



Iowa Research Online
The University of Iowa's Institutional Repository

Stead Family Department of Pediatrics Publications

4-1-2018

Effect of Study Design on Sample Size in Studies Intended to Evaluate Bioequivalence of Inhaled Short-Acting β -Agonist Formulations

Yaohui Zeng
University of Iowa

Sachinkumar Singh
University of Iowa

Kai Wang
University of Iowa

Please see article for additional authors.

DOI: <https://doi.org/10.1002/jcph.1045>

© 2017, The Authors.

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

J Clin Pharmacol. 2018 Apr;58(4):457-465. doi: <https://doi.org/10.1002/jcph.1045>. Epub 2017 Dec 27.

Hosted by Iowa Research Online. For more information please contact: lib-ir@uiowa.edu.

Effect of Study Design on Sample Size in Studies Intended to Evaluate Bioequivalence of Inhaled Short-Acting β -Agonist Formulations

The Journal of Clinical Pharmacology
2018, 58(4) 457–465
© 2017, The Authors. *The Journal of Clinical Pharmacology* published by Wiley Periodicals, Inc. on behalf of American College of Clinical Pharmacology
DOI: 10.1002/jcph.1045

Yaohui Zeng, MS^{1*}, Sachinkumar Singh, MBBS, PhD^{2*} , Kai Wang, PhD¹, and Richard C. Ahrens, MD²

Abstract

Pharmacodynamic studies that use methacholine challenge to assess bioequivalence of generic and innovator albuterol formulations are generally designed per published Food and Drug Administration guidance, with 3 reference doses and 1 test dose (3-by-1 design). These studies are challenging and expensive to conduct, typically requiring large sample sizes. We proposed 14 modified study designs as alternatives to the Food and Drug Administration–recommended 3-by-1 design, hypothesizing that adding reference and/or test doses would reduce sample size and cost. We used Monte Carlo simulation to estimate sample size. Simulation inputs were selected based on published studies and our own experience with this type of trial. We also estimated effects of these modified study designs on study cost. Most of these altered designs reduced sample size and cost relative to the 3-by-1 design, some decreasing cost by more than 40%. The most effective single study dose to add was 180 μg of test formulation, which resulted in an estimated 30% relative cost reduction. Adding a single test dose of 90 μg was less effective, producing only a 13% cost reduction. Adding a lone reference dose of either 180, 270, or 360 μg yielded little benefit (less than 10% cost reduction), whereas adding 720 μg resulted in a 19% cost reduction. Of the 14 study design modifications we evaluated, the most effective was addition of both a 90- μg test dose and a 720- μg reference dose (42% cost reduction). Combining a 180- μg test dose and a 720- μg reference dose produced an estimated 36% cost reduction.

Keywords

Monte Carlo simulation, E_{max} model, asthma, cost analysis, pharmacodynamics

Inhaled short-acting β -agonists such as albuterol play a critical role in the management of asthma. They are important for relief of acute bronchospasm and prevention of exercise-induced bronchospasm.¹ With expiration of the patent for albuterol in 1989, generic formulations of albuterol metered-dose inhalers were developed and marketed starting in 1995.² These generic formulations and the innovator formulations they were based on were subsequently withdrawn from the market because of the phase-out of the chlorofluorocarbon propellants they contained.³ More recently, 4 different branded albuterol formulations that contain hydrofluoroalkane (HFA) propellants have been marketed to replace the withdrawn products: Proventil HFA (albuterol sulfate; Merck & Co, Inc, Whitehouse Station, New Jersey), Ventolin HFA (albuterol sulfate; GlaxoSmithKline, Research Triangle Park, North Carolina), ProAir (albuterol sulfate; Teva Respiratory, LLC, Horsham, Pennsylvania), and Xopenex (levalbuterol tartrate; Sunovion Pharmaceuticals Inc, Marlborough, Massachusetts). Each of these is now a target for generic drug development.

The Food and Drug Administration (FDA) has published draft guidance for the assessment of pharmaceutical equivalence and bioequivalence of the generic versions of these albuterol-containing

products.⁴ This guidance recommends performance of clinical studies that use pharmacodynamic responses to assess in vivo bioequivalence of generic and innovator bioequivalence formulations, contains minimum requirements for designing these studies, and incorporates, by reference, statistical methods for analyzing the results.⁵ “Dose-scale comparison” of the formulations is specified (ie, estimation of the clinical potency of the generic relative to that of the innovator). The FDA denotes this clinical potency of the generic product as F .⁵ For example, in the case

¹ Department of Biostatistics, College of Public Health, University of Iowa, Iowa City, IA, USA

² Stead Family Department of Pediatrics, Carver College of Medicine, University of Iowa, Iowa City, IA, USA

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Submitted for publication 27 June 2017; accepted 23 October 2017.

Corresponding Author:

Richard C. Ahrens, MD, University of Iowa, E-110-G GH, 200 Hawkins Drive, Iowa City, IA 52242

Email: richard-ahrens@uiowa.edu

*Yaohui Zeng and Sachinkumar Singh contributed equally to this article.

of the first chlorofluorocarbon-containing generic to be approved in 1995, it was demonstrated that 1.00 μg of this formulation was equal to 1.01 μg of the brand name, innovator formulation ($F = 1.01$ with 90% confidence interval [CI] of 0.69 to 1.50).²

The pharmacodynamic studies required by the FDA are technically challenging and expensive to conduct. Data collection is time intensive. In the case of methacholine challenge-based studies, measurement of each individual data point requires hours of study coordinator time. For these studies to succeed, study end points must be measured with maximal precision (minimal within-subject variance).^{2,6,7} Multiple, specialized, and highly trained study sites are needed and may be difficult for a study sponsor to identify.

Hence, strategies that maximize study precision, increase statistical power, and reduce the sample size required are needed. We postulated that adding selected study treatments to the minimum FDA-specified study design would increase statistical power and reduce sample size and study cost. In fact, the FDA has invited such experimentation with study design in relevant, published guidance.^{4,5} The purpose of this study was to test this hypothesis using Monte Carlo simulation with inputs obtained from published bioequivalence study results.

Methods

FDA Draft Guidance

This guidance specifies that assessment of clinical bioequivalence of generic (“test”) and innovator (“reference”) metered-dose albuterol formulations should be performed using a “...[s]ingle-dose, double-blind, double dummy, randomized, crossover study that is recommended at minimum to consist of:...” placebo, 90 μg (1 actuation) and 180 μg (2 actuations) of reference formulation, and 90 μg (1 actuation) of test formulation. In other words, a minimum of 3 doses of reference (0, 90, 180 μg) and 1 of test product (90 μg) are required (3-by-1 design).⁴ The FDA guidance allows 2 possible outcomes to measure albuterol effect: percentage increase in forced exhaled volume in 1 second (FEV_1) (bronchodilation) and inhibition of methacholine challenge as indicated by increase in the provocative concentration of methacholine required to produce a 20% decrease in FEV_1 ($\text{PC}_{20}\text{FEV}_1$). In this study we focused on use of $\text{PC}_{20}\text{FEV}_1$ as the outcome measure.

Methacholine challenge is intended to assess the degree of airway responsiveness present in a test subject by having him or her inhale initially small, and then progressively increasing, concentrations of the bronchoconstrictor agent, methacholine. The concentration of methacholine required to provoke bronchospasm

is used as an index of airway responsiveness present: the lower the concentration required, the greater the airway responsiveness. It has long been recognized that asthmatic individuals exhibit increased airway responsiveness by developing bronchospasm at far lower doses than do normal subjects. This hyperresponsiveness is quantitated by estimating the provocative concentration of methacholine that would produce exactly a 20% decrease from baseline in the lung function measure FEV_1 ($\text{PC}_{20}\text{FEV}_1$). $\text{PC}_{20}\text{FEV}_1$ is estimated by plotting the log-transformed methacholine concentration administered at each stage against the percentage decrease in FEV_1 from baseline at that stage and then performing interpolation between the concentrations producing just above and just below a 20% decrease. The study methodology evaluated here makes use of the fact that clinically relevant doses of inhaled albuterol temporarily increase the $\text{PC}_{20}\text{FEV}_1$ by up to 10- to 20-fold via functional antagonism. The magnitude of this albuterol-induced increase in $\text{PC}_{20}\text{FEV}_1$ serves as our pharmacodynamic response of interest.⁸

The analysis suggested by the FDA fits an E_{max} model to the “...mean, or pooled, dose-response data...” to provide point estimates of e_0 , e_{max} , and ed_{50} parameters of the E_{max} model. The E_{max} model was fit to individual data points (pooled dose-response data) in this research. FDA guidance offers 2 methods for using the E_{max} model to estimate the dose of reference product required to produce an effect equivalent to that of 1.0 μg of test product (relative clinical potency or relative bioavailability; F [Please refer to the draft FDA guidance on orlistat for the mathematical definition of F .⁵]). In the sequential approach, the E_{max} model is first fit to reference product data. The resulting model then serves as a standard curve for subsequent estimation of F . Alternatively, F can be estimated simultaneously with the E_{max} parameters (part 1 in Supplementary Online Material). In this work, we have used the latter approach. Bioequivalence (BE) is demonstrated if the 90%CI for F , constructed using the bootstrap procedure, falls entirely within the specified BE interval (0.67 to 1.50).

Changes in Study Design Evaluated

Using Monte Carlo simulation, we investigated 14 modifications of the FDA’s 3-by-1 design (Table 1) which fall into the following groups: (1) the addition of a fourth reference dose (4-by-1 design); (2) the addition of a second test dose (3-by-2 design); and (3) both (4-by-2 design).

Inputs for Monte Carlo Simulation

Studies that assess bioequivalence of reference and test metered-dose inhaler albuterol formulations using methacholine challenge typically include at least

Table 1. Study Designs Evaluated by Monte Carlo Simulation

Number	Study Design Identification	Placebo	Reference Doses (μg)					Test Doses (μg)	
			90	180	270	360	720	90	180
1	3-by-1 ^a	X	X	X				X	
2	3-by-2 [90T]	X	X	X				X,X	
3	3-by-2 [180T]	X	X	X				X	X
4	4-by-1 [180R]	X	X	X,X				X	
5	4-by-1 [270R]	X	X	X	X			X	
6	4-by-1 [360R]	X	X	X		X		X	
7	4-by-1 [720R]	X	X	X			X	X	
8	4-by-2 [180R + 90T]	X	X	X,X				X,X	
9	4-by-2 [270R + 90T]	X	X	X	X			X,X	
10	4-by-2 [360R + 90T]	X	X	X		X		X,X	
11	4-by-2 [720R + 90T]	X	X	X			X	X,X	
12	4-by-2 [180R + 180T]	X	X	X,X				X	X
13	4-by-2 [270R + 180T]	X	X	X	X			X	X
14	4-by-2 [360R + 180T]	X	X	X		X		X	X
15	4-by-2 [720R + 180T]	X	X	X			X	X	X

Treatments listed in brackets next to the study designs indicate supplementary doses added to the standard Food and Drug Administration 3-by-1 design. R and T indicate Reference and Test formulations, respectively.

^aThe 3-by-1 design represents the standard Food and Drug Administration-recommended design for inhaled albuterol bioequivalence studies.

Table 2. Estimate of Within-Subject Variance From Previously Published Studies

Author and Reference	Drug Studied ^a	Within-Subject Variance ^b
Ahrens et al ⁶	SABA	0.794
Prabhakaran et al ⁹	LABA	0.211
Parameswaran et al ¹⁰	SABA	0.130
Higham et al ¹¹	SABA	1.584
Inman et al ¹²	SABA	0.583
Giannini et al ¹³	SABA	1.489
Creticos et al ¹⁴	SABA	1.377
Langley et al ¹⁵	LABA	1.887
Allan et al ¹⁶	LABA	0.720

^aSABA indicates short-acting β -agonist (albuterol); LABA, long-acting β -agonist (formoterol or salmeterol).

^bObtained from data in publication: complete data set¹⁴; s/b ratio, where s is the within-subject error and b is the dose-response slope^{6,10}; treatment effect expressed as “doubling doses” relative to placebo,^{9,11-13,15} or within-subject variance.¹⁶

4 identical crossover treatment periods, each consisting of a single study visit. At each study visit a methacholine challenge is initiated approximately 15 minutes after the administration of metered-dose inhaler study treatment. $\text{PC}_{20}\text{FEV}_1$ is assumed to be log-normally distributed.

Inputs for Monte Carlo simulations presented here were selected based on published studies.^{6,9-16} $\text{Log}_2[\text{PC}_{20}\text{FEV}_1]$ was used as the dependent variable. Inputs for simulations included between- and within-subject variances of 2.5 and 0.5, respectively ($\text{log}_2[\text{PC}_{20}\text{FEV}_1]$ scale). This within-subject variance is near the lower end of the range observed in existing publications (Table 2).^{6,9-16} Treatment effect for each albuterol dose considered in this study was obtained based on a study by Parameswaran et al.¹⁰ This study

had very low within-subject variance and reported mean results for 4 albuterol doses (0, 90, 180, and 360 μg). Fitting an E_{max} model (Treatment effects = $e = e_0 + \frac{e_{\text{max}} * d}{ed_{50} + d}$; where e_0 = baseline response; e_{max} = maximum possible response; d = dose of reference drug; ed_{50} = dose producing 50% of the maximal response.) to these values using nonlinear regression yielded the following model parameters: $\hat{e}_0 = 0.77$; $\hat{ed}_{50} = 70.81$; and $\hat{e}_{\text{max}} = 5.33$. The simulation inputs for the treatment effect were obtained from this model and were 0.8, 3.8, 4.6, 5.0, 5.2, and 5.6, for 0, 90, 180, 270, 360, and 720 μg of albuterol, respectively.

Monte Carlo Simulation Study

Monte Carlo simulation was carried out for the FDA-recommended 3-by-1 design and for each of the 14 modified study designs being considered (Table 1). Within the assessment of each study design, sample size (N) was varied from 12 to 80 in increments of 4. A separate simulation was done for each value of N . In each of these simulations, 500 simulated data sets were generated using a linear model with repeated measures, where subject was treated as a random effect and dose as a categorical fixed effect (ie, independent of E_{max} modeling). At each albuterol dose the response for a subject was the sum of the dose effect computed from the previously fitted E_{max} model, the simulated subject effect, and a simulated error.

For each of these 500 simulated data sets, 1000 bootstrap samples were obtained (resampling with replacement by subject rather than individual data points). For each of these 1000 bootstrap samples, E_{max} model parameters (ie, e_0 , e_{max} , ed_{50}) and F were estimated using the FDA recommended, simultaneous

E_{\max} modeling approach. This was carried out by fitting a nonlinear E_{\max} curve based on individual data points pooled across subjects. The 90%CI of F was then generated based on these 1000 bootstrap samples using the percentile bootstrap method.¹⁷ Similar 90%CIs for F were constructed for each of the 500 simulated data sets. Finally, for each given study design and specific value of N , statistical power was defined as percentage of the 500 simulated data sets that successfully met the predefined BE criterion (ie, had a 90%CI for F falling entirely within the predefined BE limits [0.67, 1.5]). Failures of convergence in estimating the E_{\max} model were infrequent and were treated as failures to meet BE criterion.

Cost Analysis

Alterations in study design that add treatments to the FDA minimum 3-by-1 design require additional crossover study visits for each subject. This will tend to increase cost of carrying out the study. However, it may also reduce sample size, which will tend to reduce cost. To assess the net effect of each modification of study design, we conducted a cost analysis. This assumed 2 study screening visits and 4 to 6 randomized visits, depending on the design. Cost of individual study visits was assumed to be the same for the second screening and randomized visits. The cost of the first screening visit was 1.38 times that of the other visits due to time required for informed consent and evaluation of entrance criteria. A 67% screen failure rate was assumed, with two thirds of these screen failures occurring at the first FDA recommended screening visit⁴ and one third at the second. Cost of recruitment of each subject who entered screening was assumed to be equivalent to 0.5 times the cost of the first screening visit. This allowed site cost for altered designs to be expressed relative to that of the minimum 3-by-1 design recommended by FDA guidance. For example, a cost of 0.6 relative to the FDA minimum design indicates cost for data collection that is 60% of that of the FDA recommended design, for a cost savings of 40%.

Sensitivity Analysis

We evaluated the effect of altering both the within- and the between-subject variances assumed for the simulation on changes in sample size and cost associated with modifying the study design. The study designs selected for this analysis were:

- FDA 3-by-1 design (reference 0, 90, 180; test 90)
- Study design 2: 3-by-2 design (reference 0, 90, 180; test 90, 90)
- Study design 3: 3-by-2 design (reference 0, 90, 180; test 90, 180)

- Study design 10: 4-by-2 design (reference 0, 90, 180, 360; test 90, 90)
- Study design 14: 4-by-2 design (reference 0, 90, 180, 360; test 90, 180)

We also evaluated the effect of altering the true value of F that was assumed in the simulation. F values of 0.90, 0.95, 1.00, 1.05, and 1.11 were studied. The study designs included in this analysis were the FDA 3-by-1 reference design and the 2 designs recommended in the Conclusions section below:

- FDA 3-by-1 design (reference 0, 90, 180; test 90)
- Study design 3: 3-by-2 design (reference 0, 90, 180; test 90, 180)
- Study design 11: 4-by-2 design (reference 0, 90, 180, 720; test 90, 90)

All statistical analyses were performed using SAS (version 9.4; SAS Institute, Cary, North Carolina) and R Statistical Software (version 3.4.0; R Foundation for Statistical Computing, Vienna, Austria).

Results

Study simulation results support the hypothesis that adding selected study treatments to the minimum FDA-specified study design can increase statistical power, reduce relative sample size, and reduce relative study cost. The FDA-recommended 3-by-1 design required a sample size of 39 subjects to establish bioequivalence with approximately 80% power and 54 for 90% power. Adding a fourth reference dose of 270 μg , 360 μg , or 720 μg (4-by-1 design) reduced sample size to 33, 32, and 29 for 80% power and to 45, 39, and 37 for 90% power, respectively (Figure 1B). Adding a second test dose of 90 μg (3-by-2 design) produced similar reductions in sample size. However, adding a second test dose of 180 μg was more effective, reducing sample size to 25 for 80% power and to 32 for 90% power (Figure 1A). Adding both a second test dose and a fourth reference dose (4-by-2 design) produced the greatest reductions, with minimum sample sizes of 19 and 25 for 80% and 90% power, respectively (Figure 1C, 1D).

Cost analysis demonstrated that a cost reduction of approximately 30% could be achieved with addition of a 180- μg dose of test alone (3-by-2 design) or addition of a 90- μg dose of test and a 360- μg dose of reference (4-by-2 design). A cost reduction of about 40% could be achieved with addition of 90 μg of test and 720 μg of reference or 180 μg of test and either 180, 270, or 720 μg of reference (Figure 2).

Sensitivity analysis showed that altering the between-subject variance had no effect on the results of the simulation in the absence of missing data (results

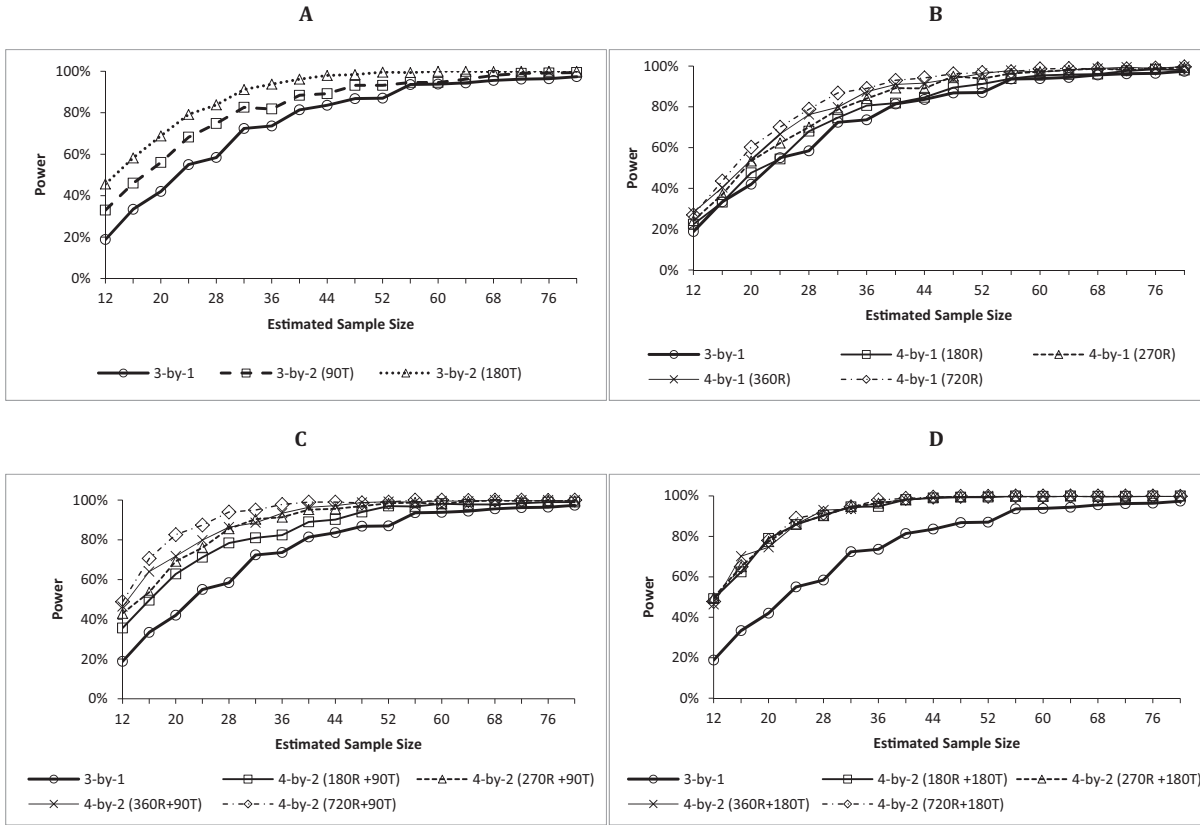


Figure 1. Relationship of sample size and power for the modified study designs and the standard Food and Drug Administration 3-by-1 design. Sample size was estimated using Monte Carlo study simulation for (A) 3-by-2 designs; (B) 4-by-1 designs; (C) 4-by-2 designs with replication of the test dose of 90 μ g; and (D) 4-by-2 designs with addition of a test dose of 180 μ g. See Table 1 for definitions of study designs being evaluated.

Table 3. Sensitivity Analysis by Varying the Within-Subject Variance for 4 Study Designs

Study Designs	Within-Subject Variance					
	0.25	0.36	0.49	0.64	0.81	1
Standard FDA 3-by-1	20	29	40	50	62	78
3-by-2 (90T)	0.80	0.83	0.85	0.86	0.87	0.90
	0.88	0.91	0.93	0.94	0.95	0.98
3-by-2 (180T)	0.60	0.66	0.63	0.64	0.61	0.65
	0.66	0.72	0.68	0.70	0.67	0.72
4-by-2 (360R + 90T)	0.60	0.62	0.58	0.62	0.61	0.62
	0.71	0.74	0.68	0.74	0.73	0.73
4-by-2 (360R + 180T)	0.40	0.52	0.53	0.56	0.56	0.55
	0.48	0.62	0.62	0.67	0.67	0.66

^aRatio of sample size for 80% power to that of the standard Food and Drug Administration 3-by-1 design.

^bRatio of study cost associated with 80% power to that of the standard Food and Drug Administration 3-by-1 design.

not shown). This is as expected with a crossover design. In contrast, sample size progressively increased as within-subject variance increased. However, the ratio of sample size estimate for a given altered design to that of the standard 3-by-1 design, an indicator of benefit of altered study design, changed little across the range of within-subject variances studied (Table 3). Consequently, results of the cost analysis also changed little across the range of within-subject variances studied. Similarly, sensitivity analysis showed that

substantial cost savings persisted across the range of *F* values between 0.90 and 1.11 (Table 4).

Discussion

Sample size for pharmacodynamic studies that assess bioequivalence of generic inhaled albuterol formulations is a function not only of the desired statistical power and expected within-subject variance but also of the specific study design used. This study used Monte

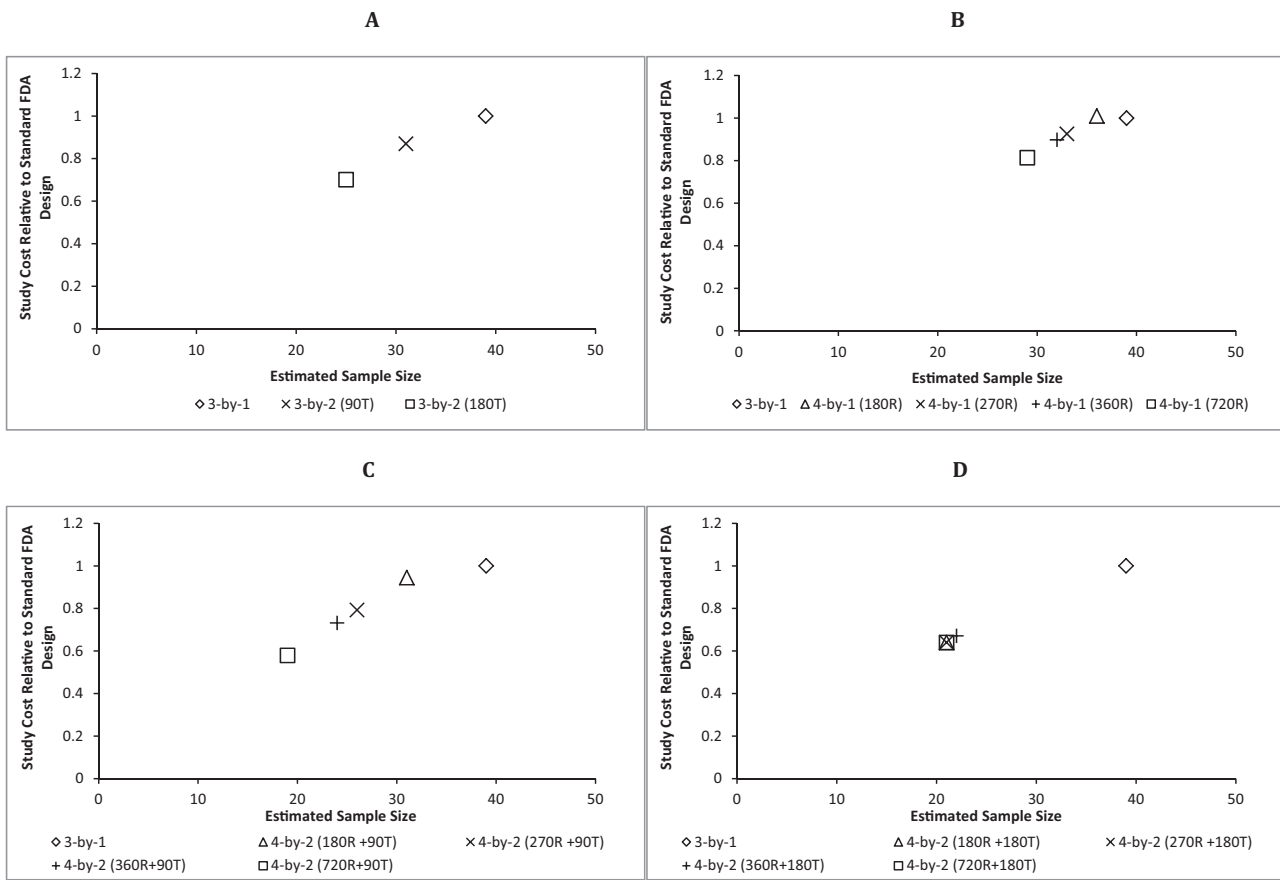


Figure 2. Analysis of the cost of modified study designs relative to the cost of the standard Food and Drug Administration (FDA) 3-by-1 design. Analysis assumes sample size required to achieve 80% power for (A) 3-by-2 study designs; (B) 4-by-1 study designs; (C) 4-by-2 study designs with replication of test dose of 90 μg ; and (D) 4-by-2 study designs with an additional test dose of 180 μg . See Table 1 for definitions of study designs being evaluated.

Table 4. Effect of Varying the “True” Value of F^3 Assumed in the Simulation on Relative Sample Size and Study Cost

Study Designs		“True” Value of F^3				
		0.9	0.95	1.00	1.05	1.11
Standard FDA 3-by-1	Estimated Sample Size	66	46	39	40	42
3-by-2 (180T)	Relative Sample Size ^b	0.48	0.61	0.64	0.65	0.79
	Relative Study Cost ^c	0.53	0.67	0.70	0.71	0.86
4-by-2 (720R + 90T)	Relative Sample Size ^b	0.45	0.46	0.49	0.55	0.60
	Relative Study Cost ^c	0.54	0.54	0.58	0.65	0.71

^aClinical potency of test formulation relative to reference formulation.

^bRatio of sample size for 80% power to that of the standard FDA 3-by-1 design.

^cRatio of study cost associated with 80% power to that of the standard FDA 3-by-1 design.

Carlo simulation to evaluate the reduction in estimated sample size associated with the addition of extra test and/or reference doses (Table 1) to the standard, FDA-recommended “minimum” 3-by-1 study design (Figure 3). This produced a corresponding reduction in cost associated with subject recruitment, screening, and randomized study. The greatest reductions in cost (approximately 40%) were obtained with the addition of a 90 μg or 180 μg test dose and fourth reference dose

(4-by-2 design). Addition of a 180 μg test dose alone (3-by-2 design) produced results that were almost as good (approximately 30% reduction), but addition of a 90 μg test dose alone was less effective.

Simulations supporting these conclusions used a single set of reasonable assumptions for the values of within- and between-subject variance. However, sensitivity analysis demonstrated that the observed, proportional decreases in sample size and cost changed

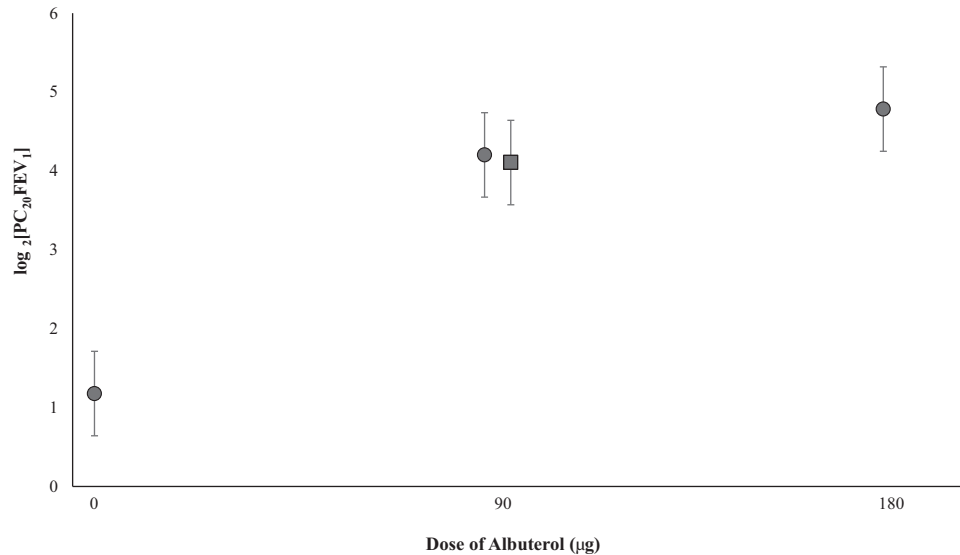


Figure 3. Example dose-response for test (solid square) and reference (solid circles) albuterol metered-dose inhalers, 90 μg per actuation. Points indicate mean $\log_2[\text{PC}_{20}\text{FEV}_1]$ with 95% confidence interval for this mean. Data are obtained from a single, representative simulated data set (see Methods section, Monte Carlo Simulation).

little across a range of between- and within-subject variances (Table 3). That is, the percentage decrease in sample size and cost produced by altering design appears to be independent of variance assumptions.

It is not surprising that obtaining more information from each subject by adding study treatments and periods to the trial reduced the required sample size. However, we did not anticipate that adding a 180 μg test dose would be more effective in reducing sample size than adding a second 90 μg test dose to the FDA-recommended 3-by-1 design. We speculate that this is because adding a second 90 μg test dose only provides information about the effect of the test product relative to the reference product, whereas adding a 180 μg test dose provides information both about the effects of test vs reference product and about the dose-response relationship (ie, the effect of 180 μg versus 90 μg of test product). Another advantage of adding a 180 μg rather than a 90 μg test dose to the study provides independent information about both the test and reference product dose-response curves. This allows the underlying assumption of parallelism between these 2 curves to be statistically tested and thus provides an additional level of rigor to the analysis.

For the Monte Carlo simulation used in this study, specific methodologic choices were made that allow a range of study designs to be evaluated with an identical mathematical approach. First, we fitted the E_{max} model to individual data points rather than means by treatment because the former is applicable to 3 or more reference treatment doses (including placebo as a dose of 0), whereas the latter applies only to 3

reference doses. Furthermore, it can be shown that for the FDA-specified 3-by-1 study design, fitting the E_{max} model to individual “pooled” data points or mean $\log_2[\text{PC}_{20}\text{FEV}_1]$ values for 3 reference treatment yields identical estimates of model parameters (part 2 in Supplementary Online Material). Second, we fitted the E_{max} model and estimated F simultaneously rather than sequentially because the former can be used with any number of test treatment doses, but the latter can accommodate only a single test dose. In cases incorporating only 1 test dose, it can be shown that the simultaneous and sequential methods produced identical results (part 3 in Supplementary Online Material).

Although work presented here focuses on the importance of choice of study design, within-subject variance continues to be a critical factor in determining sample size. Within-subject variance observed in prior work ranged from 0.130 to 1.887 (Table 2).^{6,9–16} Corresponding sample size estimates ranged from 12 to 130 subjects to achieve 80% power (assuming the FDA-recommended 3-by-1 study design and the simulation inputs for dose effects and between-subject variance listed above). Of note, the smallest variance¹⁰ was produced by a laboratory with many years of experience in conducting methacholine challenges in the context of clinical trials, whereas 1 of the largest variances⁶ came from data collected by a contract research organization that at the time had limited experience in this kind of work. This, as well as our own recent unpublished experience, suggests that selection of study sites that are highly experienced in this kind of research

is an important consideration when planning future studies.

Previously, Lalonde et al showed that interoccasion variability in E_{\max} pharmacodynamic model parameters can have an effect on ed_{50} and relative potency in studies that use bronchodilation to assess albuterol product bioequivalence.¹⁸ Unfortunately, the bronchoprovocation study designs evaluated in this work do not allow an assessment of interoccasion variability in model parameters because subjects can receive only 1 study treatment per study visit.

An additional factor that has an important effect on sample size is the “true” potency of the test product (relative to the reference product) that is assumed for study simulation. For the primary work presented here, a true potency of 1.00 was assumed (ie, the test and reference are exactly equivalent, and the true value of F is 1.00). However, in actual practice it may be prudent to assume a true F value above or below 1.00. This will act to increase the required sample size. To illustrate this, we carried out simulations as described above using the standard FDA-recommended 3-by-1 design. F values of 0.80, 0.85, 0.90, 0.95, 1.00, 1.05, 1.11, 1.18, 1.25, and 1.33 yielded sample size estimates of 150, 100, 66, 46, 39, 40, 42, 57, 100, and 250, respectively. This underscores the risk associated with studying generic formulations that are suspected of having true potencies relative to the reference product that depart from $F = 1.00$ by large amounts. For example, sample sizes appear to increase dramatically when F is changed to be 0.80 or 1.33. In contrast, our simulation results indicate that when the true F is varied between 0.90 and 1.11, estimated sample sizes do not differ by large amounts. They also indicate use of either of the modified study designs recommended below can be expected to achieve cost savings across this range of assumed true F values.

Conclusions

In summary, we used Monte Carlo simulation to evaluate the effect of altering the design of studies intended to assess in vivo bioequivalence of different albuterol formulations on study sample size and cost. Most of the altered study designs we evaluated lowered sample size and cost relative to those of the minimum design recommended by the FDA. These relative reductions in sample size and cost appear to be unaffected by changing the between- and within-subject variance assumptions. Based on these analyses, we recommend using a study design that adds either (1) a combination of a fourth reference dose of 720 μg and a second test dose of 90 μg (42% reduction in cost) or (2) a 180 μg test dose alone (30% reduction in cost).

Conflict of Interest

The authors do not have any conflicts of interest to report.

Author Contributions

Yaohui Zeng and Sachinkumar Singh wrote the first draft of the manuscript in which all authors provided their input. Yaohui Zeng, Sachinkumar Singh, Kai Wang, and Richard Ahrens designed the research. Yaohui Zeng and Sachinkumar Singh analyzed the data.

References

1. National Heart Lung and Blood Institute. *Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma*. 2007.
2. Stewart B, Ahrens RC, Carrier S, et al. Demonstration of in vivo bioequivalence of a generic albuterol metered-dose inhaler to Ventolin. *Chest*. 2000;117(3):714-721.
3. Hendeles L, Colice GL, Meyer RJ. Withdrawal of albuterol inhalers containing chlorofluorocarbon propellants. *N Engl J Med*. 2007;356(13):1344-1351.
4. US Food and Drug Administration. Draft Guidance on Albuterol Sulfate. 2016.
5. US Food and Drug Administration. Draft Guidance on Orlistat. 2010.
6. Ahrens RC, Hendeles L, Clarke WR, et al. Therapeutic equivalence of Spiros dry powder inhaler and Ventolin metered dose inhaler. A bioassay using methacholine. *Am J Respir Crit Care Med*. 1999;160(4):1238-1243.
7. Newhouse MT, Patel P, Parry-Billings M. Protection against methacholine-induced bronchospasm: salbutamol pMDI versus Clickhaler DPI. *Eur Respir J*. 2003;21(5):816-820.
8. Crapo RO, Casaburi R, Coates AL, et al. Guidelines for methacholine and exercise challenge testing-1999. This official statement of the American Thoracic Society was adopted by the ATS Board of Directors, July 1999. *Am J Respir Crit Care Med*. 2000;161(1):309-329.
9. Prabhakaran S, Shuster J, Ahrens R, Hendeles L. Methacholine challenge as a clinical bioassay of pulmonary delivery of a long-acting beta(2)-adrenergic agonist. *Pharmacotherapy*. 2011;31(5):449-457.
10. Parameswaran KN, Inman MD, Ekholm BP, et al. Protection against methacholine bronchoconstriction to assess relative potency of inhaled beta2-agonist. *Am J Respir Crit Care Med*. 1999;160(1):354-357.
11. Higham MA, Sharara AM, Wilson P, Jenkins RJ, Glendenning GA, Ind PW. Dose equivalence and bronchoprotective effects of salmeterol and salbutamol in asthma. *Thorax*. 1997;52(11):975-980.
12. Inman MD, Hamilton AL, Kerstjens HA, Watson RM, O'Byrne PM. The utility of methacholine airway responsiveness measurements in evaluating anti-asthma drugs. *J Allergy Clin Immunol*. 1998;101(3):342-348.
13. Giannini D, Di Franco A, Bacci E, et al. The protective effect of salbutamol inhaled using different devices on methacholine bronchoconstriction. *Chest*. 2000;117(5):1319-1323.
14. Creticos PS, Adams WP, Petty BG, et al. A methacholine challenge dose-response study for development of a pharmacodynamic bioequivalence methodology for albuterol metered-dose inhalers. *J Allergy Clin Immunol*. 2002;110(5):713-720.
15. Langley SJ, Allen D, McDonnell B, et al. The effect of reducing the fine-particle mass of salmeterol from metered-dose inhalers on bronchodilatation and bronchoprotection against

- methacholine challenge: a randomized, placebo-controlled, double-blind, crossover study. *Clin Ther.* 2005;27(7):1004-1012.
16. Allan R, Haughie S, Ward J, Ahrens RC. A randomized double-blind placebo and active-controlled five-way crossover study to assess the dose responsiveness of methacholine-induced bronchial hyperreactivity to single inhaled doses of Advair[®] Diskus[®] in adult asthmatics. *Am J Respir Crit Care Med.* 2017;195:A3196-A3196.
 17. Wilcox RR. *Introduction to Robust Estimation and Hypothesis Testing*, 3rd ed: Amsterdam, Boston: Academic Press; 2012.
 18. Lalonde RL, Ouellet D, Kimanani EK, Potvin D, Vaughan LM, Hill MR. Comparison of different methods to evaluate population dose-response and relative potency: importance of interoccasion variability. *J Pharmacokinet Biopharm.* 1999;27(1):67-83.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website.