資料 2-2

Leveraging Existing Clinical Data for Extrapolation to Pediatric Uses of Medical Devices

Draft Guidance for Industry and Food and Drug Administration Staff

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.
Document issued on: May 6, 2015

You should submit comments and suggestions regarding this draft document within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit written comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Room 1061, Rockville, MD 20852. Submit electronic comments to http://www.regulations.gov. Identify all comments with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this document, contact Jacqueline Francis (CDRH) at (301) 796-6405 (Jacqueline.Francis@fda.hhs.gov) or the Office of Communication, Outreach, and Development (CBER) at 800-835-4709 or 240-402-7800.

U.S. Department of Health and Human Services
Food and Drug Administration

Center for Devices and Radiological Health

Center for Biologics Evaluation and Research

Preface

Additional Copies

Additional copies are available from the Internet. You may also send an email request to dsmica@fda.hhs.gov to receive an electronic copy of the guidance or send a fax request to 301-827-8149 to receive a hard copy. Please use the document number (1827) to identify the guidance you are requesting.

Additional copies are available from the Center for Biologics Evaluation and Research (CBER) by written request from the Office of Communication, Outreach and Development (OCOD), 10903 New Hampshire Ave., WO71, Room 3128, Silver Spring, MD 20993-0002, or by calling 1-800-835-4709 or 240-402-7800, by email, ocod@fda.hhs.gov, or from the Internet at http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.

Table of Contents

Figures and Tables

Leveraging Existing Clinical Data for Extrapolation to Pediatric Uses of Medical Devices

Draft Guidance for Industry and Food and Drug Administration Staff

This draft guidance, when finalized, will represent the current thinking of the Food and Drug Administration (FDA or Agency) on this topic. It does not establish any rights for any person and is not binding on FDA or the public. You can use an alternative approach if it satisfies the requirements of the applicable statutes and regulations. To discuss an alternative approach, contact the FDA staff responsible for this guidance as listed on the title page.

1. Introduction

The Food and Drug Administration (FDA) aims to promote safe and effective device use in pediatric patients, while ensuring device approvals are based on valid scientific evidence.[1] Currently, there is a paucity of scientific evidence available to substantiate submissions for devices that are indicated for use in the diagnosis or treatment of pediatric patients. FDA believes that leveraging relevant available clinical data, when appropriate, may lead to more devices being approved for pediatric indications, which will increase the availability of medical devices with appropriate labeling to support safe and effective device use in pediatric patients. This approach will potentially streamline

[1] Valid scientific evidence is evidence from well-controlled investigations, partially controlled studies, studies and objective trials without matched controls, well-documented case histories conducted by qualified experts, and reports of significant human experience with a marketed device, from which it can fairly and responsibly be concluded by qualified experts that there is reasonable assurance of the safety and effectiveness of a device under its conditions of use. (21 CFR 860.7(c)(2))

the requirements for establishing a pediatric intended use claim, and enhance and encourage pediatric device development programs.

FDA's guidance documents, including this guidance, do not establish legally enforceable responsibilities. Instead, guidance documents describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word "should" in Agency guidance documents means that something is suggested or recommended, but not required.

2. Overview

The objectives of this guidance are: (1) to increase the availability of safe and effective pediatric devices by leveraging relevant existing clinical data for use in pre-market approval applications (PMAs) and humanitarian device exemptions (HDEs); (2) to explain the circumstances in which FDA believes it may be appropriate to leverage existing clinical data to support pediatric device indications and labeling; (3) to outline the approach FDA uses to determine whether extrapolation is appropriate, and if so, to what extent the data can be leveraged; and (4) to describe statistical methodology that can be used to leverage the data in a way that increases precision for pediatric inferences.

For the purposes of this document, "extrapolation" refers to the leveraging process whereby an indication for use of a device in a new pediatric patient population can be supported by existing clinical data from a studied patient population. That is, when existing data are relevant to a pediatric indication and determined to be valid scientific evidence, we believe that it is scientifically appropriate in certain circumstances to attempt to extrapolate such data to a pediatric use in support of demonstrating a reasonable assurance of effectiveness and occasionally safety.

This draft guidance explains when and how existing clinical data in another studied population (such as adults, or a different pediatric subpopulation) may be leveraged ("extrapolated") to support marketing approval and labeling of medical devices for use in pediatric patients. In order to make decisions about the effectiveness and safety of a medical device in pediatric patients, FDA considers the totality of the evidence available. As with any PMA or HDE, FDA will still consider clinical data (whether extrapolated or not) alongside other forms of scientific evidence from assessments of device performance (e.g., preclinical testing, engineering models, biocompatibility, virtual patient simulations, statistical models, etc.) to determine whether the sponsor has demonstrated a reasonable assurance of safety and effectiveness (or probable benefit for HDEs).

This guidance does not change the threshold for regulatory approval or valid scientific evidence. Instead, the document seeks to provide clarity and predictability for device sponsors and to ensure consistency within FDA regarding the specific criteria that should be considered when deciding whether leveraging existing clinical data to support

pediatric claims is appropriate, and if so, to what extent. When considering extrapolation, sponsors are encouraged to engage FDA early in product development planning.

This guidance should be used in conjunction with other guidance documents for pediatric medical devices and other applicable device-specific guidance documents to help ensure that medical devices intended for use in the pediatric population provide reasonable assurance of safety and effectiveness (or probable benefit, for HDE).

The scope of this draft guidance includes medical devices subject to the PMA and HDE premarket requirements. For these premarket submissions, it may be appropriate to extrapolate existing clinical data when the course of the disease or condition and effects of the device are sufficiently similar in adults and pediatric patients, and the existing data is determined to be valid scientific evidence. FDA believes that extrapolation should be limited to circumstances in which endpoints used in the adult data sources are relevant to the pediatric population, and the quality of these data is high. In this context, it is important to note that the consideration of whether to borrow existing data to extrapolate effectiveness for a pediatric population is independent from the consideration of whether to extrapolate for safety. In other words, the criteria that govern the decision of whether or not to extrapolate are considered separately for effectiveness and for safety.

This guidance facilitates efforts to address an unmet medical device need for pediatric patients. The framework described herein is one tool to make optimal use of what is already known about device effects in other populations to support indications in the pediatric population

3. Background

When considering extrapolation of existing data for pediatric indications, it is important to understand how pediatric subpopulations are identified. Section 520(m)(6)(E)(i) of the Federal Food, Drug, and Cosmetic Act (FD&C Act)[2] defines "pediatric patients" as persons aged 21 or younger at the time of their diagnosis or treatment (i.e., from birth through the 21st year of life, up to but not including the 22nd birthday). Pediatric subpopulations are defined in Section 520(m)(6)(E)(ii) (and adopted by reference in Section 515A(c) of the FD&C Act) to be neonates, infants, children, and adolescents.

Age ranges for these pediatric subpopulations are as follows:

- Neonates: from birth through the first 28 days of life
- Infants: 29 days to less than 2 years

[2] Available at http://www.fda.gov/regulatoryinformation/legislation/federalfooddrugandcosmeticactfdcact/default.htm

- Children: 2 years to less than 12 years
- Adolescents: aged 12 through 21 (up to but not including the 22nd birthday)

In 2004, FDA published a guidance document entitled "Premarket Assessment of Pediatric Medical Devices" in an attempt to clarify the types of information needed to provide reasonable assurance of safety and effectiveness of medical devices intended for use in pediatric patients and to promote the development of these devices. This document indicates that data can be extrapolated to support effectiveness and, on a limited basis, safety for premarket approval applications (PMAs) when consistent with scientific principles. The guidance states the following:

> "If it is determined that clinical data are needed, it may be that the course of the disease and the device's effects are similar in adult and pediatric patients. In such a situation, the pediatric indication may be supported by the adult data with limited additional safety data in the pediatric population."[3]

This guidance document was updated in 2014 to make clear that, as with other forms of valid scientific evidence used to demonstrate effectiveness and safety for a device intended for a pediatric population, the amount and type of extrapolated data necessary to support a pediatric indication for a device varies:

> "As is true for medical devices in general, FDA does not believe that clinical data will be necessary to demonstrate effectiveness and safety for all devices intended for pediatric populations. The agency recognizes that the amount and type of evidence required will depend on a number of factors, including the nature of the device, what is already known about the product in the adult population (if relevant), what is known or can be extrapolated about the device to the pediatric population, and the underlying disease or condition being treated. In some cases, well-designed bench and animal testing will be sufficient to evaluate the device. In others, clinical data may be needed to evaluate the safety and effectiveness of the device."[3]

Congress was aware of the 2004 version of this guidance document when it passed the Food and Drug Administration Amendments Act of 2007 (FDAAA). The House Report (H.R. Rep. 110-225) states:

> "FDA addressed premarket review of medical devices intended for pediatric patients by issuing a guidance in May 2004 entitled 'Premarket Assessment of Pediatric Medical Devices.' The guidance was published pursuant to the Medical Device User Fee and Modernization Act, which contained several provisions intended to promote the development of safe and effective pediatric devices. In

[3] Available at http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm089740.htm

> this guidance, FDA defined the age ranges for pediatric subpopulations, **identified the types of information needed to provide reasonable assurance of the safety and effectiveness of medical devices intended for use in the pediatric population**, and described the protections that sponsors should consider for pediatric subjects involved in clinical trials" (emphasis added).

Title III of FDAAA is the Pediatric Medical Device Safety and Improvement Act (PMDSIA)[4] of 2007. PMDSIA specifically authorized the use of adult data to demonstrate pediatric effectiveness[5], stating:

> "If the course of the disease or condition and the effects of the device are sufficiently similar in adults and pediatric patients, the Secretary may conclude that adult data may be used to support a determination of a reasonable assurance of effectiveness in pediatric populations, as appropriate."

In addition to allowing for the extrapolation of adult data to pediatric populations, the provision indicates that, when appropriate, data can be extrapolated from one pediatric subpopulation to another.[4]

While PMDSIA addresses the extrapolation of existing data to support a determination of a reasonable assurance of effectiveness, it does not address safety data. However, we believe there are also specific cases where it will be appropriate to consider extrapolation of existing clinical safety data to support or enhance evidence for pediatric indications for medical devices, including those defined in this guidance (e.g., the effects of the device under consideration are identical when used in pediatric and adult populations and the course of the disease or condition and associated risk factors are the same between the two populations).

Given the potential for similarity in disease or condition, device attributes and treatment effects between patient populations, and the availability of other nonclinical forms of evidence to assess safe device performance, we believe that extrapolating for safety in medical devices in specific circumstances could be appropriate and consistent with the requirement to base approval decisions on valid scientific evidence. Because the mechanism of action for devices is often well-characterized and fairly localized, non-clinical forms of scientific evidence may provide information about device performance characteristics related to safe device functioning (e.g., preclinical testing, engineering models, computer modeling, or other nonclinical data). The potential availability of these

[4] Available at http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/DevelopmentResources/UCM049870.pdf

[5] The term "effectiveness" is defined as follows: "There is reasonable assurance that a device is effective when it can be determined, based upon valid scientific evidence, that in a significant portion of the target population**,** the use of the device for its intended uses and conditions of use, when accompanied by adequate directions for use and warnings against unsafe use, will provide clinically significant results" (21 CFR 860.7).

types of data for medical devices provides further support for the use of extrapolated clinical data to demonstrate safety in pediatric patients. However, full extrapolation[6] of safety data is expected to occur rarely. The appropriateness of extrapolation for effectiveness and/or safety is considered independently on a *case by case* basis following the decision tree described in Section 6.

This guidance does not change the threshold for regulatory approval or for valid scientific evidence. When existing clinical data is relevant and appropriate for leveraging, the amount of prospective clinical data in the pediatric population needed to demonstrate a reasonable assurance of effectiveness and/or safety (or probable benefits outweigh risks, for HDE) may be reduced. If not appropriate or insufficient to meet the threshold of valid scientific evidence, data will not be extrapolated.

4. Why Extrapolate from Adult Data for Pediatric Use?

The extrapolation of adult data for pediatric use may benefit pediatric patients by increasing the availability of medical devices with appropriate labeling to support safe and effective pediatric use. Extrapolation, when appropriate, facilitates the use of available relevant data by making use of existing clinical data that may be helpful for understanding device performance in pediatrics. This is similar to the Bayesian concept of borrowing from prior adult information to come to a posterior conclusion about pediatric effectiveness or safety[7]. Extrapolation of adult data is limited to situations in which the course of the disease or condition and the effects of the device are sufficiently similar in adults and pediatric patients. For example, data from studies of devices that create intracranial arteriotomies in adults may offer insights into their effectiveness in pediatric patients between the ages of 13 and 21 because it is widely accepted that cerebral vasculature of this age group is similar to that of adults.

There are many potential challenges involved in conducting pediatric clinical trials to support pediatric indications for devices. For example:

- Small and diffusely scattered potential pediatric populations lead to small trial sizes.
- Challenges exist in enrollment and consent procedures, which could increase the length of time needed to determine safety and effectiveness.
- There are more variations in pathophysiology, physiology, anatomy, and human factors in children and within pediatric subpopulations as compared to adults.

[6] See definition of "full extrapolation" in Section 5.1.
[7] See FDA's "Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials," available at http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm071072.htm

- Reference samples may require an amount of blood too voluminous to obtain safely from a neonate or small child.

At least in part because of these challenges, relatively few devices have pediatric-specific indications and labeling. Yet off-label use of adult devices, without labeling information to guide safe and effective use in pediatric patients, is not uncommon. The use of existing clinical data when appropriate may reduce the need to prospectively conduct large pediatric clinical trials by bolstering other scientific evidence supporting a reasonable assurance of safety and effectiveness in a pediatric population. Extrapolation encourages industry to provide performance data to support a pediatric indication, which promotes proper labeling for use in pediatric patients even when limited pediatric data are available. Informative labeling of a device which promotes safe and effective pediatric use ultimately benefits patients.

5. Borrowing Strength from Adult Data

Extrapolation enables a sponsor to leverage adult data to support demonstration of a reasonable assurance of effectiveness and possibly the safety of a medical device for pediatric use. The quantitative information provided by existing adult data may be important, and thus can be incorporated either by standing in for any potential pediatric data or within a statistical model that also includes some pediatric data. The statistical model would then estimate a device effect or adverse event in the pediatric population, which can be potentially bolstered by the incorporation of additional data from adults. This is known as "borrowing strength" in statistical literature (Carlin & Louis, 2009). Such borrowing can bolster the sample size of a prospective pediatric study. The exact model used to borrow strength may vary case by case. However, for all models, the extent of leveraging depends, in part, on the similarity between borrowed data and any pediatric data that will be collected.

The extent of borrowing may also be moderated by clinical judgments that are not inherently implied by the statistical modeling. This may include consideration of the particulars of the populations and studies, and whether such data are intended to demonstrate either safety or effectiveness (or both). Effectiveness and safety often have different endpoint assessments in a study. In addition, the study design could be different for different endpoints, or there could be different considerations in the pediatric population for safety versus effectiveness. Therefore, safety is considered independently from effectiveness in deciding whether or not extrapolation may be appropriate. Section 6 provides more details about important information needed in the decision to extrapolate.

Existing clinical data from adults and some non-clinical studies may provide information about device safety which is relevant to risks in children. For some devices, the mechanism of action is expected to be similar in adults and pediatric patients. In these cases, non-clinical forms of scientific evidence may provide some information about

many device performance characteristics related to safe device functioning (e.g., preclinical testing, engineering models, computer modeling, or other nonclinical data). However, the sole use of non-clinical data as the basis for valid scientific evidence regarding safety is expected to be exceedingly rare. Likewise, existing clinical data from adults may provide information about device safety which is relevant to risks in children. Based on the nature of the similarities and differences between target populations and on the quality of the existing data, additional clinical studies in pediatric patients may be warranted to supplement the existing data to provide valid scientific evidence about device safety.

Types of existing data sources that may be considered for extrapolation include (but are not restricted to) data from a variety of clinical investigations (e.g., randomized controlled trials, single arm studies, and from any individual treatment arm), historical clinical data, reference samples, and published literature.

5.1 Full and Partial Extrapolation

Existing clinical data may be leveraged either fully or partially via statistical modeling, to support a reasonable assurance of safety or of effectiveness in a pediatric patient population. The following are the differences between full and partial extrapolation:

- **Full Extrapolation**: Existing clinical data are used directly (i.e., as a complete substitute) for prospective pediatric clinical data in support of a determination of a reasonable assurance of effectiveness or of safety for a pediatric device. No prospective pediatric clinical data are anticipated for the endpoint being fully extrapolated. However, as with any PMA or HDE, FDA will consider this alongside other data sources, such as virtual patient simulations, bench data, mechanical models, literature studies or case reports, as further valid scientific evidence supporting a reasonable assurance of safety and effectiveness in the intended pediatric population. Given the range of potential differences between adult and pediatric patients, full extrapolation of existing clinical data to demonstrate safety is expected to be rare.
- **Partial Extrapolation**: Existing data are combined via a statistical model with pediatric data sources or prospective pediatric clinical data in support of demonstrating a reasonable assurance of effectiveness or of safety for a pediatric device. The construction of such a statistical model is anticipated to require the availability of measured variables that will help connect the adult outcomes to the pediatric outcomes. If necessary variables are not available in the data sources, partial extrapolation may not be appropriate. If the model is determined to be appropriate, then the inferences obtained from it may be used to support a pediatric indication.

Full extrapolation requires a significant amount of trust in the relevance and quality of the adult data because they will constitute the sole clinical data to support effectiveness and possibly safety of the device in pediatric patients. Partial extrapolation also requires trust

in the adult data, specifically, the trust that the adult data are similar to what is expected to occur in pediatric patients. Furthermore, because the *actual* extent of partial extrapolation (or borrowing) will be determined *after* the pediatric data are gathered, there is some verification of whether extrapolation is ultimately appropriate. If extrapolation is ultimately not appropriate, then the pediatric data will need to be sufficient alone to support marketing approval. Section 6 of this document describes the approach that is used to determine whether existing clinical data sources are candidates for borrowing either fully or partially to extrapolate either effectiveness, safety, or both to a pediatric population.

5.2 Extrapolation for Effectiveness vs. Safety

FDA believes that existing clinical data can be extrapolated when appropriate to support either effectiveness or safety or both in medical devices. However, since the endpoints related to effectiveness are likely different from those for safety in a given study, and because the quality of data may differ in some circumstances, the decisions of whether to extrapolate existing data for safety or effectiveness (or both) are made independently. For example, in medical devices, there may be circumstances where FDA may conclude (based on the flowchart in Section 6.1) that full extrapolation of adult data is appropriate for effectiveness, but there is still a need for a safety study in a pediatric population.

Because of the physiological differences between adult and pediatric patients that may affect device safety and the inherent difficulties in designing and powering clinical studies that provide comprehensive assessments of safety, extrapolation for safety is expected to be rarer than extrapolation for effectiveness. However, we believe that there are cases where extrapolation for safety is appropriate in some cases to support a pediatric indication. Again, these data will be considered with the totality evidence to either support or not support a reasonable assurance of safety and effectiveness (or probable benefit in HDEs).

6. Pediatric Extrapolation Decision Process

The extrapolation approach described in this guidance document provides a framework for considering whether or not the extrapolation of existing clinical data is appropriate to support a pediatric indication, and if so, to what extent.

Extrapolation of adult data may be used, as appropriate, to support a pediatric indication if the course of disease or condition and the effects of the device are sufficiently similar in adults and pediatric patients. The appropriateness of extrapolation largely depends on three main factors: (1) the similarity of the existing adult response data and/or population characteristics to the intended pediatric population, the (2) the quality of the adult data in terms of study design, data collection, and measurement, and (3) whether extrapolated data may be used to fairly and responsibly decide whether there is a reasonable assurance of the safety and effectiveness (or probable benefit for HDEs) of a medical device (i.e.,

constitute valid scientific evidence). Broadly, factors that can affect data quality include study design, data collection and measurement, and the applicability of these data with consideration of the current standard of practice for the disease or condition being treated.

When both similarity and quality are determined to be sufficiently high, there is a greater level of certainty that the existing data can be appropriately considered for extrapolation to the intended pediatric subpopulation. If neither similarity nor data quality are high, then the existing adult data may be inappropriate to use for extrapolation purposes.

6.1 Pediatric Extrapolation Decision Tree

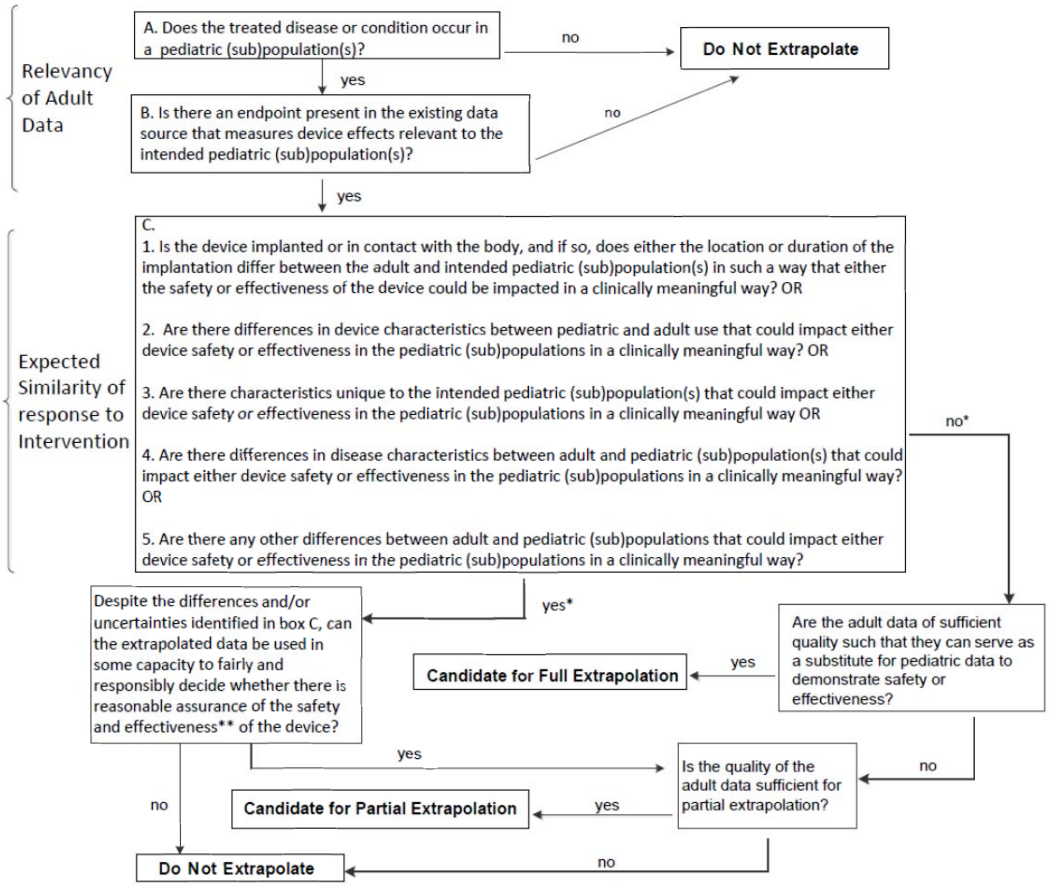
The following decision tree (see Figure 1 below) can be used by sponsors and FDA review staff as a tool to help determine whether extrapolation of existing clinical data is appropriate and, if so, whether extrapolation should be full or partial.

Please note that the approach described in the decision tree is intended as an aid to decide whether or not extrapolation can be considered in a specific situation. A conclusion from the decision tree that extrapolated data may be used does not necessarily mean that these data will support an approval decision for the PMA or HDE. If it is determined that existing data can be extrapolated in some manner to support a pediatric indication, the extrapolated data would be considered in conjunction with the totality of evidence that will either support or not support a reasonable assurance of safety and effectiveness (or probable benefit, for HDE).

The general approach of the decision tree is to first consider whether the treated condition occurs at all in the intended pediatric population and whether available adult data related to that condition and effect of the device are relevant to the intended pediatric population. One potential (and perhaps readily available) source of relevant data includes prior clinical studies done for approval of the device in adults. If these adult studies use an endpoint that is similar to the primary endpoint of interest in the pediatric population, then they may be relevant for extrapolation. If no *relevant* data are available from any prior adult studies, then extrapolation should not be used. Second, consider to what extent the adult data are similar to what may be seen in the pediatric population. For example, are there expected differences in the device characteristics, patient characteristics, or disease characteristics between the identified adult population and the intended pediatric (sub)population(s)? If there are expected differences, extrapolation might not be appropriate. The differences could contribute to a high level of uncertainty regarding the expected device effect such that the adult data cannot support a pediatric indication. On the other hand, if such differences are minimal and can be accounted for with the measurement of covariates or surrogate variables within a statistical model, partial extrapolation may be appropriate. If there are *no* expected differences, then full extrapolation could be an option if the quality of the adult data is such that *substituting* adult data for pediatric data is considered appropriate.

The decisions to extrapolate for safety or effectiveness are made by going through the tree independently for each of these factors. In the tree, there will be items that will remain constant for either decision. For example, when considering whether to extrapolate for safety, effectiveness or both, the considerations related to the similarities or differences in disease progression and device characteristics between the adult and pediatric populations may be the same. However, endpoints and the quality of data relating to these endpoints may differ when considering the safety or effectiveness components of a prior study.

Figure 1. Pediatric Extrapolation Decision Tree



* Note that if all five questions in Box C are answered "no", the direction from C is "no". If at least one of the five is answered "yes", the direction from C is "yes".

**"The agency relies upon only valid scientific evidence to determine whether there is reasonable assurance that a device is safe and effective. Valid scientific evidence is evidence from well-controlled investigations, partially controlled studies, studies and objective trials without matched controls, well-documented case histories conducted by qualified experts, and reports of significant human experience with a marketed device, from which it can fairly and responsibly be concluded by qualified experts that there is reasonable assurance of the safety and effectiveness of a device under its conditions of use. 21 CFR 860.7(c)(1)&(2)."

The questions in the Pediatric Extrapolation Decision Tree are a guide for what to consider when determining the appropriateness of extrapolation of adult data for pediatric indications. These questions are designed to promote discussion between FDA review

staff and sponsors while facilitating consistency among FDA review staff. Considerations of extrapolation of any type should be discussed with FDA staff throughout the device protocol planning stages. It is highly recommended that the pre-submission pathway be used to explore such options.[8] A 522 post-market surveillance study may be required, particularly in situations where full extrapolation of safety data is agreed upon by FDA staff and device manufacturers.[9]

6.2 Questions in the Pediatric Extrapolation Decision Tree

This section provides more detail about using the questions in the Pediatric Extrapolation Decision Tree to make a decision regarding the appropriateness of extrapolation. The first two questions are asked to determine whether extrapolation should be considered at all. Within the tree, these are referred to as "Relevancy" Questions because they pertain to whether adult data are relevant for extrapolation.

Question A: Does the treated disease or condition in question occur in pediatric (sub)populations?

If the answer is no, extrapolation of adult data is not appropriate. If the answer is yes, proceed to question B.

Question B: Is there an endpoint present in the existing data source that measures device effects relevant to the intended pediatric (sub)population(s)?

In order to borrow confidently from adult data there should be either: (1) the same variable measured in the adult data as would be expected to be measured as the primary endpoint in the intended pediatric population, or (2) a variable measured in the existing adult data that is *sufficiently related* to the primary endpoint expected to be measured in the pediatric population. For the latter case, a reliable and valid model might be used to predict the endpoint from the pediatric population using the endpoint from the adult population. Reliability and validity of the model should be established from prior investigations. One possibility is to use a validated surrogate endpoint in the adult data set(s) that has been shown to predict a different (perhaps, longer term) endpoint of interest. For example, a device that is used to treat diabetes may rely on validated adult and pediatric surrogate endpoints such as serum glucose levels or HbA1c to measure actual device outcomes.

If the answers to questions A. and B. are yes, continue along the decision tree. The next five questions are addressed as a set (Questions C.). Within the tree, we label these

[8] http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM311176.pdf

[9] http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMA/pss.cfm

questions as pertaining to "Similarity". The questions in box C essentially ask whether there are differences between the adult and pediatric populations, or the devices used in each population, that could impact the safety and effectiveness of the device in pediatric (sub)population(s). In other words, the questions in Box C serve to address whether or not the course of the disease and the effects of the device are sufficiently similar in adults and pediatric patients, and if so, to define what those similarities (as well as differences) are.

To determine that the effectiveness and safety of the device is similar across adult and pediatric populations, a basic consideration is that the direction of benefit from the device on the outcome should be the same across populations. That is, if the device has a positive effect on adults, then it should also have a positive effect on the intended pediatric population, for the endpoint under study. Some devices are intended to benefit an adult population but are not expected to benefit a pediatric population, and might even worsen the pediatric patient's condition. For example, a device used for damaged joints in adults is considered for the same indication in children; however, because children do not have closed growth plates, the device could cause significant problems for children who are still actively growing. Therefore, the direction of benefit is not the same. The *magnitude* of the device effect should also be similar. Evaluating the extent of similarity of the magnitude across populations may involve research into published literature, and should be considered on a case by case basis.

The questions in C. should be used to help answer whether the device is expected to have a similar effectiveness and/or safety result across populations. Differences tend to increase the amount of uncertainty in statistical inference when extrapolating from adult to pediatric patients. If all of the five questions are answered "no" for either safety or effectiveness or both, then full extrapolation can be considered if the adult data are of sufficiently high quality. If any of the questions in C. are answered "yes", then the review team should determine whether the adult data provide useful information for partial extrapolation by revisiting answers to the questions within C. as well as any additional important information.

Questions Box C.

Question C-1: Is the device implanted or in contact with the body, and if so, does either the location or duration of implantation differ between the adult and intended pediatric (sub)population(s) in such a way that the safety and effectiveness of the device could be impacted in a clinically meaningful way?

If the location or duration of implantation differs and the difference is expected to impact device safety or effectiveness, then full extrapolation is probably not feasible. However, partial extrapolation may still be viable if the quality of adult data is sufficiently high such that statistical and clinical modeling can account for the difference, and FDA can fairly and responsibly use such data to conclude that there is reasonable assurance of effectiveness and/or safety of the device.

Question C-2: Are there differences in device characteristics between pediatric and adult use that could impact either device safety or effectiveness in the pediatric (sub)population(s) in a clinically meaningful way?

For instance, sometimes device modifications (e.g., design, materials, and mechanisms of use) must be made in order to use a device in a pediatric population. To the extent these modifications could impact device safety or effectiveness in a clinically meaningful way, the answer to this question will be yes. Possible differences might include, but are not limited to, differences in human factors issues (e.g., self-administration versus administration by a guardian), reference or normal values, size, scaling of the device, blood sampling or sample quantity issues for in vitro diagnostic devices, energy, delivery, device function, or device materials. This question is also related to whether conditions for preclinical or clinical testing differ between adult and intended pediatric (sub)population(s) and whether the device needs to change over time to accommodate growth and development. Sometimes device characteristics and patient characteristics are intertwined. For example, if the normal value (for diagnostics) or performance (for therapeutics) of the device depends on a body measurement or unique physiology that differs between adult and intended pediatric (sub)population then the device measurement could differ.

If the answer to question 2 is "yes", then full extrapolation is not feasible. However, as described above, partial extrapolation may still be viable in certain circumstances.

Question C-3: Are there characteristics unique to the intended pediatric (sub)population(s) that could impact either the effectiveness or safety of the device when used in the pediatric (sub)population(s) in a clinically meaningful way?

Some devices might require special considerations that affect only pediatric patients; for example:

- Growth of the child during the device performance period
- Specimen sample size or quantity
- Reference or normal values
- For serologic in vitro diagnostic devices, specific challenges in certain subgroups due to differing immune status
- Analytical issues which affect interfering substances for in vitro diagnostic devices
- Drug dose or metabolic differences for therapeutic drug monitoring devices
- Pediatric human factors
- Increased impact of time exposure to younger subjects (e.g., long-term toxicity differences between populations)

The kinetics or physiology might differ between adults and children, which might then influence the interpretation of test results or treatment modality, ultimately impacting the effectiveness or safety of the device across populations.

An example where patient characteristics might affect interpretation of data concerning device effectiveness is a device indicated for weight loss. In this case, an adolescent and an adult may have different body sizes and/or masses that may impact evaluation of a device's effectiveness. For an adolescent study subject, weight gain could be attributed to normal growth, whereas for an overweight adult, weight gain would more likely demonstrate the failure of the device to have its intended effect.

If the answer to question C-3 is yes, full extrapolation is probably not feasible. However, as described above, partial extrapolation may still be viable in certain circumstances.

Question C-4: Are there differences in disease characteristics between adult and pediatric (sub)population(s) that could impact either device safety or effectiveness in the pediatric (sub)population(s) in a clinically meaningful way?

For some devices, there might be differences in disease characteristics between adults and children, which are highly likely to affect how the device performs or how test results are interpreted. The prevalence or severity of disease characteristics might differ between adults and children, or the natural course of the disease might differ. For example, a diagnostic device could indicate the need for medical intervention differently for children than for adults because analyte levels considered safe may differ for each population.

If the answer to question C-4 is yes, full extrapolation is probably not feasible. However, as described above, partial extrapolation may still be viable in certain circumstances.

Question C-5: Are there other differences between adult and pediatric (sub)population(s) that could impact either device effectiveness or safety in the pediatric (sub)population in a clinically meaningful way?

This question allows for consideration of other differences that are not addressed by the first four questions.

If the answers to questions Box C are all "no", and if the adult data are of sufficiently high quality, then full extrapolation could be considered, and it is possible that no pediatric data would be needed to achieve approval for the pediatric indication

Study design and sampling plan are factors that could influence data quality. A registry or single-arm study is of lower quality than a randomized controlled (and blinded) trial. Responses from registries or single-arm studies may be biased in favor of the device because the subjects know they are receiving a new treatment that they hope to be better than the current standard of care. Allowing study subjects to choose their own treatment arms instead of randomly assigning them to treatments may be similarly biased. The "Guidance for Industry, Clinical Investigators, and Food and Drug Administration

Staff—Design Considerations for Pivotal Clinical Investigations for Medical Devices,"[10] issued in 2013, compares study designs in terms of general quality, and represents the agency's proposed approach on this topic.

If the answer to any or all of questions in Box C is "yes", then the decision on whether partial extrapolation is appropriate will depend on whether some prospectively collected pediatric data can be obtained and/or whether an appropriate statistical model can be constructed such that pediatric outcomes can be predicted reliably by borrowing strength from the adult data. As stated above, statistical models may be used to combine relevant adult data with pediatric data in order to increase precision in inferences made from a pediatric study. These models can then potentially account for differences identified in the decision tree (see Section 7 and Appendix B for a discussion). In this way, the borrowed or extrapolated data have the potential to be used in some capacity to fairly and responsibly conclude that there is reasonable assurance of the effectiveness and/or safety of the device. If it determined that existing clinical data cannot be fairly and responsibly used in some capacity to conclude that there is a reasonable assurance of effectiveness and/or safety, extrapolation should not be considered.

It is important to reiterate that any anticipated differences between adult and pediatric populations may not be realized until after the pediatric study is finished, if a study is recommended. Therefore, the *realized* extent of partial extrapolation is determined *after* data become available, and the statistical model is fit to the adult and pediatric data.

If there are other device- or disease-specific questions not addressed in the Pediatric Extrapolation Decision Tree that could assist the FDA review team in its review, those questions may also be considered under Question C-5 in the tree. These situations may be more complex and require thoughtful collaboration between the FDA review team and the sponsor to determine whether extrapolation might be feasible. Borrowing of data may be achieved for some areas, while the sponsor may need to collect data in other areas. See Appendix A for examples.

7. Factors That Could Limit Extrapolation

This section provides a series of general factors that can aid in responding to the questions posed by the decision tree and determining whether, and to what extent, extrapolation is appropriate.

Factors that may preclude extrapolation of any adult data include but are not limited to the following:

[10] Available at http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm265553.htm.

- There is little knowledge of the disease or condition in pediatrics.
- The device is not FDA approved or cleared for adults.
- Endpoints cannot be directly borrowed.
- Statistical models cannot account for differences.
- Human factors and growth can affect safety in pediatric patients (these factors don't exist in adults).
- Appropriate labeling cannot be written for the pediatric population or subpopulation(s) targeted.
- The practice of medicine has changed since the device was initially approved to such an extent that historical data would likely be different than prospectively collected data.
- Appropriate risk mitigation cannot be assured.

Factors that may limit extrapolation to a partial extent and thus require conducting a prospective study of pediatric patients include, but are not limited to, the following:

- The age difference between the pediatric (sub)population and the available adult data is too great, making it difficult to infer similarity in risk or effectiveness. In such cases, it may be more appropriate to extrapolate to a pediatric age that is closer to the mean age of the adult population. For example, it might be more appropriate to extrapolate young adult data to an adolescent indication than to a neonate indication.
- Other supportive pediatric data are outdated and may not properly represent current treatment trends and practices.
- There are important differences between the adult and pediatric (sub)population(s) such that the adult data cannot substitute for data from a potential pediatric study to fairly and responsibly conclude that there is reasonable assurance of the safety and effectiveness of the device in the pediatric population.

Whether any of these factors would preclude extrapolation or limit it to a partial extent depends on how the differences are expected to influence potential conclusions of the new study.

8. Uncertainty in Extrapolating Data

Extrapolation does add uncertainty into FDA's assessment of the effectiveness and safety of a device. Whether extrapolating partially or in full, there remains some uncertainty even though statistical modeling may be used to account for observed differences and increase precision of inferences. The extent of this uncertainty depends on the

differences between the two populations and the quality of the data. FDA considers this uncertainty as a factor when making benefit-risk determinations.

The Benefit-Risk Guidance[11] should be consulted for understanding how extrapolated data might be weighed within a benefit-risk framework when considering device approval. Because there may be greater uncertainty when using borrowed data, it may not carry the same weight as stand-alone pediatric studies.

Regardless of the method used, extrapolation will only be permitted when it can be done in a manner that supports reasonable, scientifically sound conclusions about medical device effectiveness and safety based on valid scientific evidence.

9. Statistical Methodology for Extrapolation

When the use of extrapolation is determined to be appropriate, a sponsor may have several options for how to extrapolate the adult data. Available options could depend on whether a prospective study of pediatric patients is needed and feasible, and/or whether sufficiently robust pediatric data can be obtained in other ways such as from prior studies run by the sponsor, studies in the literature, or pediatric registries.

Many of the methods available for borrowing strength across studies employ the Bayesian approach to statistics, which espouses learning from evidence as it accumulates. Bayesian statistics use Bayes' theorem to combine prior information with current information on a quantity of interest such as the primary endpoint. The idea is to consider the prior information and the current study results as part of a continuous data stream in which inferences are being updated each time new data become available. Prior information typically comes from results of previous comparable studies. Therefore, Bayesian methods are quite applicable for partial extrapolation from prior adult studies. Refer to FDA's "Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials,"[12] issued in 2010, for an introduction and more details on Bayesian statistics in medical device studies, including Bayesian hierarchical modeling, described briefly below.

[11] Available at http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM296379.pdf.

[12] Available at http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm071072.htm

9.1 The Bayesian Hierarchical Model and Exchangeability of Studies

Bayesian hierarchical modeling may allow an increase in effective sample size in a new study by "borrowing strength" (information) from prior studies. With a hierarchical model, as the differences among the study results decrease, more information is borrowed among studies, and a smaller sample size is needed for the new (pediatric) study. A typical hierarchical model might have two levels: a patient level and a study level. In a two-level structure, studies have different but related treatment effects (e.g., mean differences between treatment and control group) or mean outcomes. The relationship among the studies is referred to as "exchangeable studies", and has a mathematical definition described in more detail in the FDA Guidance document referenced above. Practically speaking, when two or more studies are exchangeable with one another, it means one could not distinguish the studies only by looking at the study results because there is nothing known a priori that would imply one study achieved a better average outcome from the device than any other study. For a two-level hierarchical model, study treatment effects or means are exchangeable, and patients are exchangeable within studies. It is important to note that patients are not assumed to be poolable across studies.

The assumption of exchangeability facilitates borrowing among studies in a hierarchical model. Statistically, exchangeability implies that the variability of responses within each study is comparable (similar magnitude) to the difference in responses among the studies. This assumption might not hold for extrapolation because adults and children could respond differently to a treatment, and so the responses among studies could be quite different than the responses within each study. If this is true, then a weaker form of exchangeability (partial exchangeablity, discussed in Section 9.2) may hold. Ultimately, the actual extent of borrowing will depend on the data within the model. Therefore, if the device effect is actually observed to differ between adult and pediatric studies, the studies will not borrow much from each other, and the extent of extrapolation will be limited.

In order to determine whether studies are likely to have exchangeable device effects, the FDA review staff and sponsors should identify differences in the studies that could hinder exchangeability. They should compare previous studies with the proposed study for similarity in relevant factors, including the following:

- Device used
- Patient population, including anthropometric measurements, when relevant
- Protocol
- Inclusion/exclusion criteria
- Prognostic factors
- Patient management

- Ability of the patients to comply with instructions for safe and effective device use
- Proximity in time
- Operator training/experience

Exchangeability is assessed by the clinicians and engineers from FDA as well as the sponsor. The sponsor should be prepared to discuss exchangeability or partial exchangeability among studies given covariates. Partial exchangeability may still hold even if differences in any of the above factors (or others) limits or precludes the assumption of unconditional exchangeability of adult studies with the proposed pediatric study. However, if the identified differences are known to be associated with one or more measured variables, and the measured variables have sufficient overlap between populations, adjustments can be made to a hierarchical model so that the studies might still be exchangeable after accounting for those variables. The next section provides an overview of one commonly used adjustment when the adult and pediatric studies have differences that affect the outcome of the study. Appendix B provides more statistical details as well as other adjustments.

9.2 Age-Related Covariates Associated With Device Outcomes

As mentioned above, there are likely to be one or more differences that could prevent the assumption of exchangeability between adult and pediatric studies. If these differences can be identified and measured, it is straightforward to account for them in a hierarchical model. When this is done, we can say that the studies are exchangeable, except for measured differences on certain variables. Often the differences will be related to the size or growth of the patient. A simple example might be a new limb prosthetic. The effectiveness and safety of the prosthetic might differ depending on the size or weight of the patient. However, within a given patient size (e.g., height), the performance characteristics might be the same, regardless of whether the patient is an adult or child.

It is imperative that FDA clinical reviewers and sponsors identify covariates that are associated with device performance and that might be responsible for any perceived differences in outcome for adults versus children or adolescents. A first step after identifying potential covariates associated with device performance is to determine how the covariate affects the primary outcome of the study, and then how age of the patient is related to the covariate. Identified covariates should have sufficient overlap between adult and pediatric populations so that the relationship between the covariate and age on study outcome can be connected across populations.

For example, a device whose effect is related to hormone level may have very different magnitudes of effect for adults than for children because they have different hormone levels. If patients are categorized into low, medium, and high hormone levels, then within each category, the adult studies might be exchangeable with the pediatric study. Presumably, if hormone level is highly associated with the effect of the device, the

915 sponsor is likely to have patient-level data on the level of the hormone in adults. Patient-
916 level information in children as well would enable the sponsor to construct a model that
917 relates hormone level to outcome, and thus condition on hormone level to assume
918 exchangeability between the adult and pediatric studies. That is, except for hormone
919 level, there are no known (and measured) differences between adults and children that
920 would allow one to identify an outcome as belonging to either an adult or pediatric
921 patient. If there were, then these measured covariates would also be added to the model.
922 The structure of the model would be agreed upon by both the sponsor and FDA.
923 Moreover, once data become available, the assumed model would be checked against the
924 data to ensure it is still valid.
925
926 When premarket pediatric data are needed, there are several suitable study designs and
927 analyses to consider, depending on circumstances related to the feasibility of collecting
928 the data. The "Draft Guidance for Industry, Clinical Investigators, and Food and Drug
929 Administration Staff—Design Considerations for Pivotal Clinical Investigations for
930 Medical Devices,"[13] issued in 2013, discusses several concepts and principles related to
931 designing medical device studies.
932
933

[13] Available at http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm373750.htm This draft guidance, when finalized, will represent FDA's current thinking on this topic.

Appendix A. Examples of the Decision Process for Extrapolation

The examples in this section are intended to demonstrate the use of the Pediatric Extrapolation Decision Tree. The examples are not predictive of FDA decisions but may be considered guides for how FDA weighs the appropriateness of extrapolating existing clinical data to support pediatric indications.

A.1 A Hypothetical Example of Full Extrapolation for Effectiveness

A gel that is used as a pleural air leak sealant is proposed to be indicated for a pediatric population (aged 2–21). The gel is currently approved for adults in the closure of remaining visible air leaks incurred during open resection of lung parenchyma, after standard sutures have been applied; the same condition can occur in pediatrics. Suppose that extrapolation of effectiveness from adults to pediatrics is under consideration. The measure of device effectiveness used to gain approval was that the patient remained free of air leaks 1 month post-surgery, after application of the gel. The same measure would be used for pediatrics. Therefore, the first two questions in the decision tree (A. and B.) are answered yes.

The gel is intended to be applied in the body in the same location for both age groups, for roughly the same duration (eventually the gel gets resorbed and excreted). Furthermore, the gel itself does not have different characteristics for adults than for children. With respect to the purpose of the gel, the disease characteristics (air leaks) are similar for both adult and pediatric patients. However, the size of the air leak and therefore the amount of gel used and perhaps the size of the syringe to deploy the gel could differ between adult and pediatric patients. In this example, the Agency has determined that these differences do not impact device effectiveness in the pediatric population in a clinically meaningful way. The gel has been demonstrated to be equally effective when covering smaller areas as larger areas, and the size of the syringe is not relevant to effectiveness. Therefore, the answers to Box C were all “NO”, and full extrapolation of effectiveness data could be considered for this device. In this case, the FDA might decide that adult effectiveness data could be substituted for prospective pediatric study (i.e., full extrapolation) if the adult studies are of sufficient quality.

In separately assessing whether the existing data could be extrapolated to demonstrate safety in the pediatric population, the potential for adhesions was felt to be of concern due to the expected needs for reoperation in this population, based on the preclinical testing results. For this reason, safety extrapolation was not performed and a separate study for safety in pediatrics was recommended.

A.2 A Hypothetical Example of Partial Extrapolation with Relevant Age-Associated Differences between Populations Accounted for Via Modeling

A diagnostic device is approved in adults as an aid to diagnosing a particular disease or condition through the quantitative measurement of a particular measurand. This measurand is the same one used to diagnose both adults and children. In the adult study, the device was compared to the currently used diagnostic test, which is generally considered a reference standard method, to provide reasonable assurance of safety and effectiveness. This reference standard method requires the collection of a large amount of blood. An indication is sought for pediatric patients as young as 2 years old. Use of the reference standard method as the comparator for the pediatric population was considered an unsafe option, due to the need to collect large amounts of blood from young children.

When referencing the flow chart to decide whether or not extrapolation is appropriate, it is apparent that the condition occurs in both adults and pediatrics and that there is an endpoint that is relevant to both populations. It is not known whether the values obtained from the comparator reference standard are the same between adult and pediatric patients. Because these values could differ, the difference in results between the device and the comparator method may have a different magnitude for adults than for children. Accordingly, the difference in blood volume that precludes use of the reference standard method as the comparator is a unique characteristic of the intended pediatric population, which could have a clinically meaningful impact on the safety or effectiveness of the device. Specifically, the difference in the use of the reference standard may change the diagnostic result which, if erroneous, could impact patient safety. Therefore, full extrapolation is not appropriate. We thus proceed to consider whether partial extrapolation is appropriate.

In this example, the device characteristics, device matrix, and interfering substances are considered the same for the adult and pediatric population. It is also known that the reference standard values expected for adults and children can be calibrated to be comparable by accounting for body size, among other measured patient-level variables that may be correlated with age. Because calibration using measured variables is possible, the Agency and sponsor agree that the adult reference standard data can be borrowed statistically to bolster the expected reference information in pediatrics. To the extent that the calibrated reference standard values are similar between the adult and pediatric populations, more adult data can be borrowed. Therefore, because the observed differences between the adult and pediatric populations can be accounted for in a statistical model, the extrapolated data may be used in support of demonstrating a reasonable assurance of the safety and effectiveness of the device. Because the data from the adult population was of high quality in terms of study design, these data are

considered a viable candidate for partial extrapolation.

This example highlights that borrowing from adult data can be done not only for the device group in a clinical study, but also for control groups or reference standard values. In many cases, a control or comparator is not available for pediatrics but it is available for adults. As illustrated, partial extrapolation can potentially be used in these cases.

A.3 A Hypothetical Example of Partial Extrapolation

Suppose a company wishes to extend an indication to adolescents for their marketed device X to treat a condition Y. The device is approved for use in adults. There are several available adult data sets from the US pre-market application as well as from marketing applications in other regions of the world. The endpoint used in the available studies is identical to the endpoint desired in the adolescent population. However, the adults were followed for eight months, and the FDA recommends following adolescents for at least 12 months. There are no other identified differences between populations with respect to the anticipated effectiveness or safety of the device. Thus, Box C has been answered "no". However, full extrapolation is not recommended because the eight-month adult data are not sufficient to serve as a substitute for twelve-month pediatric data.

Based on additional information from studies published in medical journals about how the device performs beyond eight months in adults, the sponsor was able to borrow from the adult studies and use statistical modeling to predict adolescent response at 12 months. The predictive model also incorporated some prospectively collected adolescent data out to 12 months. Thus, the data quality when paired with the statistical model was determined to be sufficient to allow for partial extrapolation. However, with the leveraged adult data the sample size estimated for the adolescent study was smaller than it would have been otherwise. Once the adolescent study is completed, the model will be verified to ensure that assumptions are met and borrowing is indeed appropriate.

A.4 Hypothetical Examples where Extrapolation is not Recommended

A.4.1 Hypothetical Example where Extrapolation is not Recommended because of Quality of Data

A pre-amendment device is not indicated for pediatric use. When submitting their device's annual report to FDA, the sponsor cites case report studies which the sponsor believes suggest an indication for pediatric use may be appropriate. The disease to be treated is essentially the same in adults and pediatric patients, and the endpoints used to evaluate clinical outcomes are also the same. There are also no apparent expected differences between the pediatric and adult response to device. Therefore, the answer to Box C is "no."

However, the adult data available for extrapolation are decades old. Both the practice of medicine and relevant study design considerations have significantly changed. As such, despite the similarities between the adult and pediatric populations, it is likely that FDA would determine that the adult data in this case are not of sufficient quality for either full or partial extrapolation.

A.4.2 Hypothetical Example where Extrapolation is not recommended because of Relevant Differences

Consider a generic device which is indicated for a rare adult disease. A sponsor would like to extend the indication to a pediatric subpopulation because the endpoints between the two groups are similar. The only difference in the response to device intervention relates to how pediatric growth may impact the safety and effectiveness of the device. However, the device may need to be removed or adjusted for growth, which requires surgical intervention and introduces additional risk for pediatric patients. In addition, the anticipated impact of pediatric growth on device safety and effectiveness is largely unknown, and there is limited clinical experience in adults so the data are not sufficient to reliably inform modeling. Partial extrapolation is not feasible because the differences between the adult and pediatric populations cannot be accounted for, clinically or with modeling. Therefore, extrapolation is not recommended in this scenario.

A.5 An Example of an Actual Extrapolation

Patients with systemic, left-sided, congenital heart valve disease pose significant challenges for physicians. There are limited technological solutions available for these patients. Few replacement heart valves are indicated for pediatric patients, and commercially available bioprosthetic valves for aortic and mitral valve replacement may not be available in sizes appropriate for infants and children.

The clinical impact of congenitally deformed valves is significant and often lifelong. Treatment decisions are almost always impacted by the effects of rapid growth, active lifestyle, and accelerated deterioration of biological prostheses. Pediatric valve replacement is a high-risk procedure involving higher operative mortality, high reoperation rate, and late morbidity compared to adult patients undergoing the same operation.

The reasons for the higher operative mortality are multiple and complex. Most often, the available prosthesis is too large for the child's anatomy, resulting in delay in referral for surgery. When surgery is undertaken, additional steps are often required to enlarge the site of implantation to accommodate the prosthesis.

Clinical studies have routinely been conducted on the adult patient population. However, pediatric patients have typically been excluded from replacement heart valve trials for several reasons, including:

- Limited patient pool requiring a replacement heart valve, which can lead to prolonged recruitment to achieve required enrollment numbers
- Complex health histories (many leading to early mortality)
- Co-morbidities confounding the adverse event profiles for the study, making it very difficult to assess overall safety of the valve
- Limited valve sizes available
- Following valve replacement, the pediatric patient continues to grow, ultimately necessitating reoperation and the placement of a larger valve
- Uniformity of an identifiable patient population is extremely challenging to achieve, again leading to prolonged study recruitment

Below we trace the pathway to extrapolation of adult data using the decision tree, beginning with effectiveness. First, the disease is identified as being the same for pediatrics as for adults. Additionally, the primary endpoint for effectiveness would be similar in both a pediatric study and adult study (Decision Tree Questions A and B). Therefore, the adult data is considered relevant.

While a heart valve for a pediatric patient is implanted in the same location as for an adult, the duration of implantation of a particular size will be shorter for a pediatric patient due to normal pediatric patient growth. This could influence the effectiveness of the device for pediatric use. Therefore, the answer to question #1 in Box C is "yes". Furthermore, one of the most important patient characteristics unique to pediatrics is that the patient continues to grow after valve replacement, necessitating additional operations to implant larger valves. This difference can also influence effectiveness. Question #3 is also answered yes because pediatric patient growth could impact effectiveness of the heart valve. Therefore, there are some differences related to effectiveness.

However, despite the various differences that could influence effectiveness, these can be explained clinically as associated with valve size rather than age per se. Additionally, there is extensive relevant adult data of sufficient quality available for the sizes of interest and the different positions (aortic, mitral) to inform a statistical model to account for this. It was thus possible to incorporate a clinical relationship between valve size, position and device effectiveness into the statistical model used for extrapolation, which could be used to fairly and responsibly support demonstration of a reasonable assurance of effectiveness of the device in the pediatric population. Therefore, a partial extrapolation was considered plausible for effectiveness. FDA agreed that a sample size of 15 pediatric patients per size per position (aortic, mitral), when combined with the borrowed adult data, could potentially suffice for demonstrating clinical effectiveness of the device for the proposed pediatric indication.

In assessing whether the existing data could be leveraged to extrapolate for safety, the primary difference with pediatric device use is that patient growth after valve

replacement necessitates additional operations to implant larger valves. As such, the answer to question #2 in Box C would be "yes". This exposes pediatric patients to additional operations, which pose an incremental risk. Therefore, safety data adequate to evaluate this incremental risk for pediatric patients was necessary. FDA concluded that the number of pediatric patients that would be prospectively enrolled to confirm effectiveness would be sufficient to evaluate safety as well. In addition, a post-approval study was recommended to assess the long-term safety and effectiveness of the device in pediatric patients.

This example illustrates how available relevant adult clinical data were leveraged to bolster new pediatric data in a manner that constitutes valid scientific evidence. When considered alongside other forms of scientific evidence from assessments of safe device functioning (e.g., preclinical testing, engineering models, biocompatibility, etc.), appropriate partial extrapolation was used to support demonstration of safety and effectiveness of new pediatric heart valves.

Appendix B: Details on Statistical Modeling for Extrapolation

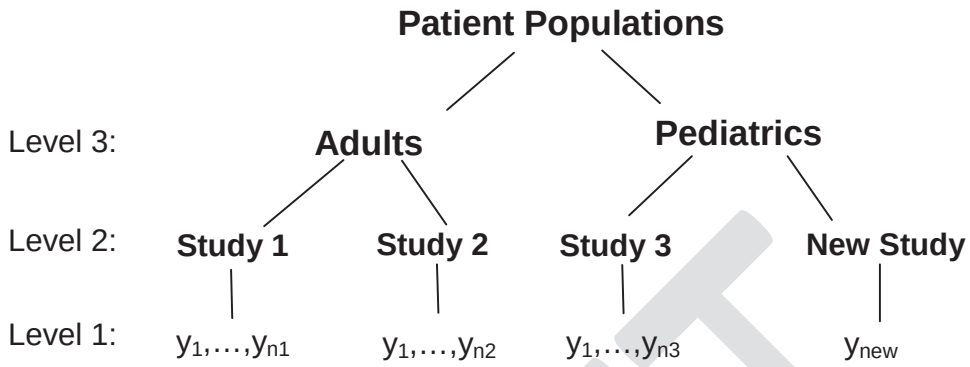
In this appendix, we present further details of statistical modeling that might be performed for partial extrapolation. As described in the text, a goal for partial extrapolation is to borrow strength or information from adult data while still accounting for the important differences between adult and pediatric populations. Accounting for baseline characteristics is a common way to distinguish studies that should not be considered exchangeable. This technique was described above. However, if there are multiple adult studies from which to borrow, then placing another level in the two-level hierarchy to include subgroups of studies might further temper borrowing between adult and pediatric studies when they should not be considered on the same level.

We introduce a simple three-level hierarchical model, followed by an overview of other possible methods for borrowing strength along with pros and cons of the methods.

B.1 A Three-Level Hierarchical Model

In the proposed three-level hierarchical model (see Figure 2), the third level involves the two patient populations (adults and children), each having studies that are exchangeable with one another. The adult studies are exchangeable among themselves, and the pediatric studies are exchangeable among themselves. To facilitate borrowing between the adult and pediatric studies, they are connected by assuming exchangeability between the two patient populations regarding the device effect on the endpoint of interest. That is, prior to knowing anything about what type of effect a device will have, it is presumed that if there is evidence of the effect of the device on a population, it would not be possible to tell which population it was, adult or pediatric.

Figure 2. Three-Level Hierarchical Model Structure Example: Studies Within Patient Populations Have Different But Related Effects



Level 1: Patients (y) exchangeable within studies
Level 2: Studies exchangeable within patient populations
Level 3: Patient populations are exchangeable

In Figure 2, patients are represented by their values (y) on the endpoint of interest. There are n_1 patients in Study 1, n_2 in Study 2, and n_3 in Study 3. (For simplicity, the figure represents single-arm studies). The adult population produced Studies 1 and 2, and the pediatric population produced Study 3. The two patient populations are assumed to come from a common superpopulation of patient populations. The figure also includes a branch for a future pediatric patient from the pediatric population. With Bayesian hierarchical models, not only is it possible to borrow strength to estimate individual study means and their population means, but it is also possible to estimate a predicted value for a new pediatric patient from the pediatric population, using the Bayesian predictive distribution. The Bayesian predictive distribution is the distribution of an unknown outcome, which can potentially be observed in the future. It is essentially the posterior distribution of a yet to be observed outcome (Carlin & Louis, 2009).

B.2 Age-Related Covariates Associated With the Device Effect or Outcome

Figure 2 above is highly simplified because it assumes no differences across patient populations that would affect the safety or effectiveness of the device. As with the two-level model, in practice, there are likely to be one or more differences that could prevent the assumption of exchangeability between adult and pediatric populations (the third level in the hierarchy). If these differences can be identified and measured, it is straightforward to account for them in the model. Essentially, the model will dictate that the populations are exchangeable, except for measured differences on certain variables. Differences could be static or dynamic (time-varying) over the trial period. Often the differences will be related to the size or growth of the patient. The structure of the model should be agreed upon by both the sponsor and FDA. Moreover, once data become

available, the assumed model would be checked against the data to ensure it is still valid. Section 9.2 also discusses accounting for covariates.

B.3 Extrapolation From a Single Adult Study

When extrapolating from adult studies, it is advantageous to have several prior studies to use in an analysis to facilitate more precise estimation of the device effect in pediatrics. However, it is often the case that only a single prior adult study exists. Although the example above described borrowing from two adult studies, similar methodology can be used when there is a single prior adult study available. FDA's "Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials" (2010) discusses limitations with the use of Bayesian hierarchical models with a single prior study.

Several authors have developed methods for incorporating a single historical study as prior information in a Bayesian model, where the weight placed on the historical study varies with the similarity of the historical study and the current study data as they are collected (e.g., Hobbs et al., 2011, 2012). Some of these methods have similar limitations with hierarchical models in that fairly informative priors must be used to describe the relationship between the historical and current studies. However, the specification of the priors might be conceptually easier than with a hierarchical model.

In limited cases it might be reasonable to prespecify, as a percentage, the amount of borrowing from the prior adult data set(s). The method of power priors (Ibrahim & Chen, 2000) uses a prior that is constructed from the likelihood of the prior data raised to a power, where the power falls between 0 and 1. The power indicates the downweighting of the prior data, so that a power of 0.5 implies that 50% of the information from the prior likelihood is borrowed. Unfortunately, when the power must be fixed in advance, it cannot change based on later observed data from a new pediatric trial. Placing a prior on the power parameter itself, thereby potentially allowing the data to determine the amount of borrowing, has been shown in practice to have limited success (see, for example, the discussion in Hobbs et al., 2011).

B.4 Additional Methods for Extrapolation

While Bayesian methods are described in this document, non-Bayesian methods can also be used for borrowing strength. The structure of the hierarchical model is not inherently Bayesian, and it can be used without the interpretation of posterior probability. However, in many cases the overall conclusions will remain the same, and the Bayesian interpretation of posterior probability is often simpler to understand.

As mentioned above, the Bayesian hierarchical model can be difficult to use when there is only one observed prior adult study. The between-study variance either must be prespecified (and just like with the prespecified power parameter for the power prior, it cannot be changed once the pediatric trial is run), or an informative prior must be placed

on the between-study variance, potentially limiting the range of values it can realize once the pediatric study is run.

In addition to hierarchical models, one could use propensity score methods for extrapolation from adult data (Rosenbaum & Rubin, 1983; Yue, 2007, 2012). A propensity score for a subject is the probability of the subject being assigned to the device group in a medical device clinical trial, rather than to the control group, conditional on a set of measured baseline covariates (but not on the measured outcome variable). In a randomized trial, with 1:1 randomization, this probability is by definition 0.5, independent of any covariates. In a nonrandomized study, the probability often depends on observed covariates. If it depends only on observed covariates, then for the same values on those covariates, two subjects have the same probability of being assigned to the device group. For a set of subjects with the same probability of receiving the device over the control, an estimate of the treatment effect will be unbiased, just as it would be in a randomized trial. Accounting for the propensity score in a regression model or matched analysis can then yield an overall estimate of the device effect that is unbiased despite the trial being nonrandomized.

If adult data are available from a previous trial, adult subjects could be grouped with pediatric subjects based on their propensity scores (say, in quintiles). Those subjects with the same propensity score quintile would be compared (device versus control) to obtain a device effect within each propensity score grouping. An overall estimate of the device effect can be obtained using regression adjustment. In general, this adjustment is similar to that described in Sections 9.2 and B.3 (Age-Related Covariates Associated With Device Effect). However, the propensity score is a single representation of all measured baseline covariates. The single-dimensional representation makes it easy to use in modeling, but the form of the model might be difficult to determine from a summary measure rather than from individual covariates. Moreover, there is no simple way to account for variability across studies that the hierarchical model can incorporate.

References

American Academy of Pediatrics, American College of Obstetrics and Gynecology. (2007). Appendix D: Standard terminology for reporting reproductive health statistics. In *Guidelines for perinatal care*, 6th. ed. (pp. 389–404). Elk Grove Village, IL: American Academy of Pediatrics.

Carlin, B., & Louis, T. (2009). *Bayesian methods for data analysis*. Boca Raton, FL: CRC Press.

Hobbs, B. P., Carlin, B. P., Mandekar, S. J., & Sargent, D. J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics, 67,* 1047–1056.

Hobbs, B., Sargent, D., & Carlin, B. (2012). Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis, 7*, 1–36.

Ibrahim, J. G., & Chen, M. H. (2000). Power prior distributions for regression models. *Statistical Science, 15*(1), 46–60.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70,* 41–55.

Yue, L. Q. (2012). Regulatory considerations in the design of comparative observational studies using propensity scores. *Journal of Biopharmaceutical Statistics, 22*(6), 1272–1279(8).

Yue, L. Q. (2007). Statistical and regulatory issues with the application of propensity score analysis to nonrandomized medical device clinical studies. *Journal of Biopharmaceutical Statistics, 17*, 1–13.