Extent of safety database in pediatric drug development: types of assessment, analytical precision, and pathway for extrapolation through on-target effects

Margaret Gamalo\textsuperscript{a}, Yihua Zhao\textsuperscript{b}, Aijun Gao\textsuperscript{c}, Jingjing Ye\textsuperscript{d}, Junjing Lin\textsuperscript{e}, Eiji Eshida\textsuperscript{f}, YounJeong Choi\textsuperscript{g}, and Robert Nelson\textsuperscript{h}

\textsuperscript{a}Pfizer, Collegeville, PA, USA; \textsuperscript{b}Flatiron Health, New York, NY, USA; \textsuperscript{c}Alexion Pharmaceuticals, Boston, MA, USA; \textsuperscript{d}BeiGene, Cambridge, MA, USA; \textsuperscript{e}Takeda Pharmaceuticals, Cambridge, MA, USA; \textsuperscript{f}Sunovion, Great Falls, VA, USA; \textsuperscript{g}Roche/Genentech, San Francisco, CA, USA; \textsuperscript{h}Johnson & Johnson, Spring House, PA, USA

ABSTRACT
Pediatric patients should have access to medicines that have been appropriately evaluated for safety and efficacy through revised labelling. Given this goal, the adequacy of the pediatric clinical development plan and resulting safety database are critical factors to inform a favorable benefit-risk assessment for the intended use of the medicinal product. While extrapolation from adults can be used to support efficacy of drugs in children, there may be a reluctance to use the same approach in safety assessments, wiping out potential gains in trial efficiency through a reduction of sample size. To address this issue, we explore safety review in pediatric trials, including specific types of safety assessments and precision on the estimation of event rates for specific adverse events (AEs) that can be achieved. In addition, we discuss the assessments which can provide a benchmark for the use of extrapolation of safety that focuses on on-target effects. Finally, we explore a unified approach for understanding precision using Bayesian approaches as the most appropriate methodology to describe or ascertain risk in probabilistic terms for the estimate of the event rate of specific AEs.

1. Introduction
Pediatric patients should have access to medicines that have been appropriately evaluated for them. Legislations have been enacted in several jurisdictions toward this goal that encourages, incentivizes, and requires studies of investigational drugs in children when pediatric use is anticipated (Bucci-Rechtweg 2017). Pediatric research has been stimulated across the globe in the last three decades as a result, with Food and Drug Administration (FDA) making labeling changes in 854 drugs as of April 2020 (FDA 2021b). To reduce the exposure of children to unnecessary clinical trials, both the United States (U.S.) and European Union (EU) regulations suggest a strategy based on assessing the relevance of existing information in the adult population to the pediatric population, in terms of the similarity of disease, drug pharmacology, and clinical response to treatment to identify the gaps or level of uncertainty that need to be addressed to extend from adult to pediatric patients the conclusion of adequate evidence of efficacy and safety (Human Medicines Development European Medicines Agency and Evaluation 2017; ICH 2017). This strategy is called pediatric extrapolation and is adopted broadly by International Council for Harmonization (ICH) member parties to maximize the use of preexisting information and to reduce the amount of, or general need for, additional information to reach conclusions of effective and safe use of drugs in children (ICH 2022). The recent ICH E11A draft
guidance on the use of pediatric extrapolation explicitly discusses how this concept can be extended to the assessment of pediatric-specific safety for establishing positive benefit-risk of the drug in children.

As with any drug development program, the U.S. evidentiary standard (21 CFR 314.50) for approval of a medicinal product for pediatric patients is the same as in adults, i.e., the product needs to be safe and effective for labeled indications under labeled conditions of use. Identifying and assessing the safety reported in pediatric clinical trials is important in determining whether a product’s labeling needs to be revised (Rodriguez et al. 2008). In drug reviews, reviewers may explicitly note whether the observed treatment-related adverse events in children are reflected in the existing labeling (for previously approved products), or whether some revisions are needed. This observation is of particular importance because the disease pathophysiology, along with growth and developmental changes, can create unanticipated dose-response, adverse events (AE), or clinically significant laboratory changes in the pediatric compared to the adult patient (Kim et al. 2021; McMahon and Dal Pan 2018). It also calls for data needed to support such conclusions.

While pediatric trial sample size usually is driven by the objective of establishing efficacy, there is no clear benchmark to guide the sample size necessary for assessing safety. Unlike efficacy where a clear primary endpoint can be identified, the evaluation of safety is multi-objective where the common assessments are treatment emergent AEs (TEAEs), serious AEs (SAEs), severe AEs, and AEs leading to discontinuation of the participant from study. These adverse events may also be associated with many quantities or qualities of interest, e.g., incidence, timing, duration, reversibility, recurrence of AEs, the impact of dose (reduction), timing and frequency of treatment, effects of rescue treatments, among others. In addition, while the ICH E11A has made a recommendation for adequate safety database in adults for drugs taken chronically, this recommendation may not be feasible in children (see ICH 1995). Of note, consideration on feasibility cannot be ignored as Hwang et al. (Hwang et al. 2019) reported that only 33.8% of pediatric studies mandated by the Pediatric Research Equity Act have been completed after a median follow-up of 6.8 years, and 41.2% of drug labels include any pediatric data.

Because of the limited trial size in most pediatric clinical trials, there is a need to prioritize what objectives can be de-risked while keeping track of other risks and their attendant precision. In relation to the goal of labelling revision, it may be important to look at two areas: on-target effects and pediatric developmental or growth and maturation domains. For on-target effects, the recently drafted ICH E11A discussed extrapolation of safety data which can be considered based on the available knowledge of the known and/or potential safety issues in the reference population that are relevant to the target pediatric population. These focused extrapolation of safety data can help increase certainty about the expected safety profile of a drug in a particular pediatric population and provide guidance if additional gaps in knowledge need to be addressed in the pediatric program.

In this paper, we explore the extent of the safety database in pediatric trials. We explore factors affecting this database, specific safety assessments in children, and what assessments may be useful for driving the extent of development. We explore precision, the proper hypothesis that needs to be tested, and how large the safety databases need to be for the conclusion to be reliable. We explore the use of Bayesian approaches as the most appropriate methodology to describe/ascertain risk in probabilistic terms with available data. We provide unified expressions for the assessment of key quantities of interest such as confidence level, incidence rate, and observed data. Finally, we explore the potential for the use of extrapolation for specific types of adverse events, e.g., on-target effects, where these are expected in any population.

2. Specific types of pediatric safety assessments

Adverse events may be classified as belonging to a certain type according to their reaction profiles: augmentation of normal drug effects (Type A), bizarre effects (Type B), chronic effects (Type C), delayed effects (Type D), and end of drug use effects (Type E) (Elzagallaaei et al. 2017). Augmentation of normal drug effects is often characterized by exaggerated and adverse pharmacologic effects when a drug interacts with its intended receptor at the intended tissue of interest (“on-target effects”). On-
target effects may also occur when a drug interacts with its intended receptor, but in an unintended tissue (Golan et al. 2011). Often on-target effects are identified at compound target validation which includes various queries to identify and define potential roles that the target could play in normal physiologic processes or homeostasis leading to early hypothesis testing for target-related safety risks (Rudmann 2013). It is through on-target effects that it is determined whether the potential on-target safety liability fits the indication (benefit/risk), i.e., right safety for the right patient (Morgan et al. 2018). Because of this principle, on-target effects are predictable, tend to be mild, can resolve with dose reduction, and constitutes about 80% of all observed AEs (Elzagallaa et al. 2017). Most on-target effects are considered as adverse drug reactions (ADRs) an undesirable effect, reasonably associated with the use of a drug, that may occur as part of the pharmacological action of the drug or may be unpredictable in its occurrence (FDA 1994).

The ability to estimate and provide adequate precision for Type B or idiosyncratic/bizarre adverse events is challenging as they often occur less frequently. Many of these bizarre adverse events are “off-target” effects. The number of these so-called “off-target” effects for a small molecule is probably high because profiling of compounds uses a predefined (restricted) non-overlapping target panel consisting of a limited set of targets (typically < 100) (Lynch et al. 2017). In addition, compounds that bind to numerous, unintended targets are associated with greater toxicity. While off-target interactions are generally weaker in affinity than those with the intended pharmacological target, they may be relevant in cases of higher cellular expression of the off-target or high systemic exposure, such as in preclinical toxicity studies (where higher doses are interrogated to define the toxic profile of the compound), clinical mis-dosings, accidental or intentional overdose, drug-drug interactions (which may lead to higher systemic exposures), or other unanticipated individual variations (which also can lead to higher systemic exposures) (Whitebread et al. 2016). These events are often unpredictable, have no clear dose-dependency, and account for about less than 20% of all adverse events observed (Elzagallaa et al. 2017).

We do not always fully understand the extent of the mechanism of toxicity and safety; however, a battery of in vitro pharmacology assays is usually tested to determine off-target interactions (Lynch et al. 2017). It is proposed that the extent of safety database be driven by the ability to characterize known on-target effects in children and determine whether there is high likelihood that they are similar to adults. In particular, the rates in children should not negatively change the benefit-risk calculus in the intended population, e.g., an increase of more than 3 times the rate in adults (see Figure 1). The use of on-target effects also makes it convenient to improve the precision of estimates since there is a possibility to leverage other related populations. Furthermore, the assessment of these event rates generally comes from well-controlled trials enabling causal estimation of effects. The proposal does not mean, however, that the adverse drug reaction is the only assessment that will be conducted and that assessments for other types of AEs must still be part of the safety assessment routine. As a rule, and like the safety assessments conducted in the adult population, the strategy in children should still be to find ways to minimize or find mitigation for the pharmacological engagement of unintended molecular target(s); and/or other non-pharmacological toxicity (e.g., membrane damage).

Alternatively, but of equal importance, the extent of the safety database may be related to ensuring that safety risks related to pediatric development are rigorously assessed. Depending on the age of the target population, the mechanism of action of the drug and the objective of the study, data on developmental parameters, e.g., growth, reproductive and neurocognitive development, need to be collected. Abnormal findings in toxicology studies can guide in ensuring proper assessments are planned and captured in clinical trials.

3. Extent of safety database through precision in assessments types

To explore extent of database, i.e., sample size of patients with a given exposure, for given precision, suppose \( y_{ge} \) as the number of events out of \( n_{ge} \) patients in the safety analysis set for each arm
Figure 1. Typical comparison in safety assessment for adverse drug reactions in children. In the first figure, placebo-corrected response in adolescents is compared with adults. In the second figure, incidence rates in adolescents are compared to adults in the absence of placebo in the adolescent trial. However, contextualization is usually used with background rates available in literature or other sources.
3.1. Precision for common identified/potential risk regardless of importance

The basis for inclusion in the Common Adverse Reaction section of the United States Package Insert (USPI) is that the AE occurs > 1% in the treated group and for which the rate for a drug exceeds the rate for placebo (FDA 1994), i.e., the placebo rate must be at least 1%. Adverse drug reactions are generally known/identified or potentially related to the mechanism of action, and they are continually monitored in the development. An identified risk is an untoward occurrence for which there is adequate evidence of an association with the medicinal product of interest (FDA 1994). Many ADRs are also on-target effects and in reviews of marketing authorizations for a drug in a pediatric indication, there generally is a mention of a comparison of ADRs between adults and target pediatric cohort. Others note differences and sometimes do not mention of labeling changes at all. In these assessments, the hypothesis of interest is similarity or that there is no clinically relevant difference in the incidence of a specific ADR or on-target effect in children. Suppose \( g \in \{1, 2\} \), where \( g = 1 \) refers to the reference population and \( g = 2 \) is the target population. Then assessing consistency of the incidence rates or an “equivalence testing” is more suitable, i.e.,

\[
H_{0,k} : \theta_{2,k} \geq \hat{\theta}_{1,k} + \varepsilon \quad \text{vs} \quad H_{1,k} : \theta_{2,k} < \hat{\theta}_{1,k} + \varepsilon,
\]

(1)

where \( \theta_{2,k} = \theta_{2,ITx,k} - \theta_{2,Cx,k} \), for \( 0 < \varepsilon \leq \hat{\theta}_{1,k} - z_{\alpha} \text{var}(\hat{\theta}_{1,k}), \theta_{1,k} \) is the effect of the investigational drug in the reference population, and \( k \) indexes the ADRs. In other instances, consistency is defined as

\[
H_{0,k} : \theta_{2,k} \geq f \hat{\theta}_{1,k} \quad \text{vs} \quad H_{1,k} : \theta_{2,k} < f \hat{\theta}_{1,k}
\]

(2)

where \( f > 1 \) corresponding to a multiple of the rate observed in the reference population. Of note, because \( \theta_{2,k} \) is small, the second hypothesis (2) may be more appropriate as it may be challenging to establish a margin about a small rate.

For ease of exposition and notation, we will drop the index \( k \). As defined above, the risk difference is \( \theta_{2} = \theta_{2,ITx} - \theta_{2,Cx} \). If the prior for \( \theta_{2,x} \) is \( \pi_{a_{2}, z} = \text{Beta}(a_{2}', b_{2}') \) then the posterior distribution for \( \theta_{2,x} \) given observed data \( (r_{2,x}, n_{2,x}) \) is \( q_{\theta_{2,z}} = \text{Beta}(r_{2,x} + a_{2}', n_{2,x} - r_{2,x} + b_{2}') \). This observed data could be any available information on the ADR before conducting the pediatric trial. Let \( a_{z} = r_{2,z} + a_{2}', b_{z} = n_{2,z} - r_{2,z} + b_{2}' \) so that the posterior distribution of the risk difference \( p_{\theta_{2}}(x) \) can be expressed using parameters \( a_{z}, b_{z} \) which if \( z \in \{ITx=1, Cx=2\} \) has the form

\[
p_{\theta_{2}}(x) = \begin{cases} 
A^{-1}B(a_{2} + b_{1}) x^{a_{2}+b_{1}-1} (1-x)^{a_{2}+b_{1}-1} 
\quad \times \mathcal{F}_{1}(b_{1}, a_{1} + a_{2} + b_{1} + b_{2} - 2, 1 - a_{1}; b_{1} + a_{2};
\quad 1 - x, 1 - x^2) 
& \text{if } 0 \leq x < 1 \\
A^{-1}B(a_{1} + a_{2} - 1, b_{1} + b_{2} - 1) 
& \text{if } x = 0, a_{1} + a_{2} > 1, b_{1} + b_{2} > 1 \\
A^{-1}B(a_{1} + a_{2} - 1, b_{1} + b_{2} - 1) 
\quad \times \mathcal{F}_{1}(b_{2}, 1 - a_{2}, a_{1} + a_{2} + b_{1} + b_{2} - 2; a_{1} + b_{2};
\quad 1 - x^2, 1 + x) 
& \text{if } -1 < x \leq 0 
\end{cases}
\]

(3)

defined on the interval \((-1,1)\) (see Pham-Gia et al. 1993)). Here \( A = \text{Beta}(a_{1}, b_{1}) \text{Beta}(a_{2}, b_{2}) \) and \( \mathcal{F}_{1} (\cdot) \) is Appell’s first hypergeometric function computed as
\begin{equation}
F_1(u, v_1, v_2; w; x_1, x_2) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{(u)^{i+j}(v_1)^{i}(v_2)^{j} x_1^i x_2^j}{(w)^{i+j} i! j!}
\end{equation}

with Pochhammer symbol \((r)^s = \Gamma(r + s)/\Gamma(r)\) (see (Pham-Gia and Turkkan 1998)). The double series representation is convenient for finding the confidence level or probability \(C = \text{Pr}(\vartheta_2 < f\hat{\theta}_1) = \int_{-1}^{f\hat{\theta}_1} p_{\vartheta_2}(x)dx\), for the estimate of the specific adverse event in children not exceeding a certain multiple of what the risk difference between treatments \(\hat{\theta}_1\) was observed in adults. This probability can be split into the following depending on where \(f\hat{\theta}_1\) is in the \([-1, 1]\) interval, i.e.,

\begin{equation}
C = \text{Pr}(\vartheta_2 < f\hat{\theta}_1) = 1 - \int_{f\hat{\theta}_1}^{1} A^{-1} B(a_2, b_1) x^{b_1+b_2-1}(1-x)^{a_1+b_1-1} \times F_1(b_1, a_1 + a_2 + b_1 + b_2 - 2, 1 - a_1; b_1 + a_2; 1 - x, 1 - x^2)dx
\end{equation}

for 0 \leq f\hat{\theta}_1 < 1 and

\begin{equation}
C = \text{Pr}(\vartheta_2 < f\hat{\theta}_1) = \int_{-1}^{0} A^{-1} B(a_1, b_2) (-x)^{b_1+b_2-1}(1+x)^{a_1+b_1-1} \times F_1(b_2, 1 - a_2, a_1 + a_2 + b_1 + b_2 - 2; a_1 + b_2; 1 - x^2, 1 + x)dx
\end{equation}

for \(-1 \leq f\hat{\theta}_1 < 0\). The relationship of the quantities of interest \(C, \hat{\theta}_1, f, (a_2, b_2)\) can be characterized from equations (5)-(6). These integrals can then be approximated through the trapezoidal rule or Monte Carlo methods to avoid numerical tolerance level issues. Alternatively, asymptotic normal approximations on each of the posterior distributions of \(\vartheta_2, x\) can also be used if there is difficulty in computing the above expressions. A very low probability would indicate that the estimates of the placebo-corrected event rate for a specific adverse event in a pediatric cohort is not consistent with what was observed in its respective reference population. Of note, the case of single arm trials is discussed in the subsequent section. These probabilistic assessments provide information on the degree of current uncertainty and how information and extent of post-marketing surveillance can be guided appropriately to increase precision of estimates based on current information.

Figure 2 shows the chance of ruling out clinically meaningful increase in the placebo-corrected AE rate \(\varepsilon\), expressed as fold increase over reference placebo-corrected rate, at different placebo incidence rates for a common ADR in the reference population according to hypothesis (1). This figure is created by setting \(\hat{\theta}_{1,Cx} = 0.01\) or 0.03 and \(\hat{\theta}_{1,Tx} = f\hat{\theta}_{1,Cx}\) so that \(\hat{\theta}_1 = \hat{\theta}_{1,Tx} - \hat{\theta}_{1,Cx} = (f-1)\hat{\theta}_{1,Cx}\). On the other hand, \(\hat{\theta}_{2,Tx} = \hat{\theta}_1(1+m)\) whereas \(\hat{\theta}_{2,Cx} = \hat{\theta}_{1,Cx}\) so that \(\hat{\theta}_2 = \hat{\theta}_{2,Tx} - \hat{\theta}_{2,Cx} = \hat{\theta}_1(1+m) - \hat{\theta}_{1,Cx} = ((f-1)(1+m) - 1)\hat{\theta}_{1,Cx}\). Furthermore, we set \(f = 2\)

\begin{figure}[h]
\centering
\begin{subfigure}{0.45\textwidth}
\centering
\includegraphics[width=\textwidth]{a.png}
\caption{Reference: Pbo = 0.01, iTx = 0.02}
\end{subfigure}\hspace{1cm}
\begin{subfigure}{0.45\textwidth}
\centering
\includegraphics[width=\textwidth]{b.png}
\caption{Reference: Pbo = 0.03, iTx = 0.06}
\end{subfigure}
\caption{Probability of consistent placebo (Pbo)-corrected incidence proportion in the target children’s cohort relative to reference cohort.}
\end{figure}
and $m = [-1, 1]$ so that the confidence level $C$ can be calculated for different values of $m$ (consequently, upper limit of $\vartheta_2$) and sample size in the target population. Given this setup, suppose that a specific common ADR for the investigational drug was observed to have a difference in incidence proportion of 1% in the pediatric population when the background placebo rate is 1%, then to rule out that there is an $\epsilon = 0.015$ increase over the placebo-corrected rate in the reference population with a probability of 80%, a sample size of at least approximately 180 pediatric participants randomized 1:1 is needed between the treatment and placebo control. As seen in Figure 2, when the extent of the safety database is small, there is a low probability that a meaningful absolute increase in the placebo-corrected incidence proportion can be ruled out. In fact, to have sufficient confidence that there is no increase by a certain margin, the placebo-corrected rate in the pediatric cohort has to be lower than the placebo-corrected incidence rate in the reference population. Figure 2 also shows that if the background placebo rate in the reference population gets higher, the sample size requirement in the target population also gets higher. Hence, if there is some evidence that a common ADR has an increased incidence, the probability assessments show that it may be important to ensure that the extent of the safety database is increased proportionally to the level of confidence we can be assured of the precision of the estimate.

Figure 3 shows the chance of ruling out different multiples of placebo-corrected incidence rates at different placebo reference incidence rates of a common ADR in the reference population per hypothesis (2). For example, suppose that a specific adverse reaction was observed to have a placebo-corrected incidence proportion of 1% in the pediatric population when the reference placebo rate is 1%, then to rule out that there is no doubling of that placebo-corrected rate in the target population with a probability of 80%, a sample size of more than 200 pediatric participants randomized 1:1 is needed between the investigational treatment and placebo control (see Figure 3(b)). When the background rate increases, all the more that a higher sample size is needed to rule out even a small fold increase (see Figure 3(a)-(b)). Hence, if there does not seem to be a difference in adults and children, the rate in children can be made more precise by using the adult information as prior and potentially reducing the extent of the safety database needed. Of note, if testing is expanded to include any adverse event, any noticeable increase in incidence relative to placebo and over adults will need to be evaluated whether it rises to the level of an important identified risk. Poisson regression can also be used if the exposure time is used for adjustment.

Since there will be many important identified or potential risks, one may also focus only on the medically most serious, i.e., important identified or potential risks can be ranked according to medical seriousness or ranked according to severity and disease burden (see for e.g. Seifu (Seifu 2023)). These can then be analyzed through prioritized outcomes (see Finkelstein and Schoenfeld, 1999), Pocock et al. (Pocock et al. 2012)). Pocock et al. (Pocock et al. 2012) introduced the concept of the win ratio (WR) in the analysis of composite endpoints based on clinical priorities. Pairs of subjects are compared based on prioritized component outcomes. The win ratio statistic is then calculated as the number of “winners” divided by the number of “losers” in the test group. To make this operational

Figure 3. Probability of ruling out an $f$-fold increase in the difference in incidence proportion relative to placebo.
in terms of assessing similarity in the context of the assessment of safety, the win odds (WO) can be used instead where ties are handled by counting them as a half win to each subject within a pair (Dong et al. 2019). In particular, if $\Psi$ is the WO parameter. The non-inferiority (NI) hypothesis for $\Psi$ is written as $H_0: \Psi \leq \delta$ vs $H_1: \Psi > \delta$, where $\delta$ is NI margin with $0 < \delta < 1$ (Peng 2020). For example, when $\theta = 0.8$, the null hypothesis is that the probability of winning for a subject in the test group is less than or equal to 80% of that for a subject in the control group, with ties splitting evenly. While this technique amplifies events to increase statistical power, it will not give probabilistic statement as it does not involve a posterior distribution but a testing procedure whether there is sufficient evidence for the dissimilarity between the rate of the specific adverse event of interest or not. It can give an estimate of the WO and associated confidence interval.

### 3.2. Precision for uncommon important identified/potential risks

Less common important identified/potential risks or adverse events of interest are usually observed in the context of long-term extension trials. In these trials, a placebo may no longer be present as exposing a child to no treatment for a long time may not be ethical or will render the trial not competitive and infeasible. In this case, control is no longer available and analysis usually involves characterizing incidence proportions and exposure time. Poisson regression or Cox regression are used when exposure to the drug is incorporated. In labeling, however, exposure time is not usually included and only crude incidences are stated and so we will focus the discussion on incidence proportions and its associated posterior distribution. The discussion on assessment of precision in rates using Poisson regression or hazard ratios which incorporates exposures will be a future discussion and beyond the scope of this manuscript.

Because a small increment in the incidence proportion cannot be associated with deleterious effects absent a control, the hypothesis being tested here is ruling out multiples of incidence proportion in the reference population, i.e., consistency defined as

$$H_{0,k}: \theta_{n,k} \geq \hat{f}_{1,k} \quad \text{vs} \quad H_{1,k}: \theta_{n,k} < \hat{f}_{1,k}$$

where $\hat{f}_{1,k}$ is the estimate of the incidence proportion of the investigational drug in the reference population, and $f > 1$ corresponding to a multiple of the rate observed in the reference population. In addition, it is also of interest whether there were unique safety risks identified in children and that if none are seen, what is the associated confidence level. Suppose $y_2$ as the number of events out of $n_2$ patients in the safety analysis set for the investigational treatment arm, $g \in \{1, 2\}$. Let $\theta_2$ be the parameter of interest corresponding to the incidence proportion of the specific event for the target population. If the prior for $\theta_2$ is $\pi_{\theta_2} = Beta(a', b')$ then the posterior distribution for $\theta_2$ given observed data $(r_2, n_2)$ is $q_{\theta_2} = Beta(r_2 + a', n_2 - r_2 + b')$. Then quantities of interest $C, \theta_1, f,$ and $r_2$ can be completely characterized from

$$C = Pr(\theta_2 < \hat{f}_{1}|n_2, r_2) = \frac{1}{B(r_2 + a', n_2 - r_2 + b')} \int_{0}^{\hat{f}_{1}} x^{n_2 + a' - 1}(1 - x)^{n_2 - n_2 + b' - 1} dx$$

which is related to the problem explored by Liu (2016). Because these events are infrequent, it is desirable to use a “near-zero informative” prior

$$\pi_{\theta_2} = Beta(a \approx p_a^2/(1 - p_a^2), 1).$$

When $0 < p_a^2 < 0.5$, the prior distribution has a mode close to zero if $p_a^2$ is close to zero. The distribution is then weighed more toward zero. When $p_a^2 = 0.5$, the prior follows a uniform distribution. When $0.5 < p_a^2 < 1$, the prior weighs more toward one. Usually, the value of $p_a$ can be based on external information from a similar population, e.g., $p_a^2 \in [0.001, 0.1]$. Of note, this prior can also be applied to the case of common adverse reactions in Section 3.1 for adverse drug reactions.
If C were fixed at 0.80, then we will be able to calculate \( f \) given a background iTx rate \( \theta_i \) in the reference population as a function of \((r_2, n_2)\). We will call this \( f_C \) as the minimum fold based on probability threshold C. Figure 4 shows the fold increase in the specific AE rate when the background iTx rate for that AE is fixed at 0.01 at different confidence levels or probability thresholds C, i.e., \( f_C \) as a function of \( n_2 \) for each possible observed \( r_2 \). For example, when the probability threshold is 80%, if no event is observed for that specific AE, then we can rule out a doubling of the rate for that AE when the sample size is about 150 participants. Of note, such a sample size is still relatively large for many pediatric trials. Once the probability threshold is raised, then the doubling of the rate can no longer be ruled out. Furthermore, when the observed frequency of events increases, the more difficult it is to rule out a doubling of the incidence proportion even with a total sample size of 300 patients. In Figure 4, it can be inferred that if the true incidence rate is small, the benefit gained in increasing sample size is not quite meaningful. On the other hand, if the true incidence rate is high, there is benefit in increasing sample size as the information (probability) gain for ascertaining \( f \)-fold increase is steep.

Ignoring \( f \), Figure 5 shows the probability contours at different sample sizes and observed frequency of specific AEs, i.e., the figure shows the confidence level C as a function of the plane created by \( r_2 \) and \( n_2 \). This then gives the likelihood of observing a particular frequency at a given sample size when the true incidence rate proportion is specified. At any of the sample sizes explored, and when the true incidence rate of the AE is small (less than 5/1000), the ability to observe at least 1 event is at most about 40%. This maximum probability happens when the sample size is 300 patients. When the true incidence rate is 1/100, the probability of observing at least 1 event is 60%. Figure 5 shows that when the true incidence is small (< 1/100) which is the upper bound for uncommon AEs the probability of observing at least one event in a 300 participant trial is low. Conversely, because these events are supposedly uncommon, observing it gives a hint that the true rate may be higher or lends credibility to the fact that they are drug effects rather than chance findings. In this example, we can recognize the importance of the Bayesian methodology for borrowing information from other populations or cohorts to increase the precision in the estimation of the incidence proportion of specific uncommon adverse events that are identified or are recognized as potential risks.

Figure 4. Incidence rate cutoff per observed incidence frequency as a function of sample size using different probability thresholds. The limits of the y-axis are deliberately modified to accommodate the extent of the curves.
3.3. **Precision for developmental safety assessments**

The unique feature of assessing safety in children is in the consideration of developmental safety where there is relativity in pediatric growth data analysis. All age subgroups of the pediatric population grow from the moment of birth until a certain time point during adolescence (or beyond adolescence for certain individuals) and each child grows at their own pace. Investigators should plan to evaluate the pediatric subjects’ growth in terms of anthropometric measurements. Usually, this includes the collection of data on sex, chronologic age (CA), bone age (BA), body weight, height, and head circumference and the extent of assessments may be influenced by the nature of the disease and the objective of a trial (Choi et al. 2022). In relation to these measurements, there is a need for outcomes that provide a sensitive measure of change since the trial cannot run for a long time and because growth happens slowly over time.

Health care professionals sometimes convert height, weight, or body mass index (BMI) measurements to standard deviation scores (SDS) and permit the understanding whether a measurement is normal by comparing it with the normal range of measurements for other children of the same age and sex (Cole et al. 1998; Inokuchi et al. 2011). A lower standard deviation score (in absolute value) means that the measurement is closer to the average or mean, while a high standard deviation score means that the value is further from the average or mean. A negative SDS indicates that the value is below the average or mean, and a positive value means it is above the average or mean. The percentile number (or centile) shows how a child’s height compares to children of the same age and sex.

For ascertaining the effect of the drug on growth measurements, the change from baseline SDS, i.e., \( \Delta SDS_{t_n} = SDS_{t_n} - SDS_{t_0} \), is determined over time, \( t \in T = \{ t_0, \ldots, t_n \} \) and the expectation is that \( E[\Delta SDS_{t_n}] = 0 \). Typically, no formal hypothesis testing is done, and conclusions are formed based on ocular inspection. The relevant clinical context is also important as the drug, known to reverse disease progression, may have beneficial impact on growth. Even if a hypothesis test is done, only the deviation at a landmark timepoint is considered \( H_0 : \Delta SDS_{t_n} = 0 \) vs. \( H_1 : \Delta SDS_{t_n} < 0 \), i.e., testing procedure starts with normality and in the process wants to rule out abnormality at \( t_n \). However, not rejecting this hypothesis may just imply that the sample size is small. Hence, it may be more
relevant to establish non-inferiority/consistency relative to placebo or context/benchmark for a development outcome, i.e., the realistic hypothesis is

\[ H_{0, t_n} : \Delta SDS_{t_n} \geq -\tau \quad \text{vs.} \quad H_{1, t_n} : \Delta SDS_{t_n} < -\tau \quad (10) \]

where \( \tau \) would correspond to a meaningful threshold. For example, suppose that the data in children pertaining to \( \Delta SDS \) are \( D = (x_1, \ldots, x_n) \) with \( x_i \sim \mathcal{N}(\mu, \sigma^2) \). Let the prior \( \pi(\mu, \sigma^2) = 1/\sigma^2 \), then the posterior distribution for the mean change in SDS is the normal-inverse-\( \chi^2 \) \( q(\mu, \sigma^2|D) = \mathcal{N}_{\chi^2}(\bar{x}, n, n-1, s^2) \) where \( \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \) and \( s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \). Integrating out the variance \( \sigma^2 \) from the posterior distribution \( q(\mu, \sigma^2|D) \), the assessment of the confidence level for a threshold \( -\tau \) corresponds to the cumulative distribution of a non-central \( t \) distribution with \( n-1 \) degrees of freedom, \( t_{n-1}(\mu|\bar{x}, s^2/n) \). Because the non-central \( t \) distribution is asymmetric and the mean is to the left of the median when \( \bar{x} < 0 \), the only way to rule out a shift of \( \tau = 0.5 \) with confidence level of at least 80% when the size of the sample is 50 is if the observed mean change SDS is at least \( \bar{x} \approx -0.04 \). Furthermore, even if one increases the sample size to 100 the observed mean change SDS does not change appreciably to rule out a shift of 0.5 with confidence level of 0.80. In other words, from a statistical standpoint, it is challenging to rule out deleterious shifts if we do not observe a shift in the mean.

More realistically, however, the hypothesis test that corresponds to the meaningful clinical question is whether the maximum of the average absolute change in SDS across time deviates from 0 as it is not known when the effect on development happens, i.e., if \( \Delta SDS = \max_{t \in T} \Delta SDS_{t_n} \), we are interested in showing whether

\[ H_0 : \tilde{\Delta} SDS \geq -\tilde{\tau} \quad \text{vs.} \quad H_1 : \tilde{\Delta} SDS < -\tilde{\tau}. \quad (11) \]

It is known that for \( x_1, \ldots, x_n \) that are independent and identically normally distributed with \( \mathcal{N}(\mu, \sigma^2) \), the probability density function is \( f(x; \mu, \sigma) = n\sigma^{-1}\phi((x-\mu)/\sigma)\Phi((x-\mu)/\sigma)^{n-1}. \) Suppose this assumption of independence is reasonable over time, the confidence level can be written similar to (8) when specification of the prior distributions for \( \mu \) and \( \sigma^2 \) and calculation of the posterior distribution. This will be explored more in subsequent research. In addition, it may be good to look at the increase in the variance and whether there is dose dependency in the disruption of growth patterns.

Because it may be challenging to create a meaningful threshold for a deleterious shift in mean of SDS and because while the means may not change, the drug can cause shifts without increasing the mean or variances, i.e., the effect of the drug on the growth can be quite different for everyone. In this case, it may be informative to examine shifts of SDS over time in patients. Intervals at baseline may be defined into eight SD groups as described in Tanaka et al. (Tanaka et al. 2021) so that the SDS \( z_{i0} = j \) if \( \delta_{j-1} < y_{i0} < \delta_{j} \), \( j = 1, \ldots, 10, i = 1, \ldots, n \), and \( \{\delta_1, \delta_2, \ldots, \delta_{10}\} = \{-\infty, -3.5, -2.5, -1.5, -0.5, 0.5, 1.5, 2.5, 3.5, \infty\} \).

Subsequent evaluation of change in the height SDS from baseline to landmark times of assessment can be grouped into the following five SD groups, i.e., \( d_{ik} = z_{ik} - z_{0} = j \) if \( \eta_{j-1} < y_{ik} < \eta_{j} \), \( j = 1, \ldots, 6 \) and \( i = 1, \ldots, n \), and \( \{\eta_1, \eta_2, \ldots, \eta_{6}\} = \{-\infty, -1.5, -0.5, 0.5, 1.5, \infty\} \). From here, a Bayesian probit regression can be used for estimation and inference.

Assessments for other measures, e.g., Tanner staging describing secondary sex characteristics for assessing maturational age, Bayley III Scales of Infant and Toddler Development to measure motor, cognitive, language, social-emotional, and adaptive behavior development in babies and young children for assessing normal neonatal and infant neurocognitive development (Choi et al. 2022) can be analyzed through usual analytic strategy for dichotomous or polytomous endpoints. Nevertheless, a meaningful hypothesis or scientific question is necessary to facilitate the proper interpretation of the results.
4. Exploiting archetypes of pediatric clinical development for informing safety

In terms of efficacy, the extent of pediatric clinical development is also influenced by whether extrapolation of efficacy can be supported or whether the “line of reasoning” has been established in the population intended for its use. Establishing the line of reasoning can be related to the current treatment landscape, referred henceforth as archetypes, for which the product belongs to at the time of pediatric study planning as proposed by Gamalo et al. (Gamalo et al. 2022). These archetypes are described as follows:

(A) First-in-class or first-in-indication in adults: In this archetype, there are no approved drugs in adults. However, there may be Phase 2 studies in adults establishing proof of direct benefit for the drugs in the class. There are no studies for the class of drugs in children.
(B) Established class where adult trials have been conducted: In this archetype, there are approved drugs in the same class in adults. There are no approved drugs in the class for children.
(C) Established class where adult and pediatric trials have been conducted: In this archetype, there are approved drugs in the same class for adults and children.

Table 1 gives some examples of drugs that belong in these archetypes at the time their clinical development is planned or executed. For example, finerenone is an investigational, non-steroidal, selective mineralocorticoid receptor antagonist (MRA) that has been shown to reduce many of the harmful effects of mineralocorticoid receptors (MR) overactivation. This is a new mechanism of action for the treatment of type II diabetes and therefore belongs to Archetype A. In the migraine indication, when rimegepant, oral calcitonin gene-related peptide (CGRP) antagonist, started pediatric clinical development, another drug of the same mechanism of action has been approved in adults previously. Hence, it belongs to Archetype B. In both Archetype A and B, the line of reasoning in children for that mechanism of action has not yet been established. The only difference is that proof of direct benefit in Archetype B is broadly supported since it has also been shown effective in other drugs in the same class. For these archetypes, the requirement for demonstration of efficacy in a robust manner for children is needed and that the extent of safety database may also be large.

Drugs being developed for children where only PK/PD or dose ranging study and a safety study is required belongs to Archetype C as in the case of many anti-convulsants as treatment for partial onset seizures. To gain precision for the estimation of adverse reactions in compounds under this archetype can use near-informative priors as the mechanism of action is responsible for many adverse reactions. For example, brivaracetam, a chemical analog of levetiracetam, was approved on August 27, 2021 for ages 1 month and older. In the development of brivaracetam, the extent of the safety database in children was from two single-arm open label studies: N01263 and N01266 with \( n = 346 \) to cover the expanded age indication for the oral formulation while the review of study EP0065 \( (n = 50) \) will examine all patients for coverage of the expanded indication for the intravenous formulation, 1 month to \(<16 \text{ years} \) (FDA 2021a). Levetiracetam, on the other hand, was approved for the same indication in children aged 4 to 16 on June 21, 2005, with a double-blind randomized placebo-control trial in \( n = 198 \) (FDA 2005a) and for ages 1 month to 4 years on December 16, 2011 forming a line of reasoning for extrapolation in this disease. The development in support of the approval for ages 1 month to 4 years includes 4 studies, including a double-blind randomized placebo-control trial in \( n = 116 \), two open

| Indication          | Archetype A | Archetype B | Archetype C |
|---------------------|-------------|-------------|-------------|
| Migraine            | perampanel  | rimegepant  | lasmiditan  |
| Partial Onset Seizures | finerenone  |             | brivaracetam|
| Type II Diabetes    |             |             | lixisenatide|
| Atopic dermatitis   | upadacitinib| lebrikizumab|             |
| RR-MS               | evobrutinib | ocrelizumab  | siponimod   |
label long-term safety studies $n = 255$ for 1 year and $n = 223$ for up to 7.5 years (FDA 2008). In fact, one can argue that for drugs in Archetype C, the extent of safety database is smaller unless a broader indication is sought after or that a certain unique signal is seen that may be of importance in children. Furthermore, the requirement for a double-blind randomized placebo-control trial is absent.

For more broader indications (even if related) were no other drugs of similar mechanism have been studied or when a safety concern exists, the extent of clinical development may be larger to support both efficacy and safety in these groups of patients. For example, in the agreed Pediatric Investigational Plan for eslicarbazepine acetate, a voltage-gated sodium channel (VGSC) blocker chemically related to carbamazepine and oxcarbazepine, the extent of development in children included 8 double-blind RCTs and open label RCTs (EMA 2019). Similarly, developments tend to take remarkably similar considerable extent, if the line of reasoning has not been fully established, e.g., conduct of adequately powered randomized controlled trials treatment and an open label extension in trials for the prevention of migraine headaches – a disease where the line of reasoning has not been established in children. For example, the extent of development for lasmiditan and rimegepant seems similar, even though lasmiditan belongs to Archetype C. In this case, the development becomes like drugs falling in Archetype A which needs a PK + randomized controlled efficacy and/or safety study. There are cases for drugs in Archetype C where even if the line of reasoning has been established in children, and the development could be smaller, the actual extent of safety database still follows the traditional paradigm, e.g., ofatumumab for relapsing remitting multiple sclerosis.

Different geographies have diverging views on being able to leverage existing safety from other relevant populations and indications. In the EU, the extrapolation guidance allows extrapolation of short-term safety data from adults to children (Human Medicines Development European Medicines Agency and Evaluation 2017). In this case, Archetype C provides further opportunity in leveraging existing information on on-target effects. In the US FDA, on the other hand, there is an understanding that safety cannot be extrapolated as adverse effects can be different in children than from adults (FDA 2005b). Hence, even if there is a possibility to extrapolate, the extent of safety database will drive developments. Arguably, however, the adoption of the proposed landscape, e.g., Archetype C, implicitly allows for some extrapolation of safety from other populations, e.g., adults, and the realization that development, particularly may not be competitive once similar drugs are in the market. The use of these archetypes together with the concept of on-target effects could provide a convincing argument for extrapolation of safety and in the use of quantitative measures to increase precision of estimates of incidence rates both for common ADRs and for identified or potential risks that are less common as the use of informative priors can be further facilitated.

5. Discussion

One objective of the FDA’s evaluation of safety in pediatric studies is to determine whether a product’s labeling needs to be revised to communicate to prescribers the benefit-risk information in children. In the reviews that have been done for investigational drugs for pediatric indications, most of the conclusions are based mainly on descriptive analysis and no inference is involved. Some may conclude that the safety profile for children differs from that for adults in ways that, at a minimum, warrant discussion in the product’s labeling. In others, reviewers explicitly note whether the findings of treatment-related adverse events in children are reflected in the existing labeling (particularly for previously approved products) or whether some revisions are needed (see for e.g. Field, Boat, et al. 2012). There are also times where the reviews state only that no unexpected adverse events had been noted. Such statement probably can be interpreted as suggesting that the safety profile was similar to that for adults if the product had been previously studied in adults (Boat, Field, et al. 2012). Other reviews may indicate unique safety concerns on musculoskeletal, neurocognitive, and maturation as indicated in relevant toxicology studies and/or outcomes in clinical studies.

Based on these observations, it is fair to say that almost all pediatric clinical trials are not designed to test specific hypotheses about safety nor to measure or identify adverse reactions
with any pre-specified level of sensitivity. But often a requirement is imposed on the extent of the safety database without sufficient scientific rationale on the specific hypothesis to be addressed. Even with such a requirement, these databases may still not be adequate to answer objectives pertaining to differences in safety profiles and permit revision of labeling. This requirement may also wipe out the efficiency gained in demonstrating efficacy of the drug via efficient means.

The use of on-target safety expressed as Type A reactions, where mechanistic explanation can supplement statistical uncertainty, presents a good opportunity for a streamlined safety database that has an opportunity to answer important questions on safety. Furthermore, it also fits with the principle of right patients and right drug and is a concept that is consistent with extrapolation. Of note, the ICH E11 (R1) defines pediatric extrapolation (for purposes of this document, simply “extrapolation”) as an approach to providing evidence in support of effective and safe use of drugs in the pediatric population when it can be assumed that the course of the disease and the expected response to a medicinal product would be sufficiently similar in the pediatric and reference population (e.g., adult, or other pediatric). On-target effects will generally be similar between adults and children barring any known age-related safety concerns; and hence, data can be extrapolated to optimize use of relevant information toward the pediatric population. This facilitates the use of Bayesian methods which provide a convenient approach for assessing probabilistic similarity and potential increase of precision needed. Because posterior distributions contain all the relevant information about a quantity of interest, they provide an intuitive representation of the associated tolerable uncertainty, enabling better decisions. Of note, even in the case for bizarre effects that are still target related, the use of Bayesian approach is still warranted.

For a clear quantitative assessment of incidence rates of AE, it is important to have the appropriate hypothesis or scientific question and decision criteria in the assessment of safety so that conclusions made are sensible. To assess these hypotheses, we have laid out a unified expression for the assessment of safety incidence rates in relation to confidence levels through posterior distributions of certain quantities of interest. This expression provides a relationship between the confidence level for ruling out certain placebo adjusted AE rate, a multiple of the placebo adjusted AE rate, sample size, observed event rates, and priors. By using on-target effects, it is possible to leverage information from related sources. Only in cases where we can use available related information can meaningful conclusions of safety happen. Otherwise, there will be a non-negligible uncertainty in these adverse reactions.

Archetypes of pediatric clinical development provide useful information to support extrapolation both from the efficacy and safety perspective and can be used to facilitate efficient use of data. In established classes, i.e., Archetypes B and C, the line of reasoning that is evolving or created gives further support for the ability to leverage data through Bayesian statistics or other meta-analytic tools. These need to be explored to speed development and ensure feasibility of trials which are competing with available drugs in the market for patients. Arguably, the on-target effects in these Archetypes are well established. Archetypes also provide benchmark on the extent of the safety database needed. In general, drugs falling into Archetype C have reduced safety database needed.

There are other questions that will need to be addressed in the future in relation to streamlined and quantitative assessments of safety. In many cases, detecting rare adverse drug reactions with latent onset may not be ruled out because trials are too short in duration to capture ADRs associated with long-term use. Unknown effects not seen in pivotal trials may be better characterized in post-market settings with appropriate mechanisms for rapid reporting. Furthermore, comparative data in pediatrics may not be available for proper characterization from background rates in the population due to the feasibility of exposing children to no treatment for an extended period. Comparative safety data is of importance when the background rate of important ADRs is high and well-established treatment already exists with effect on a clinically meaningful outcome. It is hoped that all these questions can be
catalogued, and proper statistical methods identified so that assessments of uncertainty can be clearly determined.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

The author(s) reported there is no funding associated with the work featured in this article.

**ORCID**

Margaret Gamalo http://orcid.org/0000-0003-2428-3298  
Jingjing Ye http://orcid.org/0000-0003-0086-2228  
Junjing Lin http://orcid.org/0000-0002-7448-2984

**References**

Boat, T. F., M. J. Field, eds Safe and effective medicines for children: Pediatric studies conducted under the best pharmaceuticals for children act and the pediatric research equity act. 2012.  
Bucci-Rechtweg, C. 2017. Enhancing the pediatric drug development framework to deliver better pediatric therapies tomorrow. *Clinical Therapeutics* 390 (10):1920–1932. doi:10.1016/j.clinthera.2017.07.043.  
Choi, Y., E. Ishida, S. Fürst-Recktenwald, J. Ye, D. Wajsbrot, and M. Gamalo. 2022. Analysis and reporting of pediatric growth and development assessment from clinical trials: overview and challenges. *Journal of Biopharmaceutical Statistics* 21:1–14. doi:10.1080/10543406.2022.2154357.  
Cole, T. J., J. V. Freeman, and M. A. Preece. 1998. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Statistics in Medicine* 170 (4):407–429. doi:10.1002/(SICI)1097-0258(19980228)17:4<407:AID-SIM742>3.0.CO;2-L.  
Dong, G., D. C. Hoaglin, J. Qiu, R. A. Matsouaka, Y.W. Chang, J. Wang, and M. Vandemeulebroecke. 2019. The win ratio: On interpretation and handling of ties. *Statistics in Biopharmaceutical Research* 12 (1):99–106. doi:10.1080/19466315.2019.1575279.  
Elzagallai, A. A., M. J. Greff, and M. J. Rieder. 2017. Adverse drug reactions in children: The double-edged sword of therapeutics. *Clinical Pharmacology & Therapeutics* 1010 (6):725–735. doi:10.1002/cpt.677.  
EMA. EMEA-000696-PIP02-10-M06, 2019. https://www.ema.europa.eu/en/documents/pip-decision/p/0272/2019-ema-decision-14-august-2019-acceptance-modification-agreed-paediatric-investigation-plan_en.pdf [Accessed: 2022-03-21].  
FDA. *Conducting a Clinical Safety Review of a New Product Application and Preparing a Report on the Review*, 1994. https://www.fda.gov/media/71665/download [Accessed: 2022-01-30].  
FDA. *Clinical Review: 21-035(S-040) and 21-055(S-007)*, 2005a. https://www.fda.gov/media/152421/download/ [Accessed: 2022-03-20].  
FDA. Specific Requirements on Content and Format of Labeling for Human Prescription Drugs; Revision of “Pediatric Use” Subsection in the Labeling: Final Rule, 2005b. https://www.govinfo.gov/content/pkg/FR-1994-12-13/html/94-30238.htm [Accessed: 2022-03-06].  
FDA. *Joint Clinical and Statistical Review: SNDA 21-035 S073*, 2008. https://www.fda.gov/media/71491/download [Accessed: 2022-03-20].  
FDA. *Clinical Review: 205836-SD 289, 205837-SD 122, 205838- SD125, 2021a*. https://www.fda.gov/media/152421/download [Accessed: 2022-03-20].  
FDA. *FDA New Pediatric Labeling Information Database*, 2021b. https://www.accessdata.fda.gov/scripts/sda/sdNavigation.cfm?sd=labelingdatabase [Accessed: 2021-05-26].  
Field, M. J., T. F. Boat, editors. 2012. Safety and efficacy assessments in studies conducted under b pca and prea. In *Safe and Effective Medicines for Children: Pediatric Studies Conducted Under the Best Pharmaceuticals for Children Act and the Pediatric Research Equity Act*. US: National Academies Press. doi:10.17226/13311.  
Finkelstein, D. M., and D. A. Schoenfeld. 1999. Combining mortality and longitudinal measures in clinical trials. *Statistics in Medicine* 180 (11):1341–1354. doi:10.1002/(SICI)1097-0258(19990615)18:11<1341:AID-SIM129>3.0.CO;2-7.
Gamalo, M., C. Bucci-Rechweg, R. M. Nelson, L. Vanh, A. Porcalla, H. Thackray, F. Cooner, L. Cutler, D. Friend, and R. Portman. 2022. Extrapolation as a default strategy in pediatric drug development. Therapeutic Innovation & Regulatory Science 56 (6):1–12. doi:10.1007/s43441-021-00367-9.

Golan, D. E., A. H. Tashjian, and E. J. Armstrong. 2011. Principles of pharmacology: The pathophysiologic basis of drug therapy. Lippincott Williams & Wilkins.

Human Medicines Development European Medicines Agency and Evaluation. EMA Reflection paper on extrapolation of efficacy and safety in paediatric medicine development, 2017. https://www.ema.europa.eu/en/documents/scientific-guideline/draft-reflection-paper-use-extrapolation-development-medicines-paediatrics-revision-1_en.pdf [Accessed: 2022-02-06].

Hwang, T. J., L. Orenstein, A. S. Kesselheim, and F. T. Bourgeois. 2019. Completion rate and reporting of mandatory pediatric postmarketing studies under the us pediatric research equity act. JAMA Pediatrics 1730 (1):68–74. doi:10.1001/jamapediatrics.2018.3416.

ICH. The Extent of Population Exposure to Assess Clinical Safety: For Drugs Intended for Long-term Treatment of Non-Life-Threatening Conditions, 1995. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e1a-extent-population-exposure-assess-clinical-safety-drugs-intended-long-term-treatment-non-life [Accessed: 2022-14-30].

ICH. ICH E11(R1) step 5 guideline on clinical investigation of medicinal products in the pediatric population, 2017. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e11r1-guideline-clinical-investigation-medicinal-products-pediatric-population-revision-1_en.pdf [Accessed: 2022-02-06].

ICH. Pediatric Extrapolation (Step 2), 2022. https://www.ich.org/page/efficacy-guidelines [Accessed: 2022-01-30].

Inokuchi, M., N. Matsuo, J. I. Takayama, and T. Hasegawa. 2011. BMI z-score is the optimal measure of annual adiposity change in elementary school children. Annals of Human Biology 38 (6):747–751. doi:10.3109/03014460.2011.620625.

Kim, C., K. Park, A. W. McMahon, F. G. Green, D. J. Green, and G. J. Burckart. 2021. Drug safety in labeling for pediatric drug development and dose selection in submissions to the us food and drug administration. Journal of Clinical Pharmacology 61 (S1):S133–140. doi:10.1002/jcph.1864.

Liu, F. 2016. A simple bayesian approach to quantifying confidence level of adverse event incidence proportion in small samples. Journal of Biopharmaceutical Statistics 260 (3):499–506. doi:10.1080/10543406.2015.1052489.

Lynch, J. J., III, T. R. Van Vleet, S. W. Mittelstadt, and E. A. Blomme. 2017. Potential functional and pathological side effects related to off-target pharmacological activity. Journal of Pharmacological and Toxicological Methods 87:108–126. doi:10.1016/j.vascn.2017.02.020.

McMahon, A. W., and G. Dal Pan. 2018. Assessing drug safety in children—the role of real-world data. The New England Journal of Medicine 3780 (23):2155. doi:10.1056/NEJMpt1802197.

Morgan, P., D. G. Brown, S. Lennard, M. J. Anderton, J. Carl Barrett, U. Eriksson, M. Fidock, B. Hamren, A. Johnson, R. E. March, et al. 2018. Impact of a five-dimensional framework on r&d productivity at AstraZeneca. Nature Reviews Drug Discovery. 170(3):167–181. doi:10.1038/nrd.2017.244.

Peng, L. 2020. The use of the win odds in the design of non-inferiority clinical trials. Journal of Biopharmaceutical Statistics 300 (5):941–946. doi:10.1080/10543406.2020.1757690.

Pham-Gia, T., and N. Turkkan. 1998. Distribution of the linear combination of two general beta variables and applications. Communications in Statistics—Theory and Methods 270 (7):1851–1869. doi:10.1080/03610929808832194.

Pham-Gia, T., N. Turkkan, and P. Eng. 1993. Bayesian analysis of the difference of two proportions. Communications in Statistics—Theory and Methods 220 (6):1755–1771. doi:10.1080/03610929308831114.

Pocock, S. J., C. A. Ariti, T. J. Collier, and D. Wang. 2012. The win ratio: A new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. European Heart Journal 330 (2):176–182. doi:10.1093/eurheartj/ehr352.

Rodriguez, W., A. Seien, D. Avant, C. Chaurasia, T. Crescenzi, G. Gieser, J. Di Giacinto, S.M. Huang, P. Lee, L. Mathis, et al. 2008. Improving pediatric dosing through pediatric initiatives: What we have learned. Pediatrics. 1210 (3):530–539. doi:10.1542/peds.2007-1529.

Rudmann, D. G. 2013. On-target and off-target-based toxicologic effects. Toxicologic Pathology 410 (2):310–314. doi:10.1177/019262312464311.

Seifu, Y. 2023. A note on benefit-risk assessment focused trial designs: Application to a pediatric design. Journal of Biopharmaceutical Statistics. Accepted.

Tanaka, T., N. Kato, S. Yokoya, A. Ono, T. Isojima, H. Yokomichi, Z. Yamagata, S. Tanaka, H. Matsubara, M. Ishikuro, et al. 2021. Changes in height standard deviation scores during early life are affected by nutrition. Pediatrics International. 63(6):710–715. doi:10.1111/ped.14577.

Whitebread, S., B. Dumotier, D. Armstrong, A. Fekete, S. Chen, A. Hartmann, P. Y. Muller, and L. Urban. 2016. Secondary pharmacology: Screening and interpretation of off-target activities—focus on translation. Drug Discovery Today 210 (8):1232–1242. doi:10.1016/j.drudis.2016.04.021.