Original Research

DIA DEVELOP INNOVATE ADVANCE

Therapeutic Innovation & Regulatory Science
1-6

DOI: 10.1177/2168479016651661
tirs.sagepub.com

Stratification, Hypothesis Testing, and Clinical Trial Simulation in Pediatric Drug Development

Ann W. McMahon, MD, MS[1], Kevin Watt, MD, MPH[2], Jian Wang, PhD[3], Dionna Green, MD[3], Ram Tiwari, PhD[4], and Gilbert J. Burckart, PharmD[3]

Abstract

Background: Pediatric drug development is plagued by small sample sizes, unvalidated clinical endpoints, and limited studies. *Objectives:* The objective of this study was to determine whether age stratification within the pediatric population could be used to (1) assess response to a pharmacologic intervention and to (2) design future trials based upon published stratified disease data using clinical trial simulation (CTS). *Methods:* Data available from the literature for Kawasaki disease (KD) was used in the model. Age-stratified CTS for a theoretical new drug was conducted. *Results:* Population-specific differences due to age might affect trial success if not taken into account. CTS predicted inflammatory indices, and inclusion cutoff significantly altered the trial outcome. Finally, altered pharmacokinetics/pharmacodynamics in varying age groups of KD patients may alter drug exposure and response. *Conclusions:* If assumptions regarding a pediatric disease process, such as KD, do not include age stratification with inclusion or response, then the wrong decision could result with regard to age-appropriateness or approval of a drug.

Keywords

Kawasaki disease, age groups, modeling, pharmacokinetics pharmacodynamics

Introduction

Investigators and regulators involved in planning, conducting, and evaluating pediatric clinical research struggle with small sample sizes, unvalidated endpoints, and limited and sometimes unsuccessful trials. Within the context of drug development, a failed trial is one in which the trial is unsuccessful in establishing safety and efficacy in the study population.[1] When a trial fails, it is important to investigate the root cause(s) for its failure, and what the best course of action is for repeat or future trials. One possible reason for the trial failure could be not taking age into account and/or not stratifying the study population by age. If the study population is not stratified by age, variation in disease manifestations in the different strata may be missed. Furthermore, if the cohort is not age stratified, one might miss observing the heterogeneity in therapy response across different age strata. Finally, if the study population is not stratified by age, the need for adjustments in drug dosing by age could be missed because of differences in pharmacokinetics (PK) or pharmacodynamics (PD).

Age stratification in the pediatric population in a drug development trial does not ensure that the disease process and efficacy is assessed. In fact, out of more than 300 products studied under the Best Pharmaceuticals for Children Act (BPCA) and the Pediatric Research Equity Act (PREA), we could only identify 3 products (Dymista, Sabril, Topopmax) where the assessment of efficacy was different between pediatric age groups. Therefore, the assumption that differences in disease processes between pediatric age groups may occur requires special consideration in developing those trials.

One way to test the impact of age stratification on the study population and trial outcome is to simulate the clinical trial. Clinical trial simulation (CTS) is a modeling approach that makes assumptions about different aspects of a clinical trial and postulates outcomes based on those assumptions. Thereby, the effect on outcome of changes in various elements of the trial can be tested prior to the actual conduct of the trial. By simulating variations of these elements, CTS can aid in

[1] Office of Pediatric Therapeutics, Office of the Commissioner, Food and Drug Administration, Silver Spring, MD, USA
[2] Duke University Medical Center, Durham, NC, USA
[3] Office of Clinical Pharmacology, Office of Translational Sciences, Food and Drug Administration, Silver Spring, MD, USA
[4] Office of Biostatistics, Office of Translational Sciences, Food and Drug Administration, Silver Spring, MD, USA

Submitted 24-Feb-2016; accepted 3-May-2016

Corresponding Author:
Ann W. McMahon, MD, MS, Office of Pediatric Therapeutics, Office of the Commissioner, Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993, USA.
Email: ann.mcmahon@fda.hhs.gov

Table 1. Hypotheses That Were Tested Using Clinical Trial Simulation.

Hypothesis	Design	Illustrates
The disease is the same in 2 different age strata, but the disease manifestations are different.	Enroll patients based upon their laboratory results. Assuming a uniform response to therapy across ages, what effect will this alteration in recruitment have?	Study population is changed when CRP $\geq$8 mg/dL is used as an inclusion criterion, 65% infants with KD who would not develop CAA are excluded.
The disease is different in 2 different strata and will respond differently to therapy.	Assume a difference in response to therapy. To what extent can this difference jeopardize the success of the trial?	Drug *X* + IVIG decreases risk of coronary artery aneurysm in infants but not children, but if infants and children are analyzed together the effect of drug *X* will be decreased
The dosing is appropriate in one age strata but not in another.	Assume that a single dose per BSA is inappropriate in one or more age group. How far of a change in response would be required to make the whole trial fail?	Box plots of AUC among different age groups of pediatric patients after receiving 6 mg/kg dose of drug *X*. Infants have overexposure to drug *X* at 6 mg/dose.

Abbreviations: AUC, area under the curve; BSA, body surface area; CRP, C-reactive protein; IVIG, intravenous immunoglobulin; KD, Kawasaki disease.

selecting trial designs that have a higher chance of successfully answering the question being investigated (such as whether or not a drug is safe and effective). CTS has been used to better understand the reasons for failed trials,[2,3] and is also increasingly being utilized to answer trial design questions by testing hypotheses.[4–6] Without CTS, the investigation might require a large sample size to gain evidence within each stratum to investigate a design question such as whether or not the trial may fail because of a lack of age stratification. Demonstrating the importance of age stratification in diseases and for drugs targeting those diseases is critical when the age of the child determines the natural history of the illness. One example of such an illness is Kawasaki disease (KD).

KD is an acute febrile illness involving generalized inflammation of the small and medium-sized blood vessels. It occurs primarily in childhood, and it is estimated that approximately 3000 children are hospitalized annually with KD in the US. The highest incidence is reported in Asian countries. The etiology of KD remains unknown. It has been postulated that the causative agent is likely infectious, although both genetic predisposition and autoimmune reactions have also been suggested. KD is characterized by an array of clinical manifestations, including prolonged fever, rash, nonexudative bilateral conjunctivitis, erythema and edema of the hands and feet, ''strawberry'' tongue, and cervical lymphadenopathy. Although mainly a self-limited illness, KD can lead to coronary artery abnormalities (CAA) in some children and is the leading cause of acquired heart disease during childhood.

Age appears to be an important factor in the pathogenesis of KD and for the risk of sequela such as CAA. A recent study by Song et al[7] identified age-specific clinical characteristics and laboratory findings in a large cohort of infants and children with KD. The objective of the investigation herein was to pose 3 different scenarios, using CTS as one tool to study the critical nature of age stratification in drug development using KD as the disease of interest. Data from Song et al were used in the simulation designs described below.

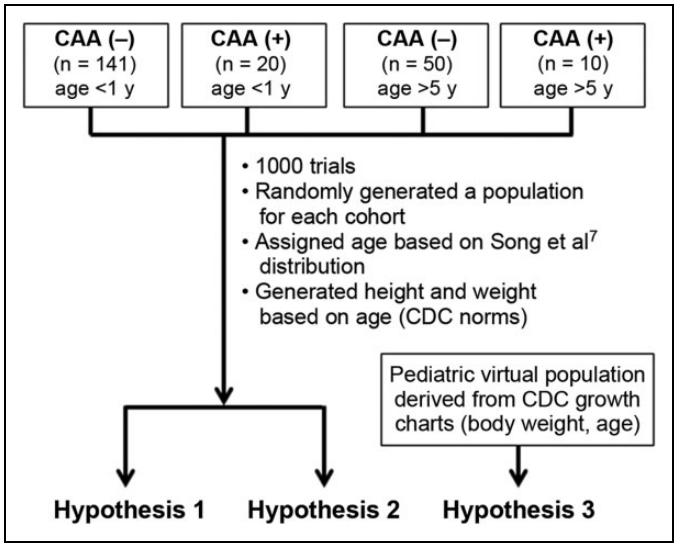

Figure 1. Diagram of the patient population in which the 3 hypotheses were addressed in the 3 simulation exercises.

Methods

The methods in this exercise used 3 different strategies specific to 3 different hypotheses and 3 different designs (see Table 1). We simulated 1000 trials, which included populations of infants/children with the age stratification for each population based on Song et al[7] (Figure 1).

To simulate the demographics for each cohort, we used the PK-Sim (Bayer Technology Services GmbH, Leverkusen, Germany) population generator. The population generator first used a uniform random-number distribution to determine age for each simulated individual based on the age distribution reported in Song et al.[7] The population generator then assigned a height according to the height distribution defined in the National Health and Nutrition Examination Survey (NHANES) database for the given age.[8] With age and height defined, the algorithm then assigned a weight based on these values.

Hypothesis 1: The disease is the same in 2 different age strata, but the disease manifestations are different

If we assume that the Song et al[7] data represent real-world observation, any substantial change in the number of children in each age strata could bias a trial's results. In order to test this hypothesis, we used a C-reactive protein (CRP) concentration of $\geq$8 mg/dL as a hypothetical inclusion criterion, since CRP tended to be $\geq$8 in children older than 5 years with coronary artery disease and <8 in infants younger than 1 year. Because CRP was significantly higher in children compared to infants,[7] this inclusion criterion would exclude a substantial number of infants.

In order to generate a population based on the new inclusion criteria, we used the normal function in Stata 13.1 (StataCorp LP, College Station, TX) to randomly generate CRP concentrations based on the distribution reported in Song et al[7] for each of the 1000 participants in the trial population. The mean CRP value was calculated for each individual from the 1000 simulations.

Hypothesis 2: The disease is different in 2 different strata and will respond differently to therapy

We assumed that hypothetical drug *X* was submitted to the US Food and Drug Administration (FDA) for prevention of CAA in children with KD. A mean $\pm$ standard deviation absolute risk reduction (ARR) of 7% $\pm$ 1.5% in infants and 2% $\pm$ 0.5% in children was assumed. We simulated 1000 participants and generated the ARR for each age cohort based on the above distribution. For each simulated population, we calculated the number of subjects that would develop CAA by subtracting ARR from CAA incidence reported in Song et al.[7] The mean number of subjects per age cohort and CAA status was calculated to determine the ARR across the 1000 simulated populations.

Hypothesis 3: The dosing is appropriate in one age stratum but not in another

To test this hypothesis, a database for simulation was used to generate a complete range of ages (0-20 years) and weights (2.35-101 kg) for virtual subjects derived from the US Centers for Disease Control (CDC) growth charts.[8] A 2-compartment model with first-order elimination was used as the hypothetical population PK model, which incorporated a maturation function to account for changes in apparent oral clearance (CL/F) due to cytochrome P450 3A4 (CYP3A4) maturation in infants.[9] In the model, the effect of body weight and age on drug clearance (CL/F) was described as follows[10]:

$$\mathrm{CL} = \mathrm{TVCL}^{*}(\mathrm{WT}/70\mathrm{kg})^{0.75} * (1 - (1-\beta)^{*}\mathrm{e}^{-\mathrm{Age}*0.0693/\mathrm{TCL}}),$$

where TVCL represents the typical value of clearance for the adult population, β is the fractional clearance at birth, and TCL is the maturation half-life of CYP3A4 enzyme. For the simulation below, $\beta = 0.2$ and TCL = 3 months. These assumptions, regarding the maximum change in CL/F and the maturation rate, are consistent with differences in CYP3A4 expression in newborns.[9] Simulations were performed for 4374 subjects receiving a 6-mg/kg intravenous bolus dose of drug *X*. The predicted areas under the curve (AUCs) among different prespecified age groups were calculated.

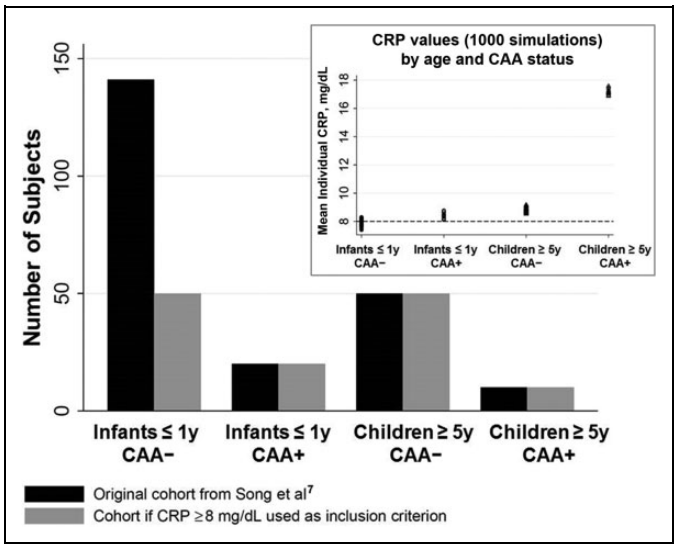

Figure 2. Hypothesis 1: When C-reactive protein (CRP) $\geq$8 mg/dL is used as an inclusion criterion, 90 infants from the original cohort who did not develop coronary artery abnormalities (CAA) are excluded. The exclusion of 64% of the original cohort of infants without CAA changes the rate of CAA from 12% to 29% in infants, making intravenous immunoglobulin appear less effective in infants. *Inset:* The reason that CRP $\geq$8 mg/dL biases the results is that distribution of CRP concentrations is age dependent.

Results

Hypothesis 1: The disease is the same in 2 different age strata, but the disease manifestations are different

The proportion of infants and children who developed CAA after intravenous immunoglobulin (IVIG) in Song et al[7] was 12.4% and 16.7%, respectively. If CRP $\geq$8 mg/dL is used as an inclusion criterion, 91 of 141 infants with KD who would not develop CAA are excluded, and no infants who would develop CAA are excluded. This scenario increases the rate of CAA from 12.4% to 28.6% in infants. Because children had higher CRP levels, using an inclusion criterion of CRP $\geq$8 did not alter the population in the older children, and the rate of CAA remained 16.7%. Consequently, in a CAA prevention trial with CRP $\geq$8 mg/dL as an inclusion criterion, IVIG appears to be less effective in infants (Figure 2).

Hypothesis 2: The disease is different in 2 different strata and will respond differently to therapy

If the approval threshold was set at an ARR of 5%, drug *X* would be approved in infants but not children. If the age

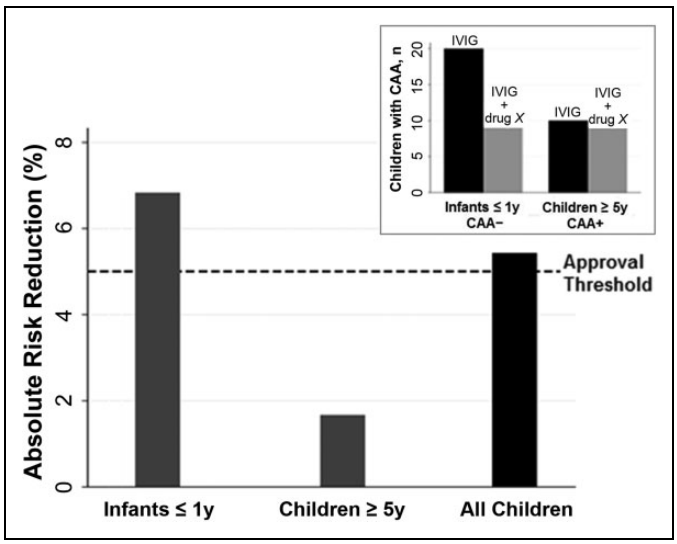

Figure 3. Hypothesis 2: The disease is different in 2 different age strata and will respond differently to therapy. If FDA sets the approval threshold as an absolute risk reduction (ARR) of 5%, drug *X* would be approved in infants but not children. If the age cohorts were combined, the drug would be approved in all children, exposing children ≥5 years to a drug with little benefit. *Inset:* The number of children in the age cohort that developed coronary artery abnormalities (CAA) on intravenous immunoglobulin (IVIG) alone and IVIG + drug *X*.

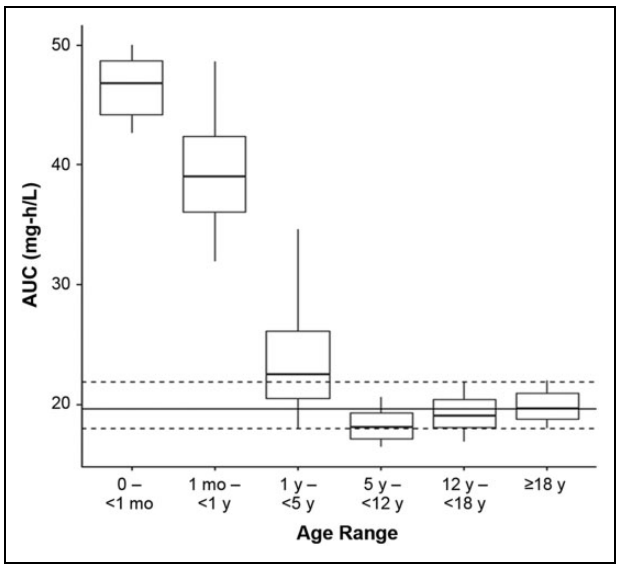

Figure 4. Hypothesis 3: Simulation of pediatric patients of different pre-specified age groups (shown on the *x*-axis) receiving 6 mg/kg intravenous bolus dose of drug *X*. The solid line represents the median area under the curve (AUC) in adult patients, and the dashed lines represent the targeted therapeutic range.

cohorts were combined, the drug would be approved in all children, exposing children aged 5 years and older to a drug with little benefit (Figure 3).

Hypothesis 3: The dosing is appropriate in one age stratum but not in another

The dotted line represents the target therapeutic range of the drug exposure in patients, assuming a similar exposure-response relationship across the age range (Figure 4). In this case, body weight–adjusted dosing at 6 mg/kg is acceptable for children older than 1 year but can lead to a more than 3-fold overexposure in infants younger than 1 year. Appropriate dose adjustments by age would be needed for safety.

Discussion

Because of the small size of many pediatric clinical trials, it is impossible to directly test the hypothesis that age stratification in pediatric trials makes a difference in the clinical trial outcome. Therefore, CTS provides a method to project the possible outcome of age differences on disease and/or response to therapy with or without age stratification. CTS utilizes 2 types of models. Pharmacologically based (PK-PD) models incorporating concentration and response data can provide insight into predictions of safety and efficacy of a drug and can inform optimal dosing and regimen selection. However, this type of model can only be employed when PK-PD measurements have been previously collected in some patient population. The second type of model simulates other important parameters in the trial, such as inclusion criteria, dropout, compliance, placebo response, non-inferiority margins, etc. It is the second type of simulation that was used in this evaluation. Kawasaki disease was selected as the case study here because of the large difference in clinical phenotype observed between infants and older children with KD.[7] Whereas there are other examples of pediatric diseases that have different clinical manifestations at different ages, KD is a pediatric illness with dramatic differences between age groups that can be objectively measured.

In the hypothesis of the first simulation design, a cutoff value that favored one stratum over another was assumed and had a sizable impact on the outcome of the clinical trial. In the literature there are examples of simulations that predict clinical subgroups of neuroblastoma by age, using predicted doubling rates of different malignant cells to predict age-specific severity of malignant syndromes.[4] If one age group were enrolled in a clinical trial, there would not be representative inclusion of different clinical severities in the subjects enrolled. Another example of the potential for age-related misclassification is microalbuminuria in type 1 diabetes.[11] Persistent microalbuminuria was significantly associated with diabetes duration only in older children. In this situation, if a cutoff of persistent microalbuminuria were chosen to be low, older individuals could be excluded from a trial and the higher rates never detected.

In the second hypothesis, the absolute risk reduction from a drug *X* varied in the 2 age groups, and there was no age stratification. This model assumes that only drug *X* is metabolized by the hypothetical route; other drugs may not behave in that manner (aka, equations) or have additional competing routes or

complexity such as biologics. This example may not really be considered CTS but simulating a typical population pharmacokinetic (PopPK) model factoring for age and maturation-related factors on oral clearance (CL/F). In this situation, age stratification would have allowed an appropriate age-specific approval. An inappropriate approval for all age groups meant that some children would be exposed to the drug unnecessarily. A potential example of the avoidance of this situation is in the lack of approval of albuterol for children younger than 4 years. Albuterol in this age group was reviewed at the FDA Pediatric Advisory Committee (PAC) on December 8, 2009.[12] There were 3 submitted trials in children younger than 4, none of which showed efficacy in that population. This is in contrast to children 4 years of age and older, who have an FDA-approved indication for treatment or prevention of bronchospasm in patients with reversible obstructive airway disease and the prevention of exercise-induced bronchospasm. If there had not been age stratification in this situation, children younger than 4 years would be inappropriately exposed to albuterol.

Finally, the third hypothesis demonstrates that age stratification for the purpose of determining drug exposure means that optimal dosing can be determined for each age group. This concept has important efficacy and safety implications. An example of this can be seen with the drug guanfacine. Guanfacine is a nonstimulant agent FDA-approved for the treatment of attention-deficit hyperactivity disorder (ADHD) at a daily dose of up to 4 mg in pediatric patients aged 6 years and older. Subsequently, following drug approval, the drug sponsor proposed to increase the dose to 7 mg/day for adolescents aged 13-17 years. This was due to findings suggesting that adolescents receiving the 4-mg/day dose were being underdosed based on differences in body weight compared to younger children. When given the same dose of 4 mg/day, exposures in children (6-12 years of age) were roughly 40% higher than that in adolescents.[13] A more appropriate dose of 7 mg/day in adolescents which achieved efficacious concentrations was identified using model-based approaches.

Benjamin et al[14] and Momper et al[1] looked at the reasons for failed pediatric trials submitted to the FDA and why they failed to gain pediatric approval and labeling. Inappropriate dose ranging was one of the findings in both studies, suggesting that appropriate age-dosing intervals could be one variable that would lead to improved success in pediatric clinical trials.

In the sample of KD patients presented by Song et al,[7] the CAA+ group older than 5 years had a CRP of 17.1 $\pm$ 5.2, whereas the CAA+ group younger than 1 year had a CRP of 8.3 $\pm$ 5.3. Duration of total fever was significantly less in the patients younger than 1 year than in patients older than 5 years. In addition, the number of symptoms experienced by the older group was significantly greater than that in the younger group. This age breakdown was maintained for our study with the assumption that any priorities that were identified in our simulation might have a differential effect on the 2 age groups. However, other reports of clinical parameters in KD (aside from laboratory values) do not show a difference between patients that are infants versus children older than 5 years.[11] Age stratification in pediatric trials is common in oncology[15] and in most pediatric drug development trials. However, the use of pediatric CTS to understand the effect of age stratification on trial outcome is rarely reported in the literature. Using PK data, Krishna et al[16] reported CTS as a means of estimating the effect of pediatric age with the goal of determining the appropriate dose of fexofenadine (adult dose, 60 mg) in children 6-12 years, 2-5 years, and 6 months–2 years. The authors found that children aged 1-12 years weighing more than 10.5 kg should get a dose of 30 mg, and children 6 months and older and less than 10.5 kg should get a dose of 15 mg. This report represents the PK-PD type of simulation exercise mentioned above.[9,10] The second type of pediatric age stratification simulation is represented by the work described herein, and appears to be absent from the literature. Clinical trial simulation is very useful for testing hypotheses that may impact an upcoming trial, is used in drug development for adults, and could be used in pediatric drug development to decrease the high rate of pediatric study failures.

The prior disease information required for CTS for age stratification is a substantial challenge. When the number of pediatric patients in the youngest age group is very small, the response is often to extrapolate drug efficacy from the older pediatric population. Therefore, exploration of the disease process and of the response to therapy depends on a high degree of awareness that these processes may be different in other pediatric age strata.

This work has several limitations. The first is that, in the case of hypotheses 1 and 2, the patient data used for simulation are themselves simulated according to the distribution observed in the population described by Song et al.[7] Since the objective of this exercise is to postulate that age stratification makes a difference to trial outcome and not to achieve a clinical objective, we believe that this data set is adequate. The second limitation is that the simulations in hypotheses 1 and 2 are "exaggerated" versions of each example, so that, for example, in hypothesis 1, if one of the parameters with a large overlap in values between children younger than 1 and older than 5 years were chosen, the difference between age categories would be less pronounced. Likewise, in hypothesis 2, more subtle differences may have resulted from choosing a difference in absolute risk reduction of less than 5%. In summary, if assumptions regarding a pediatric disease process do not include age stratification of inclusion criteria or treatment response during drug development, then the wrong conclusion could result with respect to the safety and efficacy of a drug for different ages. CTS provides a valuable tool for hypothesis testing related to age-specific disease processes such as KD, and response to intervention and outcomes during drug development. Given the high rate of failure of pediatric drug development trials, CTS may provide a means for improving the success rate of pediatric studies.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Momper JD, Mulugeta Y, Burckart GJ. Failed pediatric drug development trials. *Clin Pharmacol Ther*. 2015;98:245-251.
2. Santen G, Zwet E, Danhof M, Pasqua OD. From trial and error to trial simulation. Part 1: the importance of model-based drug development for antidepressant drugs. *Clin Pharmacol Ther*. 2009;86:248-254.
3. Santen G, Horrigan J, Danhof M, Pasqua OD. From trial and error to trial simulation. Part 2: an appraisal of current beliefs in the design and analysis of clinical trials for antidepressant drugs. *Clin Pharmacol Ther*. 2009;86:255-262.
4. Mehta CR, Gao P. Population enrichment designs: case study of a large multinational trial. *J Biopharm Stat*. 2011;21:831-845.
5. Magaret A, Angus DC, Neill KJ, et al. Design of a multi-arm randomized clinical trial with no control arm. *Contemp Clin Trials*. 2016;46:12-17.
6. Zang Y, Guo B. Optimal two-stage enrichment design correcting for biomarker misclassification. *Stat Methods Med Res*. 2015:1-16.
7. Song D, Yeo Y, Ha KS, et al. Risk factors for Kawasaki disease-associated coronary abnormalities differ depending on age. *Eur J Pediatr*. 2009;168:1315-1321.
8. http://www.cdc.gov/nchs/data/nhsr/nhsr010.pdf. Accessed December 9, 2015.
9. Kearns GL, Abdel-Rahman SM, Alander SW, Blowey DL, Leeder JS, Kauffman RE. Developmental pharmacology—drug disposition, action, and therapy in infants and children. *N Engl J Med*. 2003;349:1157-1167.
10. Anderson BJ, Allegaert K, Holford NH. Population clinical pharmacology of children: general principles. *Eur J Pediatr*. 2006; 165:741-746.
11. Alleyn CR, Volkening LK, Wolfson J, Rodriguez-Ventura A, Wood JR, Laffel LMB. Occurrence of microalbuminuria in young people with type 1 diabetes: importance of age and diabetes duration. *Diabet Med*. 2010;27:532-537.
12. http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/PediatricAdvisoryCommittee/ucm116530.htm. Accessed November 3, 2015.
13. http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/DevelopmentResources/UCM446738.pdf, Ventolin HFA FDA Presentation, page 7. Accessed November 3, 2015.
14. Benjamin DK Jr, Smith PB, Jadhav P, et al. Pediatric antihypertensive trial failures: analysis of end points and dose range. *Hypertension*. 2008;51:834-840.
15. Frazier AL, Hale JP, Rodriguez-Galindo C, et al. Revised risk classification for pediatric extracranial germ cell tumors based on 25 years of clinical trial data from the United Kingdom and United States. *J Clin Oncol*. 2015;33:195-201.
16. Krishna K, Krishnaswami S, Kittner B, Sankoh AJ, Jensen BK. The utility of mixed-effects covariate analysis in rapid selection of doses in pediatric subjects: a case study with fexofenadine hydrochloride. *Biopharm Drug Dispos*. 2004;25:373-387.