e., residual error) variance component (σ^2_{wb}) of 41–70%. Similar to FP, REML estimation yielded a higher estimation of the magnitude of σ^2_{bb} as a percentage of σ^2_{wb}, 51–83%.

Adverse events

There were no deaths or serious adverse events in the study. One subject was withdrawn due to an adverse event of cough of moderate intensity prior to dosing on period 4. The majority of the adverse events were classified as mild and occurred with similar frequency for all Advair Diskus 100/50 batches. A total of 1 ("batch 1," replicate A), 4 ("batch 1," replicate B), 3 ("batch 2"), and 3 ("batch 3") adverse events were reported for the individual treatments.

Safety data

There were no clinically relevant trends in clinical laboratory data or vital signs data in the study aside from a decline in hemoglobin expected from the large blood loss involved in the study design and a decline in creatine kinase expected from the generally sedentary lifestyle of subjects while housed in the clinical unit during the study.

DISCUSSION

The current study has demonstrated that substantial PK differences can occur between marketed batches of an FDA-approved US drug product, shown here for an FP/salmeterol dry powder oral inhalation product, Advair Diskus 100/50. These differences are large enough to cause consistent failure when Advair Diskus 100/50 is compared to itself in FDA's test for PK bioequivalence.[4] The widespread use of Advair Diskus 100/50 on the market over the past 15 years has, in broad terms, provided no indication of concern, although subtle (or otherwise) clinical consequences are unlikely to become apparent in regular market use, especially in asthma that is itself a disease characterized by variable status. Regarding systemic exposure, it is similarly possible that the batch-to-batch PK differences described here would go unnoticed in patient use of this locally acting product based on the publicly available literature evidence. Existing dose–response data for salmeterol demonstrate adverse effects on systemic pharmacodynamic markers (heart rate, QT interval, tremor, glucose, potassium) at doses of 100 μg and above,[7–9] twice the marketed dose, although it is acknowledged that salmeterol's systemic dose–response curve is not well characterized in the literature at inhaled doses below 100 μg. For FP, the relationship between PK and systemic effect demonstrates that FP blood levels achieved for the 100/50 product, even at the highest end of the observed range in healthy subjects, are more than five-fold below the level shown to cause a 5% decline in plasma cortisol,[10] a sensitive marker of adrenocortical suppression. Further, systemic exposure to FP has been shown to be two to three times lower in patients with asthma compared to healthy subjects.[11,12] Although the systemic dose is not considered a precise surrogate for the amount of drug reaching the site of action for this locally acting product, it is expected that the observed variability in PK may be reflected to some extent as variability in locally available drug as well. However, the local pharmacological effects have been demonstrated to be relatively insensitive to dose in asthma patients.[13–16] One limitation of the literature data cited above is the use of population averages, as individual dose–response characteristics may differ.

Instead, the implication of the current finding is primarily relevant to the assessment of PK bioequivalence using Advair Diskus 100/50 as the Reference Listed Drug. The implication is that a generalizable assessment of PK bioequivalence for this product requires a study design and an analysis methodology that correctly account for this substantial additional source of variability. The necessity of this is inherent in the generic drug regulations that sanction substitutability at the pharmacy level, and strive to ensure consistent and predictable response in patients taking generic drugs in the US. Although the quality of the release specifications are not necessarily implicated for the example product presented here, Advair Diskus 100/50, there may be other products and other routes of administration for which unrecognized batch-to-batch PK variability has clinically significant ramifications.

The current study was specifically designed to separately estimate within-subject between-batch variance and within-subject residual variance. In the current study, the between-batch component of variance was 40–70% of the estimated residual variance across all PK metrics. There is, therefore, a clear increased risk of both false-negative and false-positive findings when relying on standard bioequivalence study design and analysis methods that ignore the substantial batch-to-batch component of variability. The results of a conventional PK bioequivalence study comparing a single batch each of the Test and Reference products are not generalizable to the overall product comparison, and risk inadequate characterization of relative PK performance in product use over time on the commercial market.

We applied the FDA's statistical test for PK bioequivalence to the Reference-vs.-Reference comparisons among the different batches used in the current study. All pairwise batch comparisons failed the PK bioequivalence test. FP and salmeterol C_{max} batch-to-batch ratios ranged from 63% to 85% (85% being the reciprocal of 117%) across the three pairwise comparisons. To appreciate the significance of this magnitude of batch-to-batch difference, we note that success in PK bioequivalence testing typically requires geometric mean Test/Reference ratios to be within 5–10% of unity in order for the 90% confidence interval around these estimated ratios to be contained within the 80–125%

bioequivalence goalposts. An FDA review of 2,070 successful single-dose PK bioequivalence studies of generic solid oral drug products revealed that the average Test-vs.-Reference difference in C_{max} and AUC was only 4.35% and 3.56%, respectively, with 98% of successful studies showing a Test-vs.-Reference difference of less than 10%.[17] The batch-to-batch differences observed here, then, are substantial in the context of bioequivalence testing.

Aspects of study execution are not likely to explain the observed between-batch PK differences. In the bioanalysis, all plasma samples from a subject (i.e., plasma samples from all four treatment periods) were processed in the same analytical run to ensure that a single calibration curve was used for quantification. Further, the bioanalysis was blind to the randomization of treatment to period. All three batches were supplied to the European clinical site in a single insulated shipment with temperature monitoring and stored in a single temperature-controlled pharmacy for the 47 days between product receipt and completion of dosing to ensure that all batches were handled in an identical manner. All Diskus devices were removed from the protective foil overwrap within 4 hours of dosing to standardize (and minimize) exposure of the product to ambient humidity prior to dosing.

The observation of "batch 1" vs. "batch 2" bio-*in*equivalence for FP C_{max} and salmeterol C_{max} is not a consequence of study size; a larger study would have produced smaller confidence intervals, therefore accentuating the bio-*in*equivalence finding. However, the repeated failure to meet the bioequivalence requirement among the other batch-vs.-batch comparisons is potentially attributable in part to the relatively small study size. For example, the 90% confidence interval of 104.16% to 132.65% around the FP C_{max} "batch 2" vs. "batch 3" estimated ratio of 117.55% indicates that the true batch ratio on this PK metric may be sufficiently close to 100% to meet the bioequivalence requirement, and that the current failing result is a consequence of measurement noise. Equally, however, the 90% confidence interval allows that the true batch ratio may be even further from 100% than was observed here.

We acknowledge that the underlying causes of batch-to-batch PK variability are potentially relevant to a discussion of how this variability should be handled in generic drug development and approval. We note that a decline in fine particle mass (mass of particles with aerodynamic diameter $\leq$5 μm) with time, due to ambient humidity exposure, is a documented phenomenon for dry powder formulations[18] and could be suggested as an explanation for the observed differences in PK between batches. In the current study we used three batches that differed in age by 12 months ("batch 1" vs. "batch 2") and by 1 month ("batch 2" vs. "batch 3"). Consistent with a potential decline in fine particle mass as an explanation, the oldest of the three batches ("batch 1") yielded the lowest PK.

However, comparison of the two batches that were only 1 month different in age (as determined by the labeled expiry date) revealed two surprising results: (1) PK bioequivalence between these two batches was not met for either FP C_{max} or salmeterol C_{max}, and (2) the younger of the two batches ("batch 3") yielded the lower PK. Batch-to-batch variability in the PK of Advair Diskus 100/50 therefore appears to be driven by factors other than (perhaps additional to) batch age. In a complex product such as Advair Diskus, multiple sources of batch-to-batch variability are expected. These sources may include, for example, variation in formulation characteristics including the micronized lactose inert bulking agent, manufacturing process variation, or variation in the delivery performance of the Diskus device. Thus, it is not adequate to base an assessment of bioequivalence on a single batch of Reference chosen from the middle of the shelf-life, because age is not the only contributor to Advair Diskus 100/50 batch-to-batch variability. We hope that the underlying cause of the observed batch-to-batch variability will eventually be elucidated, as this understanding could lead to an educated guess regarding which other products or dosage forms might be similarly impacted.

The primary consequence of the dataset presented here is that, for products exhibiting substantial batch-to-batch variability, regulatory agencies should establish bioequivalence guidelines with consideration for the emerging scientific evidence. In the 2015 Meeting Report[2] entitled "Pharmacokinetics of Orally Inhaled Drug Products" a recommendation seemingly attributed to an FDA speaker states "To include batch-to-batch and intra-subject variability of the reference product, the comparison of one batch of the test product with two batches of the *R* product, with one *R* batch being repeated, was suggested." In essence, this is the study run here with the "test product" actually being another batch of the reference product, showing that the reference product is not bioequivalent among batches and that this observed batch-to-batch variability is additional to, not a consequence of, intrasubject variability. The Meeting Report[2] continues "For showing BE [bioequivalence] of a *T* product, a randomized, four-way cross-over design would then be necessary." It is obvious from the batch-to-batch variability of the reference product reported here that showing BE of the *T* product in this "necessary" randomized, four-way cross-over design would just be a matter of chance. We propose that with demonstrated and substantial between-batch variability, it is fundamentally necessary for a PK bioequivalence assessment to include multiple batches to avoid inferences regarding bioequivalence, or lack thereof, that are highly dependent on the particular Reference and Test batches selected for a given study. This is needed to ensure that the product average is reasonably well estimated and to allow an estimate of both of the substantial sources of variability, within-subject residual error, and within-subject between-batch variation, to contribute to the determination of confidence in the estimated product difference.

METHODS

The PK of FP and salmeterol were observed in 30 healthy adult subjects in a clinical study performed under clinical trials authorization from the UK Medicines and Healthcare products Regulatory Agency and approval by the Office for Research Ethics Committees Northern Ireland (Lisburn, UK). Written informed consent was obtained from all subjects, and the study was conducted in accordance with the principles of the Declaration of Helsinki.

Study design

A single dose of 100 μg FP in combination with 50 μg salmeterol formulated with lactose as a dry powder was administered by oral inhalation as Advair Diskus 100/50 to healthy adult males and females in a single-

center, randomized, open-label, 4-sequence, 4-period crossover study. Assessment of PK bioequivalence was the primary study objective. Three Advair Diskus batches were investigated, with one of the batches replicated to create the four treatments. The replicated batch was prospectively labeled as two different treatments by pharmacy staff prior to dosing. The number of subjects was selected to provide 80% power in the bioequivalence assessment assuming true PK differences between treatments do not exceed 5% and a within-subject variance of ~0.0625. Subjects were randomly allocated to treatment sequence using the statistical analysis system (SAS, Cary, NC) computer package PROC PLAN. The study was performed at Quintiles Drug Research Unit at Guy's Hospital, London, UK. Recruitment initiated 04 June 2015, and the last visit occurred on 31 July 2015.

All three batches of Advair Diskus 100/50 (GlaxoSmithKline, Research Triangle Park, NC) were purchased directly from the US market and used within labeled expiry.

All study treatments were administered under supervision. For each of the four study periods, participants remained in the clinic for the duration of dosing and PK observation. The dosing procedure followed the instructions provided to patients in the Advair Diskus Medication Guide: exhalation, quick and deep inhalation with 10-second breath-hold, and mouth rinse. Subjects were fasted overnight for at least 10 hours prior to dosing until 4 hours postdose. Water was allowed *ad libitum* during the study except for 1 hour prior through 1 hour postdose. Crossover treatments were separated by a washout period of not less than 7 days.

Clinical study participants

Participants (≥18 years, body weight ≥50 kg, body mass index 19 to 30 kg/m^2 [inclusive]) had no history of asthma, a fractional exhaled nitric oxide (FENO) value of ≤47 ppb at screening, and a forced expiratory volume in 1 second (FEV_1) of ≥90% of predicted at screening. Female subjects were only entered if they were not pregnant or breast feeding and were using medically acceptable methods of contraception.

Pharmacokinetic samples

In each treatment period, serial blood samples (6 mL) were drawn for the determination of FP and salmeterol concentration predose and at 3, 4, 5, 6, 8, 10, 15, 20, 30, and 45 minutes and 1, 2, 4, 8, 12, 16, 20, 24, 28, 32, and 36 hours after the inhalation. Blood samples were centrifuged within 30 minutes of collection at ~2,000*g* for 15 minutes at 4°C. All plasma samples were stored in a –20°C freezer. The bioanalysis of FP and salmeterol was conducted by Covance (Salt Lake City, UT) using a validated liquid chromatography/tandem mass spectrometry (LC-MS/MS) method with a quantitative range from 1.00 to 200 pg/mL for each analyte.

Safety measures

Safety was monitored by recording adverse events (AE), physical examination, vital signs (blood pressure, pulse rate, body temperature), blood hematology, serum biochemistry, and urinalysis.

Pharmacokinetics analysis

PK parameters for FP and salmeterol were calculated using standard noncompartmental analysis. The estimated PK parameters were the maximum observed plasma concentration (C_{max}) and time to C_{max} (T_{max}), the area under the concentration-vs.-time curve to the last time of quantifiable concentration [$AUC_{(0\text{-}t)}$] calculated using the linear trapezoidal method and extrapolated to infinite time [$AUC_{(0\text{-}inf)}$] and the elimination rate constant (λ_z) and corresponding half-life.

Because terminal phase plasma concentrations can fall below the bioanalytical method's lower LOQ when sampling continues for at least three terminal half-lives, there is the potential for a treatment bias in $AUC_{(0\text{-}t)}$ due to differences in the time period over which $AUC_{(0\text{-}t)}$ is calculated.[5] Therefore, an additional exposure metric was analyzed. This unbiased exposure metric, $AUC_{(0\text{-}tcommon)}$, was determined as the area under the plasma concentration-vs.-time curve to the last time for which all profiles for a given subject and analyte had a concentration ≥LOQ, calculated using the linear trapezoidal method.

For comparison of the replicated batch to the other batches, the within-subject natural log-scale PK parameters from the two replicates were averaged to give a more robust estimate of the PK performance of the replicated batch.

Statistical analysis

Statistical analysis of treatment means. PK parameter least-square geometric means were determined for individual batches using an analysis of variance (ANOVA) model with fixed effects for treatment, period, and sequence, and a random subject-within-sequence term, using the natural logarithms of the data. ESTIMATE statements in SAS PROC GLM were used to estimate treatment differences in geometric means, both between difference batches and within the replicated batch.

Estimation of variance components. Variance component estimation was based on a type 3 analysis using SAS PROC MIXED that provided a full ANOVA table indicating sources of variation (including residual error variance), associated degrees of freedom (DF), sums of squares, mean squares and also expected mean squares, the error term and error DF for each of the expected mean squares. These outputs allowed Method of Moments (MM) estimation of the variance components for the random effect terms specified in the PROC MIXED model code (i.e., "subject within sequence," "treatment," "treatment within subject by sequence," and residual error). The variance component for "treatment" then corresponds to σ^2_{bb} (i.e., between-batch within-subject variance component) and the residual error corresponds to σ^2_{wb} (i.e., between-replicates within-batch, within-subject variance component). A supplementary analysis using the PROC MIXED option **method-REML** to provide restricted maximum likelihood (REML) estimation for the variance components was performed.

All calculations were performed using SAS for Windows v. 9.3 or higher.

ACKNOWLEDGMENTS

E.B.G. is an employee of Oriel Therapeutics, a Novartis company. K.J.C. and L.Z.B. are paid consultants to Oriel Therapeutics. B.J. is an employee of Novartis Pharma AG. The authors thank Dr. Robert J. Meyer for valuable discussion regarding data interpretation.

CONFLICT OF INTEREST/DISLOSURE

E.B.G. is an employee of Oriel Therapeutics, an indirect wholly-owned subsidiary of Novartis AG. K.J.C. and L.Z.B. are paid consultants to Oriel Therapeutics. B.J. is an employee of Novartis Pharma AG. The study was paid for by Oriel Therapeutics.

AUTHOR CONTRIBUTIONS

E.B.G., K.J.C., B.J., and L.Z.B. wrote the article; E.B.G. and K.J.C. designed the research; E.B.G. performed the research; E.B.J., K.J.C., B.J., and L.Z.B. analyzed the data.

1. Guidance for Industry. Statistical Approaches to Establishing Bioequivalence. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), January 2001.

2. Hochhaus, G., Horhota, S., Hendeles, L., Suarez, S. & Rebello, J. Pharmacokinetics of orally inhaled drug products. *AAPS J.* **17**, 769–775 (2015).
3. Lee, S.L. *et al*. Regulatory considerations for approval of generic inhalation drug products in the US, EU, Brazil, China, and India. *AAPS J.* **17**, 1285–1304 (2015).
4. Draft Guidance on Fluticasone Propionate; Salmeterol Xinafoate. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), September 2013.
5. Background Information for Advisory Committee for Pharmaceutical Science: Concept and Criteria of BioINequivalence. October 20, 2004. <http://www.fda.gov/ohrms/dockets/ac/04/briefing/2004-4078B1_06_BioINequivalence.pdf>. Accessed February 2016.
6. Fisher, D., Kramer, W. & Burmeister Getz, E. Evaluation of a scenario in which estimates of bioequivalence are biased and a proposed solution: tlast (common). *J. Clin. Pharmacol.* E-pub ahead of print.
7. Bennett, J.A. & Tattersfield, A.E. Time course and relative dose potency of systemic effects from salmeterol and salbutamol in healthy subjects. *Thorax* **52**, 458–464 (1997).
8. Kirby, S. *et al*. Salmeterol and fluticasone propionate given as a combination: lack of systemic pharmacodynamics and pharmacokinetic interactions. *Eur. J. Clin. Pharmacol.* **56**, 781–791 (2001).
9. Maconochie, J.G. & Forster, J.K. Dose-response study with high-dose inhaled salmeterol in healthy subjects. *Br. J. Clin. Pharmacol.* **33**, 342–345 (1992).
10. Mackie, A.E. & Bye, A. The relationship between systemic exposure to fluticasone propionate and cortisol reduction in healthy male volunteers. *Clin. Pharmacokinet.* **39**, 47–54 (2000).
11. Brutsche, M.H. *et al*. Comparison of pharmacokinetics and systemic effects of inhaled fluticasone propionate in patients with asthma and healthy volunteers: a randomised crossover study. *Lancet* **356**, 556–561 (2000).
12. Daley-Yates, P.T., Tournant, J. & Kunka, R.L. Comparison of the systemic availability of fluticasone propionate in healthy volunteers and patients with asthma. *Clin. Pharmacokinet.* **39**, 39–45 (2000).
13. Daley-Yates, P.T., Austin, D.J. & Bentley, J.H. Exhaled nitric oxide (eNO) versus adenosine-5′-monophosphate (AMP) challenge for demonstrating dose response to inhaled corticosteroid. *Am. J. Respir. Crit. Care Med.* **187**, A1926 (2013).
14. Lumry, W.R. *et al*. Fluticasone propionate hydrofluoroalkane inhalation aerosol in patients receiving inhaled corticosteroids. *Ann. Allergy Asthma Immunol.* **96**, 51–59 (2006).
15. Currie, G.P., Fowler, S.J. & Lipworth, B.J. Dose response of inhaled corticosteroids on bronchial hyperresponsiveness: a meta-analysis. *Ann. Allergy Asthma Immunol.* **90**, 194–198 (2003).
16. Kuna, P., Thyroff-Friesinger, U., Gath, I. & Jones, S. Randomized equivalence trial: a novel multidose dry powder inhaler and originator device in adult and adolescent asthma. *Allergy Asthma Proc.* **36**, 352–364 (2015).
17. Davit, B.M. *et al*. Comparing generic and innovator drugs: a review of 12 years of bioequivalence data from the United States Food and Drug Administration. *Ann. Pharmacother.* **43**, 1583–1597 (2009).
18. Borgström, L., Asking, L. & Lipniunas, P. An *in vivo* and *in vitro* comparison of two powder inhalers following storage at hot/humid conditions. *J. Aer. Med.* **18**, 304–310 (2005).