Optimizing the Use of Postapproval Pregnancy Safety Studies

Hybrid Public Workshop

National Press Club, 529 14th St. NW, Washington, DC 20045

September 18-19, 2023

Table of Contents

This hybrid public workshop was supported by the Food and Drug Administration (FDA) of the U.S. Department of Health and Human Services (HHS) as part of a financial assistance award [U19FD006602] totaling $4,241,714 with 100 percent funded by FDA/HHS. The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement, by FDA/HHS, or the U.S. Government.

Welcome and Opening Remarks

Marianne Hamilton Lopez:

All right. Good morning and welcome to our hybrid public workshop, Optimizing the Use of Postapproval Pregnancy Safety Studies. I'm Marianne Hamilton Lopez, senior research director of Biomedical Innovation at the Duke Margolis Center for Health Policy at Duke University. I have the pleasure of advancing my own slides, doing this welcome, and also trying to navigate between hybrid world when I've been on Zoom for the last few years. So we're all in this together. And for those of us on the screen, good morning. This workshop is supported by the Food and Drug Administration under a cooperative agreement with Duke Margolis and will fulfill a Prescription Drug User Fee Act (PDUFA) VII commitment. PDUFA VII includes an FDA commitment to develop a framework to identify and meet the regulatory needs of identifying optimal studies to obtain safety data in the postapproval setting in pregnant individuals when there's little to no existing data.

Today's public workshop is being held to introduce a draft framework and solicit public inputs on postapproval pregnancy safety study design, types, challenges, and needs. We recognize that there are other important clinical needs for pregnant women, maternal, fetal, and infant outcomes, and have looked to incorporate those issues in various day one workshop sessions. Day two focuses specifically on the draft framework, looking to identify optimal studies to obtain safety data and pregnancy data postapproval. We invite your feedback and comments during this meeting and we'll remind you often of our hope that you submit comments on the draft framework to the docket prior to its closing day of November 30th, 2023.

Okay. This is the first iteration of the FDA's draft framework. Noting the draft nature, FDA colleagues may be limited in their ability to respond in detail. But we'll consider feedback to inform framework updates. Thank you again in advance for your active participation. So on this slide here, I'm going to review the panel discussions and presentations teed up in the agenda. We'll begin today with an opening presentation from Leyla Sahin from the FDA. The presentation will provide an overview of how work on postapproval pregnancy safety studies have progressed since the 2014 FDA public meeting on Study Approaches and Methods to Evaluate the Safety of Drugs and Biological Products during Pregnancy in the Postapproval Setting.

In session one, panelists representing a range of stakeholders will share their views on the impact of and considerations for pregnancy safety studies. The panel will discuss the balance between benefit risk when utilizing or prescribing drugs to pregnant individuals, especially those that have little safety data and how to proceed in situations with minimum data and evidence. Following a 15-minute break, we'll return for session two. In this session, drawing from their experience and practice in the field, we'll hear from researchers that have conducted a variety of postapproval pregnancy safety studies to learn about past research efforts and future directions in addressing challenges in conducting these types of studies.

We will break for lunch, and before returning for session three where FDA will present postapproval pregnancy safety studies used to approve FDA products, data used in PLLR product labeling and drug utilization data to inform the development of the pregnancy safety framework, that is presented in session four. And then after a quick 10-minute break, we will return for session four, where the pregnancy safety study framework will be presented by Dr. Wei Hua, followed by a moderated discussion with FDA staff that took part in developing the framework. We'll then wrap and adjourn day one of this workshop around 4:30.

Let me turn then to our statement of independence and some final logistical reminders before we get the meeting started. The Margolis Center for Health Policy is part of Duke University, and as such, we

honor the tradition of academic independence on behalf of its faculty and scholars. Neither Duke nor the Margolis Center take partisan positions. But individual members are free to speak their minds and express their opinion on important issues. I'll also add that this conversation, while supported through a cooperative agreement with the FDA, is not a federal advisory committee so we will not be voting or making recommendations. We have one more slide before getting to the good stuff.

As this is a hybrid public workshop with both in-person and virtual attendees, we will not be moderating the Q and A in Zoom since many of our in-person attendees may not have access to the Zoom Q and A. We'll be using the Q and A platform called Slido for live questions and comments. If you would like to ask a question during the meeting, you can use this platform to submit questions and comments. Your submissions will be monitored by our team and passed along to the moderators to be included in discussion sessions, time permitting. We suggest that you take a moment to log into the Slido by scanning the QR codes or navigating to slido.com and iterate the code pregsafe. This applies to both in-person and virtual participants.

Again, please note, we are not going to be monitoring the Q and A function on Zoom, so please use Slido for all questions and comments. Finally, a reminder that the meeting materials, including the agenda and speaker bios are available on the Duke Margolis event website. For those in person, you can access them using the QR codes you received at registration. We will also be recording this meeting and posting the recordings on the Duke Margolis website for future reference. Now it is my pleasure to introduce you to our presenter on the role of postapproval safety studies, Leyla Sahin.

Leyla Sahin:

Good morning, everybody. Well, it's great to see all of you in person and it's just amazing how many people have signed up for this workshop. Over almost 1,000 people have signed up. So hello to all of you online as well. I know how hard it is to be here in person now that we have the capability to have these meetings where you can join online, so I really appreciate everybody who traveled to make it here in person, all this way. Okay, great. All right. So since I'm the first FDA speaker, I'd like to say that on behalf of all the FDA speakers, we don't have any financial disclosures to report. The FDA presentations reflect the views of the speaker and should not be construed to represent FDA's views or policies.

The objectives of my talk are to provide some background information on why it's important to get pregnancy safety studies in the postapproval setting. I'm going to provide a high-level overview of the 2019 Postapproval Pregnancy Safety Studies Draft Guidance. I'm going to talk a little bit about FDA's efforts to advance safety data collection in pregnant individuals, and then I'm going to introduce the PDUFA VII pregnancy safety commitments. So as introduction, I'd like to take us back to 2014 when we held the FDA public meeting on Study Approaches and Methods to Evaluate the Safety of Drugs and Biologics during Pregnancy in the Post-approval Setting.

So where were we in 2014 and where are we today? In 2014, pregnancy registries were the primary method to collect pregnancy safety data in the postapproval setting. At the 2014 public meeting, we heard from our stakeholders, from all of you, that database studies should be used to complement the data collected from pregnancy registries. We really appreciate all of the input from all of you, from all of our stakeholders and all of the expertise that you provided. We appreciate the many clinicians and epidemiologists, several of whom are here at this meeting and others who have presented a past FDA meetings. And I want to mention a few of them because of the impact that they've had. Lou Holmes from Massachusetts General Hospital, Allen Mitchell from Boston University, and Elyce Cardonick from the Cancer and Pregnancy Registry. Yes, unfortunately pregnant women do get cancer. And I see Lee Cohen here in the audience too. So we acknowledge you as well, Lee.

You've all dedicated your careers to advancing the understanding of the safety of medication used during pregnancy. So as a follow-up to that 2014 public meeting, we then published the 2019 Postapproval Pregnancy Safety Studies Guidance which expanded the scope to include database studies. So what else happened since 2014? Well, we all experienced and lived through a global pandemic, and that pandemic was really helpful in highlighting the need to get pregnancy safety data in medications and vaccines. So as FDA was authorizing vaccines and therapeutic products under emergency use authorization, the public was wondering whether these products were safe to use in pregnancy.

So what else has happened since 2014? Well, to advance pregnancy safety data collection in pregnant individuals, FDA has been involved in national and international collaborations. The Sentinel Initiative has been expanding its capabilities and FDA has committed to PDUFA commitments related to pregnancy safety. So why are we all here today? Why are we all here for this meeting? Well, the reason that we're all here is for the 5.5 million individuals who get pregnant in the US every year who have to decide if it's safe to use their medication or vaccine. Pregnant individuals may need treatment for chronic conditions like depression or inflammatory bowel disease, or they may develop acute conditions in pregnancy like nausea and vomiting in pregnancy. And they should not have to be in the position of having to decide whether to take their medication or their vaccine without knowing if it's safe for them or their baby.

The current regulatory environment is that pregnant individuals have historically been left out of drug development trials. Which means that the default position is that most drugs are approved with only nonclinical reproductive toxicology data. Which is why it's so important to get human safety data to inform labeling and to inform clinical care. And this happens in the postapproval setting. Pregnancy safety studies can be required under section 505(o)(3) of the FD&C Act. It's important to note that lack of a safety signal in animal data doesn't mean that it's safe to use that medication in pregnancy.

So as a public health agency that approves drugs and vaccines based on safety and efficacy data, we view the lack of pregnancy safety data as a safety issue. Under 21 CFR 312.32(a), congenital malformations due to drug exposure in pregnancy are serious adverse events. Historically, pregnancy registries have been issued as postmarketing requirements or commitments. More recently, two types of pregnancy postmarketing requirements, a pregnancy registry, and a database study, to complement the pregnancy registry have been issued in CDER.

To get a better understanding of trends in postmarketing requirements, we had an ORISE fellow do a review of pregnancy and lactation postmarketing studies that were required by the FDA. From 2007, when FDA was granted authority to require postmarketing requirements through the end of 2020, we see an increase in postmarketing requirements in pregnancy and lactation that have been issued starting within a few years after the 2014 public meeting. However, it's also important to note that of all new drug approvals that may be used in females of reproductive potential, only 16% of those new drug approvals received a postmarketing requirement in these populations. And as I note on this slide, this may partly be accounted for by the fact that there are existing disease-based pregnancy registries rather than PMRs in certain therapeutic areas such as antiretrovirals, antiepileptics, and psychiatric drugs.

We published the Postapproval Pregnancy Safety Studies Draft Guidance in 2019 and it's currently undergoing revision based on public comments that we received. The draft guidance highlights that all these data streams are really, really important in assessing the safety in pregnancy once drugs are newly approved in the postmarketing setting. And this includes pharmacovigilance based on case reports and case series, pregnancy registries, database studies relying on electronic healthcare data, and descriptive pregnancy safety studies.

So I want to say a little bit about pharmacovigilance. So we all know the limitations of pharmacovigilance. However, it's also important to recognize that pharmacovigilance has contributed to

our knowledge of what we currently know as major teratogens. For example, thalidomide, isotretinoin, mycophenolate, these were all based on clinicians reporting specific malformations or distinctive patterns of malformations and reporting these cases in the medical literature.

Pregnancy registries are prospective observational cohort studies where pregnant individuals are enrolled and then followed until the outcome of the pregnancy occurs. There's also a disease matched comparator cohort of individuals who were unexposed to the drug of interest who were enrolled concurrently. Infants are followed up to at least a year of age. Outcomes that are assessed, include major malformations, patterns of malformations, miscarriage, pregnancy termination, still birth, preterm birth, small for gestational age, et cetera.

Advantages of pregnancy registries are that they can provide real-time prospective data collection and can serve as an early warning system for safety signals in newly approved drugs. Clinical data are obtained from the pregnant individual who can confirm that the drug was actually taken. The pregnant individual can also confirm when the drug was taken in terms of gestational timing of exposure. And this is really, really important because we know that the etiologically relevant window of organ development is week three to week eight post conception. And so this is really important information in terms of understanding exposure status when a birth defect is identified. The pregnant individual can also confirm the dose of the medication that she took and the duration.

The pregnant individual can also provide covariate data such as smoking information, alcohol, drug use, et cetera, and can also provide all the rich clinical information like past obstetrical history, family history of birth defects and genetic syndromes, et cetera, all of which are really helpful in providing context when a birth defect is identified. Clinical data from medical records are obtained from the obstetrician, the neonatologist, and pediatrician, and the medical specialist treating the condition. Pregnancy registries can capture non-live birth outcomes like miscarriage, pregnancy termination which may be done for a fetal malformation and stillbirth.

The medical records of the infants with birth defects are reviewed by clinical experts, thus allowing for clinical input and clinical judgment. And some registries have dedicated experts that actually assess all the newborns. The known limitations of pregnancy registries are that they have small sample size due to challenges with enrollment, and they take a long time to conduct. There may be selection bias because of the known demographics of patients who typically enroll in these studies. Database studies are generally retrospective cohort studies using claims data. These studies require mother-infant linkage. There's the ability to develop several disease matched cohorts using comparator populations of individuals who are unexposed to the drug of interest and who also may be exposed to comparator drugs. There's validation work that needs to be done for the algorithms that estimate the pregnancy start date and for the positive predictive value of the ICD codes.

The advantages of database studies are that they have the potential to have a larger sample size and greater power and they can potentially assess specific malformations rather than a composite outcome of major malformations overall, which is typically what pregnancy registries do. Database studies can be conducted faster potentially and there are no recruitment or enrollment challenges. Limitations of database studies include the following: the exposure and the timing of exposure cannot be confirmed because these are based on pharmacy dispensing. This is particularly problematic for drugs that are used on an as needed basis such as acute migraine medications, or pain medications, or acute asthma medications. There's potential exposure misclassification because the estimates are based on algorithms. There's potential outcome misclassification because the outcomes are based on ICD codes.

Non-live births are poorly captured because miscarriages and pregnancy terminations may be occurring outside the healthcare system without a claim. And some covariates are not well captured like obesity, smoking, alcohol use, et cetera.

The guidance discusses considerations for when exposure in pregnancy is expected to be uncommon. In these situations, an adequately powered pregnancy registry or database study may not be feasible.

For example, in the situation of a new drug that's approved to treat a rare disease or in the situation of a drug where there's strong labeling that says not to use in pregnancy or contraindicate its use in pregnancy based on the animal reprotox studies, or because of the pharmacologic class that's associated with teratogenic effects. In these situations, there may be a potential role for a descriptive pregnancy safety study. The previous terminology that's used in the current 2019 guidance is pregnancy surveillance program. But that caused some confusion so we've moved to using the terminology descriptive pregnancy safety study. This is basically a systematic collection of pregnancy specific data. It includes prospective and retrospective data collection, and it may be part of an existing disease registry like a rare disease registry.

FDA has been involved in various efforts nationally and internationally to advance pregnancy safety data collection. FDA's part of the task force on research specific to pregnant women and lactating women, also referred to as PRGLAC. This task force was required under the 21st Century Cures Act of 2016, and the objectives are to identify and address gaps in knowledge and research regarding safe and effective therapies for pregnant women and lactating women. The task force was made up of representatives across various federal agencies, industry, academia, patient groups, and professional organizations. Reports and recommendations were submitted to HHS in 2018 and then an implementation report was submitted in 2020.

There's currently an oversight committee that is being formed to monitor the implementation of the recommendations. Some of the PRGLAC recommendations related to pregnancy safety include the following: increase the quantity, quality, and timeliness of research involving therapeutic products used by pregnant women. So we can't continue to wait 10 or 15 years to get data in pregnancy. The next one is for industry. Sponsors should implement a proactive approach to data collection, and study design, and develop a systematic plan for collection of pregnancy safety data.

Pregnancy registries need to be optimized by expanding the use of disease-based pregnancy registries, where we know based on experience that these have a better track record of increased awareness and increased data collection. And the only way to do this is through sponsor collaboration. Facilitate access to data and transparency of pregnancy registries, similar to what the Antiretroviral Pregnancy Registry does by publishing their results every six months. Use and improve existing resources such as mother baby linkages and existing large databases.

FDA has also been involved in global efforts to advance pregnancy safety data collection through the formation recently of a pregnancy and lactation cluster where we meet regularly with EMA, MHRA, and Health Canada to discuss scientific and regulatory issues related to these populations.

I'd now like to introduce the PDUFA VII commitments. FDA has committed to the following pregnancy safety commitments. FDA will develop a framework describing how data from different types of postmarketing pregnancy safety studies might optimally be used, incorporating past knowledge of how different types of these studies have been used by FDA and industry in identifying gaps in knowledge needed to be filled by demonstration projects. The framework would specifically address the use of pregnancy registries and electronic healthcare data sources including Sentinel, with the goal of ensuring the most efficient means of obtaining the highest quality safety data available. FDA has committed to conducting five demonstration projects and publishing a MAPP or a guidance to be developed through fiscal year 2027.

So in summary, the lack of safety data in pregnant individuals is a public health issue that affects all of us. We really need to do a better job to get timely and high quality safety data in pregnancy in the postapproval setting. And the only way that we can advance this field is for all of us to collaborate. So

the purpose of this meeting is to seek input from all of you, from all of our stakeholders, on the development of a framework to optimize postapproval pregnancy safety studies and on the proposed demonstration projects to fill in the data gaps. And that's the end of my presentation. Thank you very much for your attention.

Session 1

Megan Clowse:

Good morning. I'm Megan Clowse, and I'm a rheumatologist at Duke University who has spent the last 20 years focused on pregnancy in women with rheumatic disease, and I'm so pleased to be here and see the interest in this topic. So session one is going to be stakeholder perspectives on the impact of post-approval pregnancy safety studies. We're going to focus on the impact for real physicians and real women who are making decisions on a daily basis about medications to be used during pregnancy.

First, we're going to hear from Mariah Leach, who is a patient who has had to struggle with these decisions herself. She also is the founder of Mamas Facing Forward, which is both an informational website, but also really a social media platform that allows women with chronic disease to really share their stories.

Next, we're hearing from Keele Wurst, who is in the industry. She is the Head of Safety Science Epidemiology at GlaxoSmithKline and the co-chair of GSK's Pregnancy Outcome Advisory Panel and has been deeply in this work for a long time.

Finally, we're going to hear from two separate physicians. The first is Katherine Wisner, a Perinatal Psychiatrist who directs the Asher Center for the Study and Treatment of Depression Disorders at Northwestern University Feinberg School of Medicine.

And finally, Geeta Swamy, who is a maternal fetal medicine expert, as well as an expert in infectious disease and vaccination in the setting of pregnancy and she has led numerous vaccination trials.

I'm going to remind everybody that if you have questions that come up, please put them into Slido so that we can have a lively discussion at the end. And now we're going to turn to Mariah to kick us off. Thanks.

Mariah Leach:

Hi, everybody. Excuse me. I'm Mariah Leach, and as you mentioned, I am a patient. I'm here to remind us that basically everything we're talking about today impacts real people with real families and real wishes to grow families. And so, I'm going to share a little bit about my own experiences with pregnancy and safety data because it changed quite dramatically across my three pregnancies.

So my first son is 11, and I was diagnosed with rheumatoid arthritis when I was 25 years old. So one of my first concerns was, "How do I start the family that I always knew I wanted to have?" And so, my first son was 11, and at the time I was basically told, "We don't have enough data. Nothing is safe, that we can say for sure is safe." And so, I was advised to stop taking all of my medications in order to become pregnant.

So I did that, and many women with RA or some unicorn women that I do not know, they're supposed to go into remission while pregnant, but it did not happen for me. So I struggled with untreated RA through my first pregnancy, but I managed. And unfortunately, after my first son was born, I had a really terrible postpartum flare and I had to basically make the choice, "Do I continue breastfeeding him or do I go

back to my RA medications?" And so, I had to give up breastfeeding very early and I went back to my RA medications.

Knowing that I wanted a second son, we tried to see if the data had improved by the time that I wanted to get my second pregnancy, and it hadn't improved much. So I went into my second pregnancy with the same sort of advice, to stop taking all of my RA medications, which I did. Only with my second pregnancy, my RA flared so terribly that we reached a point where the doctors were telling me that the uncontrolled inflammation in my body was far riskier than the RA medications, even though we didn't know that much about the RA medications.

So during that pregnancy, I ended up restarting a biologic for my RA while I was pregnant. It was a really difficult and scary decision to make, because I didn't have any sort of assurances that what I was doing was safe for my baby. But I was not managing the pregnancy and I had to get to the end of the pregnancy to meet my baby. So I did take some medications during that pregnancy with very little information. And so, my second son is now nine.

And it was a completely different situation by the time I was ready to get pregnant with my third baby. By that point, the data had improved to the point that I was able to talk with my rheumatologist and with a maternal fetal medicine specialist, and they recommended a biologic that didn't cross the placenta. I started that biologic before getting pregnant with my third baby. I managed my whole pregnancy without flaring during my pregnancy, and I didn't have a postpartum flare after my third baby was born. And so, with that baby, I was able to make my own choices about breastfeeding. And it was a very different experience for me as a mom to have this information and make a decision that I felt comfortable with and keep my disease under control.

And so, I think it's important to remember that we are here to speak about women and this is real women, real lives. And as you mentioned, I run a website and support group called Mamas Facing Forward, and this is one of the topics that comes up over and over again is that people want to get pregnant and they don't know what's safe, and sometimes their doctors don't know what's safe. And these women are struggling to figure out where to turn. And I would like to know where they should turn so I can tell them where to turn because they want this data. They want information so that they can make choices that they're comfortable with for their families and for their own treatment as well.

Megan Clowse:

Terrific. Thank you so much. Dr. Wurst.

Keele Wurst:

Thanks, everyone. I'm Keele Wurst. I'm really excited to be here to hear about the new pregnancy framework. As Mariah just highlighted, this is such an important issue for us all to work together to really give women and clinicians information that they need to make educated decisions. So I'm here to talk about the industry perspective.

From that perspective, we have a number of challenges. The first is from our clinical trial data, we have limited pregnancy exposures. This may change in the future as we start to include pregnant women in our clinical trials, but for now, it really remains a gap. When a drug is marketed, we don't know who will take this medication, so we don't know how many women will take it, how many women of childbearing potential or how many pregnant women. We also don't know how a drug will be used in clinical practice or if there are characteristics of people that might take that medication that could impact the design that we would use to study pregnancy.

As Leyla touched upon, we need such large sample sizes to study birth defects, overall birth defects alone, let alone specific birth defects that require so much sample size, because they're so rare. So historically, what we've done from an industry perspective is we have set up product-specific pregnancy registries. But as Leyla touched upon, they have limitations. They're voluntary reporting. We have high losses to follow-up, and sometimes they don't even include a comparison group. And they take a long time. They go on for years and years, and sometimes not even generating meaningful data to help women make decisions.

So we had one pregnancy registry that went on for 13 years, and it actually met its enrollment target, but most of them do not. We had an example once from our vaccines registry where it went on for 14 years, and we collected information from 1,400 women, which sounds like a lot. But then, when we take into account loss to follow-up, we actually had only 100 pregnancies with outcomes that we could analyze. But conversely, when we looked at electronic healthcare data, we were able to look at 16,000 exposed pregnancies, which is a huge difference. And so, it kind of speaks to can we use these databases more often?

We had another example of a registry where it was for asthma, and we projected enrollment based on another pregnancy registry for asthma. However, when we looked at the real-world use of our medication, the difference in the populations that used the medications was vastly different. We had an older population, less women, and so that registry really didn't recruit. In that situation, we looked at using our electronic healthcare data sources, even the large electronic healthcare data sources, and we just saw very low exposure.

So what we did there is actually looked at enhancing our pharmacovigilance process and could we get more data to actually provide women and their clinicians some information? So I'm really looking forward to hearing today and how we can incorporate all these data sources from registries, from databases, and spontaneous data to really provide meaningful information to women and their clinicians. So thanks.

Megan Clowse:

Thank you. Dr. Wisner.

Katherine Wisner:

Good morning. I am a psychiatrist who practices in obstetrical settings, and so pregnant individuals are the sole population that I work with. I've been doing this for nearly four decades, being someone who contemplates very difficult decisions with patients as they decide about medication use in pregnancy.

When I started my career, the mantra was, "You're pregnant, stop all medications." I see that really hasn't changed that much. But what happened was I was watching pregnant individuals throwing themselves against seclusion room walls. So this appreciation is developing to some extent that the diseases that justify drug treatment affect pregnancy and offspring outcomes as well, and that optimal care involves a balanced approach between the control of the disease and the risk of medications.

I'd like to share some observations. First of all, prescribers and patients always have incomplete data. It is always there. There's never every outcome possibly studied, so we're always working with incomplete data. There are no zero-risk options including pregnancy itself, which has substantial mortality, especially here in America.

When I do a consultation, I owe the patient up-to-the-minute information about the disease course in pregnancy and the drug risks to control that disease. I owe information about what domains of information are available. Do we have no information? Is it a brand new drug? Do we have a few case

reports, large case series or higher level data like large scale epidemiologic studies, that have the capacity to control for confounding variables, that is those factors that come with the disease that's being treated that affect reproductive outcomes that we need to separate from the effects of the medication.

These studies take a very long time. An example is for fluoxetine, also known as Prozac, which was introduced and marketed in 1988. It has just been in the last year that we have high quality, large scale data to show that neurodevelopmental outcomes in children are not affected by drug use. That's 35 years. So this is the kind of information we are dealing with.

A major challenge in this area is putting the risk of adverse effects into perspective. So people are often frightened by, "Well, if you take this drug, there's a five times greater risk of a particular malformation." But if the malformation occurs in one in 100,000 births, you're looking at one in 100,000 versus five in 100,000. And unfortunately, tackling some of the information on the internet and presenting real factual data is an increasing challenge.

I also find that patients are the true experts. They're the ones making the decision about whether, with your guidance, they are going to accept the medication or not. Is the risk associated with not prescribing greater than the risk associated with prescribing?

I also find that patients are often concerned about function, so will they be able to be well enough to sustain their employment, their insurance, take care of their other children? So function is an important element of the decision-making.

So what are unmet needs? When drugs are approved by the FDA, they're done with drug versus placebo comparison studies, where the effectiveness of the drug is known as well as the side effect profile. But as we've heard, pregnant individuals are not included. Post-marketing studies, however, tend to focus on adverse maternal and offspring effects, and the control of the illness piece is largely missing in most of the way these studies are designed.

We need to collect data on the disease course during treatment. Is the drug as effective in pregnant as in non-pregnant populations? It is well known that there are pharmacokinetic changes, changes in liver enzyme function. They're activated often in pregnancy, and concentrations of medications decline. So we need more information on dose optimization.

Another problem is that more and more individuals who are pregnant have complex diseases and multiple medications are often required to control those diseases. The literature is set up around single drug use, single drug effects. So we have to kind of put together what are the drugs and make some estimate without much data about what the effects may be. So thank you for allowing me to present today, and I'm really looking forward to this conference.

Megan Clowse:

Excellent. Dr. Swamy.

Geeta Swamy:

Thanks, Megan. So the benefit of going last is to not repeat, but to think about how to be additive. And so, sitting here on the panel as a physician representative or a clinical provider, one of the things that strikes me is the role that providers play, both as OB/GYNs or obstetric providers as well as other providers that might be caring for the pregnant person or individual seeking pregnancy.

And one of the things that strikes me is that we really are supposed to be providing care, providing information, providing a pathway, not standing in the way. And so, if we function as gatekeepers, then we're already doing a disservice to our patients. And I'll give you an example of that that's completely

different from this space. But really, if we look to what has happened over the evolution for the last, gosh, I guess 20+ years now, from when we used to require annual exams with a pap smear, and that was our gatekeeper to contraception. We would not give you a refill on your oral contraceptives or whatever it was unless you came in for that exam, because we wanted to make sure we screened you for cervical cancer.

You could argue that we also shouldn't want you to become pregnant if you don't want to. So why were we using one in place of the other as a gatekeeper? And it's a similar thing that I find that there can often be discussions or maybe disagreements between say, obstetric providers and say other subspecialist providers. So we might have a scenario where I know that every time I see a pregnant person who has a medical condition, I'm pretty certain that if I don't have a healthy pregnant person, it's unlikely I'm going to have a healthy fetus or healthy infant. So if my goal is to have a healthy mother and a healthy baby, it seems like we should be looking at those two together.

But then, if the individual who say is treating a person for rheumatoid arthritis or say for ulcerative colitis or MS or some other condition says, "No, no, we don't have data. I can't prescribe that medication." Someone is holding the gate, right? So we have to also think about who the audience is well beyond the patient, who really, as you said so eloquently, are becoming the experts right in front of our eyes. And we have to be thinking about how to work together.

If we look at silver linings of Covid, I would tell you that I think it's one of the first times that we have seen incredibly in sync collaboration in the vaccine world between American College of OB-GYN, Society of Maternal Fetal Medicine, American Academy Pediatrics, Infectious Disease Society of America, CDC, FDA, and others.

But it's sad that it took a pandemic to get us all to a place where we were having conversations and agreement. Why aren't we doing that on all of these other conditions? It takes an effort, but it's not insurmountable, to be honest. It really is about coming together and thinking about, "What do we know? What do we not know? What do we need to figure out?"

Even though it was very, very, very small numbers of data, the information that came from the inadvertent conceptions from the Pfizer and Moderna studies of the Covid vaccine was still helpful information. It might be small, but I can tell you when a patient looks at you and says, "What would you do?" I don't want to tell them what I would do. I want to tell them what I know and how that can help us to make a decision.

So really having as much information as we can, even if it seems like small numbers, it's not powered. It's not data that we would ever say, "Great. We're going to go with it all the time." It is still helpful, because it also helps to instill some trust in our patients and their families that we are paying attention and that we're not simply just there as, "Nope. Sorry, we can't help you today on that."

And I can't say enough about the issues related to exposure during pregnancy as well as exposures prior to or periconception. They are very different. If we continue to focus solely on the thalidomide story or birth defects, we are missing the extensive amount of time that supersedes the sort periconception time period. One doesn't equate with the other. Leyla said the exposure to organ development is up to eight weeks post-conception and so forth. So why do we then always look at the number of birth defects, the adverse events is birth defects, when we're talking many times about conditions that we want to treat over the course of pregnancy?

Another thing that I've heard routinely, not only about, "Well, you stop all medications because you're pregnant" is the other aspect is stop them at the end of the pregnancy. That also wreaks havoc in so many other ways. So rather than just say, "Let's stop medications," why don't we think about what the

best course of action is to care for the person who's standing right in front of you and then their overall health, which includes their pregnancy and the outcomes. That's my moment, Meghan.

Megan Clowse:

Terrific. Thank you. So I hope everybody puts in questions to Slido, or I can also take questions if you raise your hand. But I'm going to kick everybody off with, I know that birth defects is something that is often sort of our leading outcome, particularly in the label and so on. But what other outcomes, both as a person who has taken medicines during pregnancy, as a physician, what other outcomes matter to you when you're deciding about what to do with medicines?

Mariah Leach:

For me, I'm not even sure I knew what the possibilities were. It was just kind of a scary do you want to expose your... You've never been pregnant before the first time you're pregnant, so you're learning about pregnancy as you go. And so, I don't think I knew what the possibilities were, but they were scary because suddenly it was not just my body that I was making decisions about. It was someone else's body as well. And so, that made it really challenging with limited information to make a decision to take a medication.

Megan Clowse:

Absolutely. So it's almost like the vague unknown is worse than the actual drug.

Mariah Leach:

Yeah. And I think in the situations where there wasn't a lot of data, my doctors couldn't provide anything that made me feel better. And there was also issues with my OB/GYN would say, "Ask your rheumatologist." And my rheumatologist would say, "Ask your OB/GYN." And I was sitting in the middle going, "What? How am I supposed to make this work together?"

Megan Clowse:

Yeah.

Mariah Leach:

And so, I think that's becoming more common that doctors are working more with each other. But that is a really big hurdle that a lot of patients face, is that one doctor is telling them one thing and another doctor is telling them another thing and they're left in the middle to decide for themselves.

Megan Clowse:

Yeah. Absolutely. Absolutely. Dr. Wisner or Dr. Swamy, any other outcomes that you would add to the list?

Katherine Wisner:

I think there are two that patients often ask about. One concern is for NICU admission. What kind of condition will the baby be in when the baby's born, and is there any sort of withdrawal or discontinuation? And the second is, "Will my baby be able to function? Are there neurodevelopmental difficulties?" Those are the top two.

Geeta Swamy:

Yeah. The infant outcome is so hard, because one of the most common things that occurs during pregnancy is ultrasound. And I can't tell you how reassured patients are by ultrasound, but it's not exactly accurate, right? And that's in the setting regardless of any medications, right? People see their fetus and they suddenly feel better. But that doesn't really tell you anything about the function of the soon-to-be-born infant. So I think that's a struggle, that people really want to know that their child is going to be healthy.

But the other concern to a point that Dr. Wisner raised is looking at the impact, if we're looking at registries in the post-approval space, is also about the maternal health condition. So I think looking at the outcomes in the other way is not the occurrence of an adverse outcome, but what happens in the untreated space?

So if we have worsening of renal function because someone has gone off their medications, worsening of their MS because they had a relapse in between, and they have a relapsing remitting sort of course, and a number of things that I can name. But if we're not looking at the overall health, then you'd also wonder, "Well, what is the potential alteration to life, life course, life expectancy of that pregnant person? Have we done such a disservice that that infant then has impacts on their long-term family?" And these are real things that we know in improved surveillance and maternal morbidity and mortality are happening around us.

Megan Clowse:

Absolutely. Yeah. I would agree that women have often a desire just for a healthy baby. And that's actually not as clearly defined, I think, often as we would like it to be. But I think if you have children, you sort of think, right, none of them is actually perfect, right? At least mine are not. But nobody's perfect, right? And so, that can't be what has to be the goal, right?

And our patients with chronic diseases also don't have a perfect body based on what somebody would say is the right way to be or a healthy way to be. So they're already living in this space of, "Well, it's not perfect, but I still want to have a child, and I still can be a wonderful parent." And I think that helping women sort of weigh the pros and cons there is really important.

Are there questions from the crowd? I don't see any questions on Slido if I'm supposed to. Yeah.

Speaker 6:

For the clinicians, do you think pregnancy loss-

Speaker 7:

Wait for the mic. I'm turning it on here.

Speaker 6:

Thank you. Do you think pregnancy loss, miscarriage, is important to your patients?

Geeta Swamy:

Absolutely. But again, one of the issues I think there is, we don't often see a lot of individuals who have a pregnancy loss. So the exposure data is really poorly captured in those individuals. And then, ones that we do see, it becomes a question of... It's that sort of retrospect view, right? Maybe that was what caused it. Maybe this is what might've happened.

When you talk about the numbers of conceptions that end in early loss, the numbers that are then actually linked to some genetic causal relationship, which we're finding more and more, right, as we are getting more sophisticated genetics, that starts to dwindle down to pretty small numbers, right? So it is a concern, but I think it's also almost a very separate category, because to know about periconception exposure versus in-pregnancy exposure are really two distinct things.

Katherine Wisner:

Yeah. I completely agree with that. But there's a population we've begun working with, women in the infertility clinic who often are taking psychotropic medications and who are very concerned about that question. So I've recently begun to look more carefully at the data. But that's because they're a group who is actively trying to get pregnant and wanting to understand exposures. But it's a small group.

Geeta Swamy:

I think the other thing that comes into play there, at least in the United States, is that we have such a high number of unplanned or mistimed conceptions that we would be in, I think, an even better place if we could be improving, having the best maternal health possible going into conception. Because I think that's adding to the story as well, that a lot of times, say for example, in the space of diabetes, we often see diabetes is associated with miscarriage, very common, except that when you look at it, it really is A1C-dependent, meaning that you can still have individuals with diabetes that's well controlled that have a miscarriage. But you could have someone without diabetes have a miscarriage. You look at individuals with a high A1Cs, that's where you see higher congenital malformations, higher risk factors, so forth. So we really still have to get at that picture as well. I'm sure in the same space of having good behavioral mental health going into a conception is very different than having someone that is uncomfortable.

Katherine Wisner:

Absolutely. Yeah.

Megan Clowse:

Yeah. Mm-hmm. Yes.

Speaker 8:

Question for-

Megan Clowse:

Hang on. He needs to turn on the mic.

Speaker 7:

Go ahead.

Lee Cohen:

Thank you. A question for Dr. Wurst. Could you speak to what seems like the extraordinary variability that we see across manufacturers of medications with respect to their wish to support, whether it be pregnancy registries, other pharmacovigilance activities, because absent a post-marketing requirement, they do not have to pursue those sorts of activities. So over the last 15 years, it's been noteworthy that

there seems to be extraordinary variability across your industry. Some companies are extremely motivated to identify the reproductive safety data of a molecule, and other companies frankly have no interest whatsoever because they don't have to. I wonder if you could speak to that.

Keele Wurst:

So I can speak from, I think, a company that is actually looking at pregnancy, and we have a lot of pregnancy registries. And actually, GSK was one of the first companies to have a pregnancy registry, so it's been an important thing. I think it depends on what molecules and what vaccines you have in a portfolio. If you don't have a woman taking a medication, is it that important?

But I do think that there is movement in this area for companies to start to work together. We have an example with the EU IMI ConcePTION project, where we have a number of companies that have come together to assess, "How do we study pregnancy better? How do we define elements that we need for pregnancy registries and population-based data sources?" So I think the needle is moving, in that I think there are companies that are starting to work together.

We have another example with our asthma product that we were working with that we reached out to other companies to say, "Hey, we know that working together and creating a disease-based registry, we'll actually get better data." So I think there is variability. We acknowledge that. But I do think we are moving the field a little bit forward and people are coming around to look at this important issue.

Megan Clowse:

Oh, here.

Speaker 9:

Good morning, and thank you for these very important and very challenging discussions in this topic. I heard the comments about the importance of having any data, limited data, and I wonder if anyone could comment on maybe the challenges of that as well. Because you can see that perhaps you're now having a false reassurance of no risk, or perhaps there's a risk that's just very, very, very rare, and you're taking the patient away from receiving a medication. So I wonder if we have some discussions on that point.

Megan Clowse:

Okay. Hi.

Katherine Wisner:

Yeah. I have that situation a fair amount. And what happens then if there's very limited data or just a few case reports, what tends to happen then is the severity of the illness rises in priority. So if it's somebody who... And the other thing that happens is often the new meds, people who try those are people who have failed a lot of other trials. So you've got pretty significantly ill patients who are on this medicine. So what they tend to do is say, "I'm willing to take that risk because I'm a sole provider for my family. If I lose my job, I cannot take care of the people I love." So that gets lifted.

The other thing is, again, I'm talking about how risks are overvalued. The reality is that most babies are going to be born okay. There's rates of all of these things, but even in... The malformation rate with thalidomide was about 25%, right? So it's not a hundred percent. And I worry about really, again, overestimating risk. But like you were saying, very respectful of the woman's right to choose and that I would support her no matter what she chooses. Then what I might do, thinking about Mariah's

situation, is set a set of criteria, "If X, Y or Z happens, then we go back in with the medicine because it's too dangerous for your health."

Megan Clowse:

Mm-hmm. Yeah. If I could add, I would say that early small data can be really helpful and it can actually kind of set things back, actually. So in the rheumatology world, there was a paper that was published a long time ago now that was probably a misinterpretation of data, but it was presented as a plenary at the American College of Rheumatology meeting. Doctors still are holding those medications based on that, even though there's reams of data since then. And so, it can be helpful, but it can also be harmful if there's not more stronger data kind of stepping in right behind.

Geeta Swamy:

Yeah. And I think to that point, Megan, it is also about what you do with the data, who maybe confers to discuss what the data looks like, what the caveats might be. We have relied in our world on peer review, but peer review doesn't get at all the other aspects of the field, right? And so, I think it is important to take small amounts of data and really think through it before it's simply put out there. But I think holding it back just at bay is not helpful and not getting the right people to take a look at it and have conversation about it.

Megan Clowse:

Mm-hmm. Absolutely. There's a question up here.

Marc Stone:

All right. One thing that's interesting about this discussion is that most of it begins with a question of someone who is already pregnant and says, "Oh, you're pregnant. Stop your medicines?" or "Should I stop these medicines?" And it's a very closely related but different problem of someone who is trying to get pregnant. And how do you approach that? And of course, many of them will become pregnant and continue to take medications before they know they're pregnant. And so, it's a continuum. We're going to put everyone who's trying to get pregnant who's on a medication in a registry? How are we going to enact that? And in some ways, it's almost the same problem as saying, "Well, we don't know how this drug works in pregnant women, but this is how it works in non-pregnant people." Same thing. We go, "Well, we kind of know how the drug works in pregnant people, but we don't know how it works in someone who's trying to get pregnant."

Megan Clowse:

Right, right.

Katherine Wisner:

We did some of this thinking about pharmacokinetic studies, where we did a pharmacokinetic study prior to conception for women who wanted to conceive and then track them through pregnancy to look at, "What are the concentrations? How does the drug behave in non-pregnant and pregnant situations?" So that kind of study is out there, but I do think it's an interesting possibility to also optimize her health prior to conception. And that's the group of patients who are much more concerned about the rate of miscarriage.

Geeta Swamy:

And I think we're getting better and better about thinking through the mechanism of action that we know, understanding we don't know what we don't know, but at the same time, we are never going to have all the data we want ever. So it does become, even with mRNA vaccines, right? We had to make an expert decision based only on the laboratory science we had. We had to make a decision without the data that we would've wanted to. And instead, what happened without a recommendation is we relied on healthcare workers who said, "Nope, I want the vaccine. I don't want to get Covid," whether they were periconception, already pregnant, and so forth.

And so, I think it becomes using the knowledge that we have to the best we can and advising patients. But in the periconception or preconception space, as Katherine was saying, in the sort of fertility space, we offer a lot of preconception counseling. And now that we've gone through Covid and done a lot of now telehealth, these are not incredibly complicated visits to have. And we encourage all of our patients to do that if they have medical conditions or are on medications.

Megan Clowse:

Mm-hmm.

Speaker 7:

I think we have one more question.

Megan Clowse:

All right. We've got one more question.

Speaker 11:

Thank you. Just to briefly circle back, one more point on the perception of risk. I think it's an excellent point, Dr. Wisner. And I think it's all in the context as well. We can have a discussion of a malformation that can be one in 10,000, and yet on a completely separate point, be very accepting of a complication rate for a surgery of 20 or 30%.

Katherine Wisner:

Yes. Yeah. Right. Yeah.

Speaker 11:

And so, it is definitely the patient's decision, but it's also a point of education and context for the healthcare provider and an opportunity. So I just wanted to add that. Thanks.

Megan Clowse:

Terrific. Well, we'll finish up on shared decision-making and the importance of that, which I think is what you're getting to, is really being able to help women interpret the risks that are put in front of them. So thank you. All right.

Katherine Wisner:

[foreign language 00:36:38].

Geeta Swamy:

Oh, absolutely. It's what I say.

Session 2

Geeta Swamy:

Good morning. So I'm Geeta Swamy. I will just introduce myself briefly even though I was a panelist on the first session. So I am an obstetrician, gynecologist and maternal fetal medicine specialist at Duke University and I'm really pleased to be here with you all. I'm going to be moderating our next session entitled Stakeholder Perspectives on Challenges and Opportunities to Optimize Post-Approval Pregnancy Safety Study Types and Designs. Our speakers this morning, I'll introduce in just a moment, but in this session, you're going to hear their thoughts on factors that have the potential to enhance the generation of data from drugs and biologics in pregnant individuals. You're going to hear from individuals across academia as well as industry. They're each going to have a presentation and then we'll have a discussion amongst our panelists. So I'm just going to introduce everyone who you're probably hopefully mingling through the morning already and it's great that we're able to be here in person together as well.

So Dr. Christina Chambers is a perinatal epidemiologist and professor in the Department of Pediatrics in the School of Medicine at UC San Diego. She's the PI of the MotherToBaby Pregnancy studies and has led a number of national, international complex longitudinal cohort studies as well as clinical trials focused on environmental causes of birth defects, adverse pregnancy outcomes, and childhood disabilities and conditions.

Dr. Jessica Albano is a pharmacoepidemiologist by training and is the Vice President of Epidemiology and Analytics at Syneos Health. Dr. Albano serves as a PI for the Antiretroviral Pregnancy Registry, which is an international collaborative that's a pregnancy exposure registry, and that's been actually ongoing for over 30 years.

Dr. Christine Olson is an OB/GYN trained medical officer at the CDC's Immunization Safety Office. She served as an expert consultant in infection control and prevention measures in maternity care settings for the World Health Organizations and on multiple federal collaborative work groups. She has been leading the CDC's COVID-19 Vaccine Pregnancy Registry for the past few years as well.

And Dr. Elyse Kharbanda. Elyse is a pediatrician by training and is the Executive Director for Research at HealthPartners Institute, which is a health system-based research institute. And her work focuses on safety of vaccines used in pregnancy primarily through her work with the Vaccines Safety Datalink.

So I'm going to take a seat and turn over to Dr. Chambers to get us started.

Christina Chambers:

Thank you. Okay, thank you so much and I appreciate the opportunity to be able to speak on this panel. So as said, I'm going to give a little bit of the perspective of pregnancy registries from the MotherToBaby point of view and a number of points that I'll make today Dr. Sahin and others have already made. So MotherToBaby, for those of you who are not familiar with this, is a US- and Canada-wide network of counseling services provided by the nonprofit Organization of Teratology Information Specialists or OTIS. These were established in the 1980s and there are about 14 services currently funded by HRSA just

recently by CDC and other state sources. And these services provide a service, individualized evidence-based information to pregnant and lactating persons and healthcare providers and the public about the safety of medications, vaccines, infections, other exposures that a person may have had already or is anticipating during pregnancy and lactation.

So in the last 25 years or so, because of the dire need to have information to provide people who contact us with these questions, MotherToBaby has engaged in collaborative pregnancy studies. And we conduct these studies like a pregnancy registry, an open cohort study through a single research center at the University of California San Diego.

So MotherToBaby pregnancy registries are constructed as a cohort study design, and as Dr. Sahin mentioned earlier, these are a registry design that recruits exposed and then comparison pregnant persons in the US and Canada to compare the standard outcomes of major structural birth defect overall. And then we specifically review these for a pattern of malformations as has been suggested by others. We also look for a pattern of minor or more subtle structural anomalies among infants who receive a study related physical examination. And then we look at the other range of outcomes including pregnancy loss, preterm delivery, growth, both at birth and postnatally, and then typically over the first year of life, serious infections, malignancies, and then at least short-term neurodevelopmental screening up to one year of age. And some studies have longer term neurodevelopmental follow-up.

So the schema for the overall study design is shown here. So depending on what the exposure is, we typically recruit in the middle of the first trimester, about seven to 10 weeks, and women go through an extensive telephone interview at that point in time about all of their exposures, comorbidities, pregnancy history, medical history, and so on. And then that's updated with an additional interview in the second and third trimester. And then depending on what the outcome of the pregnancy is, as soon as the pregnancy ends, there's an outcome interview and that's the point at which we request and abstract medical records. So that's from the OB, hospital of delivery, pediatrician, any specialty physician, any pathology reports. And through that one-year follow-up period, if there's been a suspected malformation, additional records are requested to help confirm what that might be.

And then in the one-year follow-up period, most studies have this for live born infants, we offer the specialized physical exam conducted by one of a team of five pediatric geneticist dysmorphologists who do a blinded exam of the child for these subtle minor anomalies using a checklist. And this is now done typically by telemedicine since the pandemic. And then the standard follow-up is to one year when we get another pediatric medical record to abstract for growth and malformations and then the neurodevelopmental screening. And then the arrow goes out because we consent people to further follow up if necessary. So some studies go out to five years. And we typically recruit for a medication, a cohort that has exposure, meeting the inclusion/exclusion criteria, a comparison cohort that has the same underlying disease as the exposed cohort but not exposed. And then a secondary comparison group that has neither the exposure nor the disease.

Assumptions: The key assumptions that have been mentioned earlier that we're functioning under are that pregnancy registries, including ours, are typically statistically underpowered to evaluate the risk or safety for specific congenital anomalies unless the magnitude of the risk is astronomical. And even if a registry is adequately powered to rule out, say a two or threefold increased risk of major birth defects overall, that's not what we expect of a teratogen. We expect that specific malformation are clusters of anomalies would be induced and that's what the teratogenic effect would be. So that's basically what we're looking at.

So as was mentioned earlier, initial goal is to try to rule out, say this is not a thalidomide, this is not an isotretinoin, this is not a mycophenolate, this is not probably a methotrexate, and then maybe this is not a valproic acid. And then beyond that, depending on the sample size and if there is a signal, what we can

do. Of course this requires careful evaluation and accurate classification of congenital anomalies in the context of knowing specifically what the gestational timing of the exposure is as it relates to biological plausibility of it being causally related to the outcome and consistency of patterns, if any, are identified. And as has been mentioned, teratogens are often associated, not just with malformations, but with a range of adverse outcomes including increased risk for pregnancy loss, growth deficiency, and possibly neurodevelopmental deficits.

So using those points or keeping those points in mind, MotherToBaby pregnancy registries focus on a range of outcomes. So they all include the range that I described. They all include a minimum of a standard one-year follow-up in each of the cohorts. We have this addition of the study related physical exam for minor anomalies and specifically to assess these children for a pattern as this pattern of minor anomalies may occur more frequently than an increased risk for a major birth defect associated with the drug. So it's a way of screening for that.

And the source of data on exposure is what we call the truth. It comes directly from the mother. So it's always surprising to me that what a mom is prescribed is not always what she takes, even for medications that she would think she would adhere to very closely. They often do not. And so having this internal comparator groups we think is incredibly important. They're recruited and followed in the same manner and for the same follow-up period as the exposed group. And then as has been mentioned earlier, the importance of potential confounders. As these are observational studies, we assess a wide range of potential confounders including co-exposures and comorbidities that are unlikely to be obtainable or obtained reliably from any other source.

So limitations, as have already been mentioned, are time and resources required to do these. This is a limitation. I think we often have relatively limited diversity and socioeconomic status and race ethnicity in the sample who enroll in these studies. From the standpoint of teratogenicity, you could argue that if a drug is teratogenic, it's going to be teratogenic irrespective of your socioeconomic status or race ethnicity. But in terms of generalizability and so on for other outcomes, it's a deficiency that we're working on rectifying. We typically have small sample sizes. And then important to keep in mind that if the medication exposure is not continuous throughout pregnancy, we have statistical power issues that are exacerbated if the exposure is intermittent. So is the exposure sample size, it really becomes the sample size that's exposed in whatever the critical window is for the outcome that you're looking at. And then last one study is only one study. So pregnancy registries are not the be all end all and never will be. They're one source of data that can contribute to our understanding of safety.

So I want to say a word about our collaboration with the American Academy of Allergy, Asthma and Immunology, what's called the Vaccines and Medications and Pregnancy Surveillance System as an approach to addressing this issue of needing to have multiple approaches to address the same research questions. So this was established back in 2009 with the pandemic H1N1 influenza outbreak and need for safety data on vaccines that were being generated at that time.

So under this umbrella, we run the pregnancy registry cohort arm, and then there are two complementary arms, database or claims data arm, run by the Harvard Group. Krista Huybrechts is here today and Sonia Hernández-Díaz run that. And then we had a case control arm, which has been the Boston University case control study, but more recently the CDC BD STEPS. And the point of this is twofold. One is to be able to say rather than doing this sequentially, that we'll look at a pregnancy registry and then we'll follow it with this kind of study or that kind of study if there's a signal, to do these simultaneously so that we can get to answers and interpret the complementary data as a group more efficiently and quickly.

But also a second objective is that if the pregnancy registry is failing to recruit, having the claims data information to help us understand or put into context whether the prevalence of exposure is really

there to support the pregnancy registry recruitment is really incredibly helpful in determining what's feasible or not. So we have produced complementary papers, for example with the case control and the cohort or registry arm on, the example here is pandemic H1N1 vaccine in the journal Vaccine.

So the next couple of slides, I'm going to give you a few examples of pregnancy registries through MotherToBaby that were so-called "successful." And really the take home message here is that, as has been said before, many of these have taken a long time to recruit, whether it's due to the trajectory of prevalence of exposure or the participant burden of participating in a registry. So people who have to make the commitment to do this over a two-year period of time or longer and to engage in providing us with a lot of information, multiple factors may be involved. But these are ones that have been "successful." So we met our target sample size. We typically have quite low loss to follow-up, so not threatening the validity from that standpoint. And those that were completed some time ago are in the product label in the section regarding pregnancy. Ones that have just been completed in the last year, we expect will go into the label and we met our target sample size there. And then we have two that are just winding up in this next year. We completed enrollment for Pfizer COVID-19 vaccine study. It met our target sample size and are completing one on dupilumab. So these are so-called successes with modest sample sizes, but met the target enrollment.

These are three examples of ones that we might call failures. So the target sample size, despite many years of recruitment and promotion of the study and so on, was not met. We continued to recruit in the comparator groups. But as you can see here, for tocilizumab, tofacitinib and mepolizumab, we only achieved between 11 and 34 exposed pregnancies ballpark. The last one, mepolizumab, Dr. Wurst was talking about earlier, here's a drug that was part of this VAMPSS collaboration and this was an example of where we didn't know, as was expected when this drug was first marketed, we thought, well maybe it'll be like the first biologic in the asthma space that we would be able to achieve a sample size as was outlined here. And it turned out that we did not. But the claims data in VAMPSS was really useful in helping us understand that the prevalence of use was quite low and that consistent with us not recruiting as many as we had anticipated.

So when a pregnancy registry works such as ours, what are the pluses? If the prevalence of use in pregnancy is sufficient to support the pregnancy registry, we think having this ongoing open cohort platform, and I'll call it a platform because it doesn't require you setting up independently one-off a new registry from beginning to end, it's layered onto an existing platform. So it requires very little setup time, it's the amount of time that it takes to get addendum for the substudy approved by the IRB. Engagement with the pregnant woman leads to we think a high retention rate, why we have a limited loss to follow-up.

We have the ability to, we think do a good job of classifying outcomes using these multiple sources of data, including maternal report, medical records, abstraction over a year's period of time, and then this study related expert assessments. We have the ability to confirm actual exposure dose and gestational timing by capturing this from the mother. We have the ability to acquire data on covariates that are not typically available through other data sources, including EHR, such as substance use, over-the-counter medications, fever, herbal products, as well as we have the ability to capture measures of disease severity and have these disease match comparators. And we have the ability to do broad and extended follow-up if needed, including responding to a signal.

So I'll give you one example of a signal. In the etanercept registry, we did identify a pattern of three or more minor anomalies that were the same ones identified in six children who had received the study-related physical examination. And so we followed up on that signal by having those children and their parents re-examined by a different examiner to confirm that the minor anomalies that are sometimes very subtle were actually there and to see if they were in the parents as well. And then we offered

neurodevelopmental testing face-to-face for those children and five of the six completed it. And the result was that most of the parents also had the minor anomalies. The minor anomalies that were seen in the children were validated by the second exam. And all of the children who were tested performed within the normal range. So evaluation of the signal we felt led to no further concerns.

If a pregnancy registry such as ours doesn't work, what then? So why did you start it in the first place? And I know that's kind of what the framework is about here. If the prevalence of use in pregnancy cannot be reliably predicted to be rare, so you just don't know, we think it's important to plan for a feasibility period, so rather than saying we'll do this indefinitely and hope that we reach a sample size that's interpretable, to say after a given period of time and using other data sources such as claims data, is this really feasible as a source of data and not to further invest in going in that direction. As mentioned, it's important to have an additional source of population data such as the claims data or EHR to confirm or refute evidence for low use.

And as was mentioned earlier, and I'm glad that Kathy Wisner brought this up, we do need an approach and a systematic approach to interpretation of small numbers. So it's one thing to say you have three exposed pregnancies and two were preterm, and then what do you do with that? Nothing, there's not much you can do. But if you have 30 exposed pregnancies and you don't have 25% who have a pattern of malformation like mycophenolate or like isotretinoin, can you say this is not a thalidomide? Maybe you can. So coming up with some approach to being able to deal with those smaller numbers, as long as the data is of good quality, I think would help the field in general.

So final comments. I think we can build on efficiencies associated with disease-based registries, and I know Jessica Albano is going to talk about this and theirs is the perfect example of this. I'd go a big leap further and say given the opportunity to be at the pulpit here that we would greatly benefit as a nation from the efficiency and productivity of establishing a single US-based pregnancy registry, not just a disease-based registry, but a pregnancy-based registry to serve as a signal detection system for all new drugs irrespective of the prevalence of use. So not having to bank on there'll be enough exposures, but creating a pipeline whereby healthcare providers and pregnant women could all be part of this.

And to the comment from the FDA earlier, I don't see why people who are planning pregnancy couldn't be part of that to initiate enrollment in a pregnancy registry when they're planning pregnancy with some minimal baseline data so that we actually can follow them from the time of conception throughout pregnancy. So there would be a lot of fiscal efficiencies and a lot of scientific efficiencies I think as well.

And then in the meantime, I think it's important, this is a broken record, that when pregnancy registries are initiated, that they do not function in a vacuum, that we need coordinated efforts across data sources. So I will end there. Thank you.

Jessica Albano:

All right. Good morning. It's a pleasure to be here and to talk with you today about my insights on pregnancy exposure registries from a multi-product disease-based perspective. All right, by way of disclosures, I'm an employee of Syneos Health that runs the Antiretroviral Pregnancy Registry, which is a collaborative study jointly funded by the following list of manufacturers.

So multi-product or disease-based registries can be defined in two ways. Based on the patient's exposure to the specific drug or products being monitored. Usually this is without the regard to indication for treatment and includes all marketed brand and generic versions including regardless of how they are prescribed and then secondly, based on a patient's diagnosis of a particular disorder, including those who are either treated or untreated. In general, multi-product or disease-based registries can be a good consideration when the products are manufactured in combination and therefore individual drug exposures can't be separated out.

There could be complex multi-drug treatment regimens. Again, making it difficult if not impossible to separate out exposures. There's a high likelihood of polytherapy, so multiple different disease specific treatments that are given, or polypharmacy when multiple different disease are being treated at the same time in the same individual. The treatment landscape can be rapidly changing. For example, due to frequent new product approvals in the market, internal comparisons are desirable or when confounding disease or population characteristics exist. Ultimately the focus needs to be on evidence-based data collection that better meets the needs of both the physicians as well as the patients who are making the treatment decisions. We have work to do towards expediting the accrual of patient safety data in order to inform patients and healthcare providers in a timely manner.

To put this into perspective, we've heard a lot about sample size so far this morning, but if one is looking to roll out a twofold increase in the overall risk of birth defects based on a general background defect prevalence of 3%, you would need 200 live birth outcomes in order to establish a relevance of a signal or not. So that does seem reasonable, but if you put that in perspective of a more rare congenital malformation such as neural tube defects that has a background prevalence of 0.1%, you would need 2000 exposures in order to evaluate that potential signal. So we can agree that there are very few examples of pregnancy registries that have met that threshold. And while some disease pregnancy registries may have as many as a half dozen or more distinct registries that have been launched to address these questions, and even some have ended potentially even meeting their sample size goals, it is unlikely that they were individually robust enough to provide sufficient evidence of risk.

So for the remainder of this talk, I'm going to focus on the pregnancy exposure registry, the Antiretroviral Pregnancy Registry, of which I serve as the primary investigator, and some lessons learned that can be applied to this situation.

I'll start by... setting the stage here that the APR is a collaborative endeavor with multiple stakeholders that work together for the common goal. The APR is a voluntary international perspective exposure registration cohort study. The purpose of the registry is to assist clinicians and patients in weighing potential risks and benefits of HIV treatment during pregnancy. The objectives of the registry are to provide early warning signals of major heterogenicity to estimate the prevalence of major birth defects and compare that to the general population and to supplement preclinical and clinical and epidemiological study data.

The predecessor to the APR was the Zidovudine and Pregnancy registry, which was established in 1989 by Burroughs Wellcome. With the entry of additional antiretroviral products into the market, the registry expanded to become the Antiretroviral Pregnancy Registry in 1993. The APR fulfills post-marketing commitments for the FDA. Currently there are 24 sponsoring manufacturers. The registry monitors prenatal exposures to 164 drugs, including 61 brand name single-entity drugs or fixed-dose combinations and 136 generic versions covering HIV treatment and prevention, including pre- and post-exposure prophylaxis as well as hepatitis B treatments.

The primary outcome is prevalence of major birth effects. Infants are not followed after birth. The registry was not designed to formally evaluate premature birth, low birth weight, small for gestational age, or developmental delays. The analysis is multi-tiered, so the prevalence is calculated overall and by trimester of earliest exposure for all drugs being monitored at the drug class level, at the individual drug level, and for common drug combinations or treatment regimens. Comparison groups include the external background reference groups of the Metropolitan Atlanta Congenital Defects Program, or MACDP, and the Texas Birth Defects Registry, or TBDR, and then there are also internal comparison groups.

The APR also follows the rule of three, which specifies that the occurrence of three specific birth defects for any exposure or exposure combination triggers an immediate review. This is particularly useful for

drugs with fewer than 200 exposures for which prevalence cannot be calculated. However, understanding weak signals that may vary over time needs to be carefully monitored and is methodologically challenging. The dedication of a potential signal requires detailed analysis to evaluate and identify this risk. In addition to comparing to the prevalence, the two external comparison groups, having multiple available internal comparison groups, is a distinct advantage of multi-product registries. It affords further comparisons such as exposed exposure to drug A in the first trimester to exposure to drug A in later trimesters, exposure to the drug of interest in the first trimester versus no exposure to that drug at all, exposure to that drug in the first trimester versus non-drug exposure in the first trimester. Comparison one mimics the approach used in the APR for its primary analysis and the other comparisons can serve as confirmatory internal comparison groups when evaluating a signal.

As previously described, the APR began as a single drug registry in 1989 and has expanded over the past more than 30 years to become a multi-product multi-sponsor registry. As of January 2023, there are nearly 26,000 women defined as prospective cases have been enrolled from 75 countries. More than 70% of enrollments come from the US and its territories and the loss to follow-up rate is currently 11%. Given the diversity of the APR, population representativeness is a concern and reliance on a single external comparator for risk assessments is considered risky. Based on the recommendation of the APR's Advisory Committee, they evaluated external comparators and added, in the absence of a US national data source, a second external comparison group, the Texas Birth Defects Registry, to supplement the original MACDP. The APR continues to evaluate other international resources to further diversify comparator populations.

The APR as mentioned requires a threshold of 200 exposures in the first trimester before conducting a drug specific analysis to calculate prevalence. For the newest class of ARV drugs integrase inhibitors, which includes bictegravir, cabotegravir, dolutegravir, elvitegravir, and raltegravir, those that have reached the threshold have taken between four and nine years to accrue 200 first trimester exposures. As depicted in this forest plot, there were a total of 21,636 live births, including 631 outcomes with at least one birth defect, 348 of those which occurred in the first trimester. The overall APR birth defect prevalence is 2.9%, 95% confidence interval of 2.7% to 3.2%. This proportion is not significantly higher than those reported in the registry's two external comparator groups. There's also no difference in the internal comparisons of first trimester exposed pregnancies versus the second or third trimester.

To date. There's been no concern among the individual drugs analyzed, with the exception of nelfinavir and didanosine, which have modest, but statistically significantly increased risk compared to the MACDP. However, for these two drugs, we have not seen discernible patterns among the defects reported, and due to concerns with the safety of these drugs, they are rarely used in pregnant populations of child-bearing potential. Therefore, we're not seeing new exposures and this risk is likely not to be further refined.

You'll recall I started by describing the APR as a collaborative registry, one that began by choice and evolved into a post-marketing commitment over time. Essential to the success of any complex registry is a well-defined governance structure. And by a collaborative registry, I'm referring one in which multiple stakeholders work together to meet one or more specific objectives, whether this is by choice or by mandate.

The figure depicts the governance structure of the APR. It's overseen by a steering committee. It's comprised of the following groups, the Scientific Advisory Committee, which is made up of experts in the appropriate fields including academia, government, and private practice. The advisors provide scientific oversight and review and interpret the registry data. The consultant birth effect evaluator is a member of the Advisory Committee. The sponsor representatives are from the manufacturing companies and they oversee the registry management, budget approvals, and regulatory reporting. The

registrating Coordinating Center is responsible for managing the daily operations of the registry, including the enrollment, the data collection, statistical analysis, report writing, interactions with the IRB, and the facilitation and interaction of both the advisors and the sponsor groups. The primary benefit is such a governance structure is clear separation of scientific, financial, and operational objectives.

The APR Steering Committee convenes twice annually to review and interpret the analysis results. The Advisor Committee meets separately to form an independent consensus statement. In an effort to help ensure that the study results are communicated consistently and accurately, the registry requires all publications of data to include a published consensus statement as seen on this slide. The APR's ability to maintain enrollment and remain relevant over these many decades is due in large part to its internal and robust dissemination of findings. An interim report is published every six months, as Leyla mentioned earlier, and it's posted to the study website and mailed to participating healthcare reporters. A webinar focused on treating physicians is given annually. The results are also used in the US Perinatal Treatment Guidelines and drug specific data is included in the product labels and adherence with PLLR.

Now that we've seen an example of a successful collaborative registry, I want to unpack some of these lessons learned starting with the challenges. First is complexity. Whether there's three or 30 products in a multi-product registry, it can be operationally and analytically complex and require a high degree of expertise to carefully plan and implement. The second is collaboration. They require agreement from companies who are competing in the same therapeutic area to work together, adopt common processes, adhere to agreed policies and timelines. In some cases, there can be fairly significant differences in the profile of the sponsor companies involved with regard to their cultures, the SOPs, their regulatory interpretations, the level of risk they're willing to assume, and resources. The challenge is to bring all the stakeholders together collaboratively to avoid delays and conflict that could put the entire program at risk. Regarding communication, all parties must respect the established lines of communication, which should be appropriately documented.

Competition: Multi-sponsor registries don't happen at once, but over time they grow in size and with more entrants coming into the market. Innovator companies are usually the first of the table, frequently responsible for the design and set up and implementation of the study. Latecomers will often have to adopt the established processes and governance, which can be challenging and make documentation a key process and oversight. There is confidentiality to consider. It's necessary to have sensitivity to the proprietary aspects of drug discovery, marketing and lifecycle management for each of the individual sponsors, as well as direct communications that might happen outside of the registry between regulatory agencies and pharmaceutical companies. However, they may have relevance to the establishment and the conduct of the registry itself. So it's an ongoing challenge to ensure adequate confidentiality is paramount in all agreed processes.

And then finally, commitment. Various stakeholders that we've even heard across industry may not have had the same level of commitment to any given project. It's just how business works and there might be numerous reasons for this. Equal input and engagement should not be expected, so it's important to build an expectation and commitment requirements to join a multi-sponsors registry at the outset.

So despite their complexity, multi-product or disease-based registries have distinct advantage over single-product registries. We've heard some of these already today, but they're logical in that they avoid duplicate efforts in the establishment of the registry. They form reporting lines from a perspective for healthcare providers, pharmaceutical companies, and regulators that are less complex. They're economical because resources can be pooled and as well as budgets for multiple stakeholders. Not only are multi-product registries more efficient with their use of limited budgets and resources, but with their utilization of experts in the roles of advisory board members as well as birth defect evaluators who

can provide input for the overall program on a global scale. Multi-project registries reduce competition for eligible patients and streamlined healthcare provider participation. They also offer a more robust consolidated awareness effort. Multi-product registries are methodologically advantageous through the standardization of data collection, case evaluation and statistical analysis, not to mention the enhanced validity and power.

Most importantly to the advantages; however, it comes from the clinical point of view. Multi-product registries serve as a centralized resource for patients and physicians. They minimize the reporter burden of healthcare providers. They increase the incentive to participate by providing a single comprehensive reporting mechanism. They offer a coherent assessment of the available data and they provide a consistent message of the current understanding regarding risk for both patients and physicians.

Okay, thank you very much.

Christine Olson:

All right. I'm checking my watch though. I'm going to switch to good afternoon. Good afternoon. Thank you for having me here to present on CDC's COVID-19 Vaccine Pregnancy Registry. I'll note, although the slide is at the end, that I have no disclosures and the opinions expressed here are those of the speaker and not of the agency. During this session, I'll be describing the concept and implementation of CDC's COVID-19 Vaccine Pregnancy Registry, the operations of the registry, including the enrollment and data collection process, the two-phase approach used during the pandemic and the cohort of participants, and then addressing the successes and challenges of this approach.

This graphic depicts the different vaccine Safety Monitoring Systems employed by CDC's Immunization Safety Office, which include the Vaccine Adverse Event Reporting System or VAERS, which many are probably familiar with, the Clinical Immunization Safety Assessment Project or CISA, the Vaccine Safety Datalink or VSD, about which you'll hear more from Dr. Kharbanda shortly. And the two newest systems that were initiated during the COVID-19 pandemic in 2020, v-safe and the COVID-19 Vaccine Pregnancy Registry that are currently housed together under a new team, DETECT, the Data Exploration and Technology Team.

So first I'll start with a bit of history from way back in the early point of the pandemic when vaccines were first becoming available in 2020, December of that year. At that time, you'll remember that there was limited data available about COVID-19 vaccine in pregnant people and that they had been excluded from the phase three clinical trials, leaving only developmental and reproductive toxicity, or DART, animal data available as the only source of information at that time.

At the beginning of the pandemic, we didn't know if SARS-CoV-2 in pregnancy was associated with increased risks in pregnancy. Though it was well established there are changes in adaptive immunity and physiology that occur in pregnancy that could increase the risk of severe respiratory viral illness and it was reasonable to assume pregnant people may be at increased risk for severe disease and adverse pregnancy outcomes. We've seen this with other respiratory viruses, notably with H1N1. This was also likely to lower the uptake of a new vaccine product in the pregnant population. And both of these assumptions were born out over time as the pandemic progressed.

The COVID-19 Vaccine Pregnancy Registry was therefore conceived, developed, and implemented relatively quickly to coincide with the early availability of COVID-19 vaccines to help fill this information gap. It was a collaborative effort between multiple areas of CDC, including the Immunization Safety Office, Division of Birth Defects and Infant Disorders, and the Division of Reproductive Health. The new V-Safe After Vaccination Health Checker, which was a web-based voluntary safety surveillance system for adverse events, was launched in December 2020 to capture information about vaccines' responses to receipt of COVID-19 vaccine. We were able to use v-safe enrollments to identify individuals who

reported they were pregnant or became pregnant shortly after one of their primary series vaccine doses. And this allowed a readily available convenience sample of recent vaccines potentially eligible for the pregnancy registry to follow in fairly real time with enrollment in the registry beginning in January 2021 and continuing to enroll those who had enrolled in v-safe by June of 2021.

The initial goal of this pregnancy registry was to monitor for adverse outcomes of interest in pregnant people receiving COVID-19 vaccines in a systematic and fairly rapid way to identify early safety signals, characterize the safety profiles of COVID-19 vaccines in pregnant people and supplement the existing passive and active surveillance systems. The initial planned framework included up to five phone interviews during and after pregnancy collecting participant reported data, medical record acquisition and abstraction for those who consented and who reported outcomes of interest, analyzing the self-reported data initially by chronologically available outcomes. This required a clinical review step for some outcomes such as stillbirth and birth defects. And a plan for more robust analysis later after medical record information was available for confirmation of and more detail about outcomes of interest. The slide lists the planned outcomes of interest that the registry would be able to capture by obstetric and neonatal and infant categories. Those in bold are outcomes for which we determine medical records would be most helpful and were therefore requested.

I'll next describe the enrolled cohort resulting from this process and the data flow in operations. So this graphic depicts the process of enrollment that, again, began in January 2021 and included those who reported a pregnancy into v-safe between December of 2020 and June of 2021. Eligibility screening ensured that participants were 18 years of age or older, spoke English or Spanish, and were pregnant at the time of vaccination or became pregnant based on their LMP within 30 days of receipt of vaccine. During phase one, the interviews concluded questions about demographics, health during the pregnancy, pregnancy outcome, information from the birth hospitalization, and postpartum and infant health through three months of age. Participants were asked to consent for medical record review. In phase two, the last rectangle on this slide, conducted 15 months after the end of pregnancy. Participants who were not lost to follow-up and who had not opted out a future follow-up included questions about both participant and infant health with the primary focus on infant health, particularly on birth defects ascertained during the first year of life consistent with how routine birth defect surveillance is conducted.

On this slide, you'll see that over 123,000 v-safe participants ages 18 to 54 years were identified as potentially eligible for the pregnancy registry. In May 2021, a pregnancy follow-up survey was sent through the v-safe platform to confirm pregnancy into better identify likely participants among potential participants who had not yet been reached. Those who had responded that they were or had recently been pregnant and gave permission for contact were called. While those who indicated that they were not pregnant around the time of vaccination or were not interested in participating in the registry were removed from the list of potential participants. Overall, about half of participants who reported a pregnancy into v-safe from December through June were called. This is shown in blue in the pie chart on the left. About 32% were unreachable, which was our biggest barrier to enrollment, making contact. Among the 20% we were able to reach and assess for eligibility, 94% were eligible and enrolled, bar on the right. A very small proportion of those eligible actually declined.

This graphic shows the characteristics of over 23,000 total eligible and enrolled participants who reported at least one pregnancy into v-safe between December and June. Most received one of the mRNA vaccines. A majority self-identified as non-Hispanic white. The mean age at first vaccination was 33.5 years old. And 45% identified as healthcare personnel a subtype, which was not surprising given the stage rollout and prioritization of vaccination and vaccine distribution early in its availability.

This graphic demonstrates the timing of the earliest vaccination during their peri-conceptual period or during pregnancy among our participants with distribution seen across the spectrum and over half receiving their vaccines in the second or third trimesters.

You can see in this graphic that those in the registry who reported receiving the monovalent booster dose of vaccine during pregnancy, depicted in blue, were as expected those who had received their primary registry eligible dose of primary series vaccine in the early part of pregnancy, either preconceptually or during the first trimester.

For our phase two or extended follow-up interviews conducted 15 months after the end of the pregnancy, of over 21,000 phase one participants that were eligible, we were able to reach 44%. Similar to phase one, the primary barrier to data collection was having someone answer the call from the interviewer. When the participant was reached live by phone, the vast majority completed the interview, nearly 97%, which was also similar to phase one. Methods used to optimize reaching participants by phone included having an assigned caller ID to the interviewer caller line, following call attempts with texts with callback information, including asking for convenient times to try to reach the participant again.

The very simplified, overly simplified planned data flow for phase one, which is similar for phase two, is depicted here with the interview data on the left generating preliminary early outcomes such as for reported spontaneous abortion or miscarriages, followed by acquisition of abstraction and review of medical records to enhance the interview data. With this more comprehensive data set of interviewing medical record data, then producing more detailed final outcomes. On the right, you can see the insertion of the clinical review work involved in reviewing and better understanding the interview data and its corresponding medical record data. This was a particular importance for birth defect adjudication and categorization.

One note particularly about one of the most important outcomes of this work, birth defects. The process for this has been fairly labor intense and involved. We classified all reported fetal and infant health conditions as major, minor, possible/probable birth defects, or as not a birth defect. All birth defects were then coded using the MACDP process. This is what CDC uses for its routine birth defect surveillance. All pregnancies with any outcome including live birth, stillbirth, abortions, or miscarriages were included. And we have now, through phase two interviews that have just been completed, included any birth defects identified and reported through 15 months of age.

I'll move now to the registry's accomplishments and challenges. In the interest of time, I'll not be presenting specific results as that is not the focus of this meeting. Links to the publicly available data from the registry are provided at the end of this presentation in the resources slide.

The registry allowed for monitoring of reported outcomes as the enrolled participants progressed through their pregnancies and interview data accrued. Excuse me. Proportions of reported outcomes were compared to expected background rates and monitored for any deviations from expectations. We were able to provide regular data updates to the Advisory Committee on Immunization Practices COVID-19 Vaccine Safety Technical Work Group or VAST. We provided public presentations at ACIP meetings in the fall of 2021 and 2022. And there were publications on the preliminary findings of the registry in The New England Journal in 2021 with the information about no noted increase in the rate of spontaneous abortion or miscarriage used to be able to strengthen the COVID-19 vaccination recommendations made by CDC to vaccinate pregnant persons during any trimester in August of 2021.

Taking a broad view of this type of registry, in the unusual setting of a global pandemic and global vaccination program, some of the advantages were its flexibility, its potential for medical record confirmation of consenting participants' self-reported data and the relative speed for some of the earliest outcomes in pregnancy, such as spontaneous abortion or miscarriage. However, there were

some significant limitations. Obviously the cohort was biased toward the earliest eligible recipients and adopters of vaccine, which were largely healthcare personnel. It proved to be a very time and resource intense effort involving large volumes of data from two sources, both interviews and medical records. And importantly, was based on a convenient sample with no control group externally, which then necessitates comparison to available background rates of studied outcomes.

Our real-world experience highlighted some important issues to consider in future similar efforts. We did not know what stage of pregnancy someone was in at the time of vaccination from v-safe data alone as that data were not collected, so we were unable to order calls based on gestational age, which would've been very helpful. That meant many participants had already delivered by the time of the first interview. Over half of the participant reports of infant birth defects required medical record confirmation, not surprising, for further clarification as these standardized questions were not effective in eliciting the type of more detailed information during an interview that would allow birth defect subject matter experts to appropriately categorize the birth defects from interview data alone.

Medical record acquisition abstraction was very resource intense and of variable quality based on the records received. And there's an important balance to be struck between the sensitivity and specificity of the information. For example, in the safety surveillance system, one would want to include possible/probable birth defects, but this is not routinely done in national birth defect surveillance, making background rate comparisons with those data particularly problematic. And some outcomes are just not as easily captured as others either by interview or medical records.

More specifically, challenges that were unique to the interview data included the impact of more open-ended questions to improve the feasibility of interview completion meant that we captured a broad range of diagnoses, which was helpful for hypothesis generation, but resulted in decreased standardization. Use of text fields to allow for some participant responses resulted in coding challenges and some unusable data. Participant report is not the standard approach for complex medical conditions, which then required a robust protocol for classification to use the data for analysis. There are sometimes nuanced details needed for clinical adjudication and classification of outcomes that are simply not available from use of a standardized interview format.

There was some inconsistency in how participants responded to interview questions, particularly for some outcomes such as birth defects and hypertension. For example, reporting birth defects only when asked about the infant's clinical referrals rather than as a direct response to a question about birth defects or infant medical conditions, or reporting gestational hypertension or diabetes from a prior pregnancy as a chronic medical condition, which was information collected for confounding analysis. Another issue was that there was uncertainty of certain conditions that had been noted during pregnancy had resolved in-utero or were still present after birth, which particularly impacted birth defect classification. And again, this proved to be time and training intense, which impacted the speed of available results.

While medical records are often thought of as the gold standard for clinical data, there were several notable challenges in their use. Records were not always obtainable, that is there may not have been consent provided, the records were not found or they were not sent by the facility. There were also variable facility requirements we encountered for medical record release that delayed or prevented acquisition. Incomplete records were often obtained. For these outcomes, prenatal delivery and outpatient infant records were needed and constituted what we would consider a full set of records. However, often not all of these records were available.

There are discrepancies within the records, including dates of diagnoses, diagnostic terms used by providers, and repetitive procedures such as ultrasounds that had different results. And there were handwritten records and lack of standardization across multiple electronic health record platforms,

making use of what would seem a simple record, what I was used to, the prenatal flow sheet, sometimes more difficult than anticipated. It was often written in narrative form rather than graphic form and a bit on the confusing side to interpret, I think mostly a result of EHR use. Again, this proved to be a resource intense area, particularly with training, re-abstraction of records, data quality checks, comparisons of abstractions and feedback to improve the quality.

I'd like to close with some thoughts about how we might approach this differently in the future for consideration. Clearly defining outcomes of interest that are ideally narrow in scope would have been helpful, particularly for interview questions. Importantly, this requires prior determination of a more limited set of study outcomes that may not be appropriate for all purposes, like a global vaccine rollout. Minimizing the number of text fields, planning specific analyses prior to implementation to guide data collection, including specific definitions of key variables such as preexisting medical conditions that might be confounders for outcomes. Recognizing, acknowledging, and planning for likely discrepancies within data sources such as hypertension documentation in the medical record that will require additional decision-making. The two most challenging outcomes we found were hypertension, both preexisting and gestational for participants and birth defects for infants. Birth defects is a complex outcome of high importance in these types of studies and often requires comprehensive medical records and interpretation for confirmation.

Inclusion and exclusion for multi fatal pregnancies is important, for example, how to handle fetal reduction scenarios like vanishing twins. Having clear and standardized definitions of birth defects is critical. Use of participant report alone, while fast, is highly challenging to standardize and interpret. And the use of medical records available after the first year of life is standard process for routine birth defect surveillance. So challenges in acquiring records delays data collection and results and must be factored into timelines. And finally, having an available control group is important as reliance on background rates, especially when those may be fluctuating such as during a pandemic, is limiting. We did take a look at some internal comparison groups based on the timing of vaccine receipt, but for the most part have been using background rates.

I'd like to thank the registry participants who generously gave of their time and information as well as the large group of people behind this effort. And this is the resources slide that has links to our publications and presentations. Thank you.

Geeta Swamy:

As Dr. Kharbanda is getting ready, I just want to remind folks they can put questions into Slido if you have them as they're coming up and for our discussion for after.

Elyse Kharbanda:

Okay. Thanks for having me here today to present my Stakeholder Perspective: Evaluating the safety of vaccines administered during pregnancy in the Vaccine Safety Datalink. And I just want to acknowledge that I am the last formal presenter before lunch break, so thanks for bearing with me.

As a disclaimer, my work is supported through contracts with the CDC, but what I present, these are my views alone and not official positions of the CDC. And any mention of a product or company name is for identification purposes only.

So as an overview to this presentation, first I'm just going to talk a little bit about where I work at HealthPartners Institute. Then I'll provide background on the Vaccine Safety Datalink, including an introduction to our data structure and provide some information on some studies that we've conducted in the past on vaccine safety in pregnancy. Then I'll provide my perspective on postapproval monitoring of vaccine safety in pregnancy.

So first about HealthPartners, we are the largest consumer governed nonprofit healthcare organization in the nation. We were founded in 1957 and we're based in Bloomington, Minnesota. As an integrated health system, we include a health insurance plan with about 1.8 million medical and dental health plan members. We're also a care delivery system based in Minnesota and western Wisconsin. We have each year about 1.2 million medical and dental patients. And since we're talking about pregnancy, I'll say we have about 12,000 prenatal care visits in our system each year with a live birth in our health system.

And now about HealthPartners Institute. So we're a nonprofit institute within the larger HealthPartners organization dedicated to high-quality public-domain health research. The image on the left is screenshot of our webpage. There are a number of investigators who work at HealthPartners Institute who conduct work related to postapproval drug or vaccine safety in pregnancy. And some of those projects are listed on the slide. But the rest of the presentation I'm going to focus specifically on our CDC funded work and the Vaccine Safety Datalink.

So a little bit about the VSD. The VSD is a collaborative project between CDC's Immunization Safety Office. And we're now up to 13 integrated healthcare organizations and networks across the US and these sites are shown on the map on the slide. The VSD monitors the safety of vaccines in use in the United States, primarily through observational multi-site studies of rare and serious events following vaccination. The CDC was founded in 1990 and HealthPartners joined as a VSD site in 2000.

Now a little bit about the data structure for the VSD. As the name implies, a critical component of the Vaccine Safety Datalink is our data structure. And for any group working in postapproval safety monitoring, working with high quality data is key. Of the 13 current VSD sites, 11 are now data contributing sites. And as such, they create standardized data files that can be used in multi-site studies. We use a distributed data model in the VSD, so all data files reside at the site and programs are run off these data sets with limited data sets transferred to a lead site for specific studies. This site also shows the many files that each VSD site creates. Those in bold, the dynamic pregnancy episode file and the mom-baby linkage have been critical for our work in maternal vaccine safety surveillance.

It was around 20 years after the founding of the VSD in 2010 when we expanded to evaluate vaccine safety in pregnancy. There were many reasons for this shift, and these have already been discussed a bit. So pregnant women were generally not enrolled in clinical trials of vaccines. And when vaccine trials are conducted in pregnant populations, there's often insufficient power to assess rare post-vaccination safety outcomes. And as such, vaccines might be recommended for use during pregnancy with limited data on their use in this specific population. So it's clear that postapproval observational studies are needed, but these can be challenging. As the VSD includes data on over 3% of the US population, and we have comprehensive data on vaccine exposure, standardized data files, and access to medical records as needed, the VSD is an ideal network for conducting postapproval vaccine safety monitoring. An increasing number of vaccines are now recommended in pregnancy, and these vaccines are usually recommended for all pregnant women.

So this slide provides a brief history of vaccine recommendations for vaccines in US populations. Today, the influenza, Tdap, and COVID-19 vaccines are all recommended in pregnancy, and the RSV vaccine may soon be added to this list. It's clear that the past 20 years have been marked by a substantial expansion in the number of vaccines indicated are recommended in pregnancy, increasing the importance of our work in maternal vaccine safety monitoring.

So now I'm going to shift to talk about a few of the studies that we've worked on in the VSD that my group has led. So our first area of a focus was on influenza vaccines. So the study on this slide published in 2014 included nearly 60,000 vaccinated pregnancies from the 2004-05 through the 2008-09 influenza seasons. And these women exposed to influenza vaccine during pregnancy were matched to women unvaccinated in pregnancy. When using a propensity matching and analysis that accounted for

vaccination and pregnancy as a time-dependent exposure, we found that receipt of vaccination and pregnancy was not associated with increased or decreased risks of preterm or small for gestational age birth.

In the second study published in 2017, we evaluated 50,000 live born infants from 2004 to 2013 with maternal first trimester influenza vaccine exposures and found as compared to unvaccinated on first trimester adjusted prevalence ratio for having a major structural birth defect was 1.02 with a 95% confidence interval, 0.94 to 1.10.

A few key points related to our early work on maternal influenza vaccine safety. Overall, these were well powered studies, we were able to evaluate a range of outcomes and we had no safety signals. However, these studies took place before we had access to the dynamic pregnancy algorithm. We required access to birth records, which resulted in a substantial lag between when the vaccine exposures occurred and when we were able to report safety outcomes. Also, we combined data from multiple influenza seasons and ignored potential differences by manufacturer. And finally, at the time of these studies, our data sources for vaccines were limited to electronic health record and claims, and thus, these studies might have missed vaccines, administrative flu fairs, or other community-based venues.

Following recommendations for use of Tdap in pregnancy, we expanded our portfolio to explore Tdap safety in pregnancy. This slide shows one of our first studies on this topic published in 2014. The study included pregnancies ending in a live birth from 2010 to 2012 at the two California VSD sites. Overall, our data was reassuring for SG and preterm birth, but we did have a small but statistically significant signal related to chorioamnionitis with an adjusted relative risk of 1.19 and a 95% confidence interval of 1.13 to 1.26. This raised a lot of questions, including what was the clinical significance of this finding and even why were we studying this outcome? Was it biologically plausible that a vaccine in pregnancy could increase risk for chorioamnionitis? Our findings did not result in a change in vaccine recommendations in the United States, but we did conduct a follow-up study.

Our follow-up study evaluated pregnancies ending in a live birth from 2016 to 2018. The VSD site's chart reviewed all cases with the diagnosis of chorio, and these cases were then clinically adjudicated by obstetricians on our team. We started this follow-up study in 2019, and then the pandemic hit and our work shifted. We were able to finish these analyses and disseminate our study findings in spring 2023. In analyses accounting for vaccination in pregnancy to be a time varying exposure, the adjusted hazard ratio for chorio following maternal Tdap was 0.96 with a 95% confidence interval of 0.9 to 1.03. However, we also found that the ICD-10 code for chorio had a low positive predictive value for actually detecting clinical chorioamnionitis, but it had a positive predictive value of 81% for either clinical or histologic chorio. Findings were reassuring for other outcomes assessed. But similar to our flu vaccine work, this study did not differentiate between the two licensed Tdap vaccines.

Our groups also evaluated inadvertent exposures to HPV vaccine in pregnancy. Data presented to the FDA proceeding the licensure of the 9vHPV vaccine found in the vaccine trials in comparison to trials of the 4vHPV vaccine, there was a potential increased risk for spontaneous abortion when the nine valent vaccine was administered within 30 days of the LMP or pregnancy start. For our studies of inadvertent HPV vaccination and pregnancy, we compared women vaccinated either during pregnancy or peri-pregnancy to those vaccinated during a period 16 to 22 weeks prior to pregnancy, with all potential spontaneous abortion cases chart reviewed and clinically adjudicated.

This slide shows our main findings from our paper published in 2021. We found no increased risk for 9vHPV vaccination in pregnancy or peri pregnancy as compared to distal vaccination for any of the outcomes, including spontaneous abortion, preterm birth, SGA birth, or major structural birth defects. Of note, for all groups, the numbers of HPV vaccine exposed were low and thus confidence intervals are wide.

Both the pace and interest in our work on vaccine safety in pregnancy grew exponentially with the COVID-19 pandemic and authorization of the COVID-19 vaccines. It was around this time that the VSD and work led by a colleague at Kaiser Northwest expanded our ability to identify and date ongoing pregnancies in near real time through the dynamic pregnancy algorithm. In addition, many VSD sites expanded their capacity to integrate immunization information system data from their state registries into their standard VSD vaccine files. This slide shows the many studies our site has published on maternal COVID-19 vaccine safety, including the primary vaccine series and booster doses with several additional studies in the works.

In these last few slides, I'd like to provide some additional thoughts regarding post-approval surveillance of vaccine safety in pregnancy. First, regarding outcome selection, it's important to consider biologic plausibility when evaluating a potential maternal vaccine safety outcome. Is there a plausible explanation for how a vaccine administered during a given time period in pregnancy could cause a given outcome? When considering public health importance, is the outcome a potential concern for pregnant women or their providers and perhaps fears related to this outcome pose a barrier to vaccination in pregnancy? Also, additional sharing of data from pre-licensure trials or other surveillance systems would be useful for guiding our outcome selection. I would also like to advocate that some important vaccine safety outcomes may not be well suited for automated surveillance studies and may require a different approach with some examples shown on the slide.

A second area I wanted to comment on relates to vaccine exposures in pregnancy. First, we learned after completing our Tdap studies that our findings could not be used in the vaccine label as we didn't evaluate outcomes by manufacturer. Starting with COVID-19 vaccines, we have included manufacturer in all of our analyses. Similar to my previous comment about outcome selection, additional sharing of data from animal or phase one studies could potentially inform our choice of exposure windows. And finally, now that we can incorporate state immunization information system data into our VSD data files, we can attest to its importance. At HealthPartners, we found 25% of our data on COVID-19 vaccines for our full VSD population were only found in our state immunization data.

Third, moving forward, we hope that timelines for reporting results of our maternal vaccine surveillance work are realistic. We need time for vaccine data to be available and incorporated into our VSD files. We also need data used in the dynamic pregnancy algorithm that helps us identify and date pregnancies. We need for it to have time to mature. And finally, we need time for pregnancy outcomes to occur. With our work on COVID-19 vaccines, we were often asked to report on first trimester vaccination and risks for birth defects before these pregnancies were complete, before any potential birth defects could have been confirmed.

And finally, it's critical when conducting postapproval maternal vaccine safety surveillance, that analytic approaches to minimize bias should be used. Some common sources for bias are shown on the slide. It's also important to note that risk for outcome vary by gestational age of vaccination and vaccine availability can vary by season. Furthermore, these optimal analytic approaches should be applied when data is incorporated into meta-analyses.

And with that, I will end and I want to thank the large team and CDC that supported this work.

Geeta Swamy:

Thank you to all the panelists. And so we're going to open up for discussion. We'll just see if anyone in the room has any questions to get us started with? There's one over to the back here. Yeah.

Alex Hillman:

Hi. Yes. So I just wanted to maybe ask a follow-up on something that Dr. Chambers raised in her remarks and something that I think many of your presentations might support, which is this idea of a national pregnancy registry. And I think sounds like a great idea in theory, has a lot of challenges in implementation. So I'm curious if maybe you, Dr. Chambers, or others have thoughts around what that might look like, where it might live, who might staff that, and any other considerations? I know there's limitations, but just some initial thoughts on what that might look like.

Gerrit Hamre:

Can you also share your name?

Christina Chambers:

Well, that's the-

Alex Hillman:

Sorry. Ms. Alex Hillman. I'm a senior program officer with the National Academies.

Christina Chambers:

... I think that's the million-dollar question. The idea I think is appealing for all of the reasons that we worry about single drug pregnancy registries being costly and having a hard time raising awareness in the population. But I think Jessica's presentation points to the complexity of doing this. And that's for one class of medications, let alone all. But honestly, I think we're kind of all doing something like that. It's just that it's not interconnected. So there isn't, other than somebody calling you up and saying, "What are you guys seeing in what you're doing compared to what we're doing?" There isn't an infrastructure to link all of these disparate things that are happening together. And I think it's doable. And I don't know where it should sit. I am agnostic about that. It should sit in heaven.

Christine Olson:

They have the resources.

Christina Chambers:

Yeah, that's right.

Jessica Albano:

I mean there's certainly the funding aspects and the logistics of it on top of where does it even belong. But I think there's probably some steps, some intermediary things that could be done now just as far as looking at common protocols and standardization of outcomes and the way that outcomes are assessed and evaluated and coded, that would make at least even pulling the data, even if they do sit in different places, if they're collected in a more standardized way, that could make them more amenable to meta-analyses or at least capitalizing on the numbers in a collective way through pulled analysis. So that's something that could be done on a much more near future kind of timeline.

Geeta Swamy:

I have a question, but I was make one comment about that is that using common definitions seems like such an easy thing. We've been struggling for that for years across all of research, not just in postapproval through spaces. And I think one of the things that you said, Christine, about hypertensive

disorders, that seems pretty clear, right? We can see what hypertension is or not. But I will tell you some of it is the reliance on clinical care because it is hard for us to make the diagnosis much less document it because-

Christine Olson:

It's evolutionary diagnosis, which makes it harder.

Geeta Swamy:

Exactly. Right. So sometimes I wonder if we should be looking at treatment or medication use, right? If there's an anti-hypertensive used, they must have had a condition, but it might not have met the exact definition to meet the ICD coding. So some of those things that we need to be thinking about as we go to some of these other strategies I think is-

Christine Olson:

Something like that is done for hemorrhage in quantifying blood given because you cannot estimate blood loss. I mean, the surgeon always says. Well, or is close to.

Geeta Swamy:

Well, that's because we're all great.

Christine Olson:

Right, right. Yeah, so using something quanti... I agree. I think there's has to be something more objective to measure that.

Geeta Swamy:

Well, one question I was wondering, Dr. Albano and Dr. Chambers, you both of your discussions on your registries had really incredible follow-up rates, or I should say loss to follow-up being really low. Are there any successes or sort of pearls to give to folks as they're thinking about this, if they're endeavoring into these areas, research or data?

Jessica Albano:

Go ahead and I'll add on.

Christina Chambers:

We've always said, and I believe this to be true, is anecdotal, is that it's because of engaging the mom. So she's committed to taking part in this. That being said, not all people consent, but once consented, once they're in the study, they develop a relationship with the person who's calling them, the person who's collecting that information. And I think we're very successful then in retaining them in the study.

Jessica Albano:

I would say, I mean the Antiretroviral Pregnancy Registry is in a little bit of unique situation because it's anonymously reported by healthcare providers. So as long as they know the outcome, they are able to provide it. So there's less of a attrition due to loss to follow-up. With that said, I mean in keeping the providers engaged and actually adding value back to them so that they are receiving information that's actionable to them and their patients and use in their practice, I think contributes to that significantly.

Geeta Swamy:

Question.

Janet R Hardy:

Thank you. Wonderful discussion. This is Janet Hardy, most recently from Biohaven and Pfizer. So asking from the industry perspective, and since a lot of this translates into what goes into a PMR from the FDA, I love the idea of these disease-based large registries and efficient, how to make it happen? Jessica, the Antiretroviral Registry has run for an extraordinarily long time. That in itself is productive for information, but also daunting from an industry perspective. And I think if I'm not mistaken, that list you have provided of industry members, they have come and gone. It's not continuous participation throughout. Is that correct?

Jessica Albano:

For some, GSK being the originator, ViiV now owns all their antiretroviral products, so they have been part of it. Other innovator manufacturers, as they have developed their initial products and come to market, have been a part of it. Where we do see turnover is through the generics because they do have a lot more volatility in the drugs that they have. So it does turn over to a small degree.

Janet R Hardy:

Okay, thanks. And in terms of baby steps along the way to the giant registry and heaven idea, Tina, thank you. There is the possibility for some of the larger manufacturers to look at their product list and consider whether there can be disease-based within or class-based within their own products and start looking at efficiencies and innovation from that perspective. But I also want to take this back to the FDA and ask somebody from the FDA to speak to what regulatory authority you have to be moving this ahead and asking for this type of requirement since we're talking about PMRs and PMCs? Thanks.

Geeta Swamy:

I think that will be forthcoming in some of the presentations for sure.

Gerrit Hamre:

If we sorted out the issue that we had in the first presentation. Do you see if you have any questions that came in from virtual participants? I think we've got it sorted out. If not, we'll run something else up to you.

Geeta Swamy:

I don't know. You might be able to look at it and detect for me.

Gerrit Hamre:

All right, I'm going to grab it from you. While, the next person's speaking, I'm going to grab it for you and I'll get it. Just a second.

Megan Clowse:

Hi, this was really terrific and it's so great to hear about vaccines, which you actually know when people get. Oh, this is Megan Clowse, a rheumatologist. It sounds like these different registries recruit patients

really quite differently, and that all brings pros, cons, biases, and I would love to hear some thoughts about how you all think optimal ways to recruit patients and pregnancies into these registries?

Christine Olson:

I can start. So obviously with the pandemic, there was a lot of press information out there. People were very aware of what was going on. I mean, I think that's probably not the poster child for how to do this on a regular basis, but it did provide sort of a mechanism by which people were able to engage. And we still had a hard time getting the original v-safe platform up and going, getting notification out to vaccine centers, the drive-through vaccination places, the major universities, the other places, clinics where people were getting vaccinated. We provided communication tools and I think there was a bit of a slow uptake for that. I'll say when I got my vaccines, I didn't see it. I looked for it and I brought it with me the next time. I said, "What about this?" So we had that availability.

When we've been talking recently about how would you do something similar, not a registry, but how would we even engage people in using v-safe for RSV? I think one of the easiest things, at least on paper, would be to engage with prenatal care providers and just having that messaging going out and having QR codes that people could use while they're sitting in the waiting room for periods of time. Not that that's where they're going to get their vaccines necessarily, but it might be. I mean, there's that disconnect as well. The prenatal care places are not necessarily where these vaccines are going to be administered in those 10-minute appointments. But to me, there's an opportunity there, especially if it's a vaccine for pregnant people to partner with those groups and using the parent organizations to try to get that messaging out. But I don't think there's a super great answer. I think there are probably multiple pathways to try to get engagement.

Christina Chambers:

I agree. I think we do direct to consumer because we tap into people who would otherwise never come to us. So we do social media, but we also recruit through healthcare providers, as you know, Megan, because rheumatologists, for example, are motivated to do this. But we still target the mom for the reasons that I've mentioned before. I think you mentioned the partnership. We're working on a partnership with the Maternal Fetal Medicine Network to help recruit participants into pregnancy registries for the reason you said, and also for the reason of trying to increase diversity since that patient population represents a much more diverse population than who contacts us spontaneously.

Jessica Albano:

So I would add to that in addition to kind of educational aspects for both physicians and patients looking at distribution networks and works particularly well in the case of vaccines. We ran a couple of vaccine pregnancy registries that actually tapped into military bases where they knew their vaccines were being distributed and were able to collect large populations of exposed women there.

Geeta Swamy:

Any other questions?

Gerrit Hamre:

Nothing on mine.

Geeta Swamy:

Yeah. Okay. Go for it. So I would just ask you all as individuals who have been involved in the design and the conduct of the studies, if there are ways that you would think about as key to how you might do it differently? You all talked about challenges in sort of ways, but if you could wave your magic wand, what would you think about doing differently?

Christina Chambers:

I'll say, and this is a duh, that we create a really high bar for people to participate. So it's great when they do, but we think and are working on some ways to do this, of testing different incentives that would appeal to people. Altruism is the major incentive why people participate, but that is probably not the only one. And we need to test those. And we also need to create a pathway to participation that doesn't require that you sit on eight hours of telephone interviews and complete a whole bunch of paperwork. And so can we get what we call pregnancy registry light information that's informative and maybe a subset gets the more in-depth information. So that's where I would see some opportunities.

Jessica Albano:

I would say standardization and simplification. If you think about a healthcare provider who might be interested, even on the obstetric side, there are, if you look at the FDA website, I think well over 100, maybe 150 or even more pregnancy exposure registries available to report to. Well certainly, they're all going to have nuances with regard to what information they're collecting, how they're collecting it, what systems they're using, data elements, et cetera. So it's really a lot to ask for essentially voluntary work on already busy clinicians. So anything that can be done to streamline that would be very useful.

Geeta Swamy:

I think the burden on patients, pregnant individuals is significant, but the burden on providers, given all that we know about healthcare setting, physician burnout and so forth, is a tough one to rely on as well. Elyse?

Elyse Kharbanda:

Yeah, I'll just add, along with having the manufacturer data available so that we can always do the analysis by manufacturer, that I would've loved to always have more time to validate outcomes before we study them. That often our approach is we conduct the studies, we think that we have good outcomes, and then when we have a signal, then we say, oh, was this a good outcome and what was the positive predictive value? And then we do a validation study after like we did for chorio. And so I also wish sometimes we just said no and didn't get pushed to study outcomes where we felt like our data wouldn't work with this outcome and just suggested other alternative approaches should be used.

Geeta Swamy:

And I think one other key thing there is, you're right, we had lots of conversations about the Tdap and chorio, and I think even looking at some of the data that we talk about as confounders, while it's important, I'd also look and think, well, why would we think that any vaccine is going to cause gestational diabetes? We know what the underlying causality of that is. Is it really a factor we should be even collecting or worrying about, right? If we have birth weight, you can see that there might be very significant LGA or something like that. So I do think we have to get smarter about what it is. Because we can talk about all kinds of ways to incent providers, patients, whatnot, but at the end of the day, people want to be efficient and timely and still get their information. So I think as researchers and clinicians, we're going to have to figure out how to do that in better ways as well. All right.

Gerrit Hamre:

One more quick question.

Geeta Swamy:

All right. Anyone out in the audience?

Lee Cohen:

Just Lee Cohen, Massachusetts General Hospital and PI on the National Pregnancy Registry for Psychiatric Medications. Tina, I just wanted to follow up with a comment that you said. I think we have an incredible opportunity to engage potential subjects digitally that we didn't have before. And frankly, during the pandemic, we found that our ability to procure medical records digitally, because we had to do it, we did it on the fly. And it was a terrific lesson that we were able to get quality data across a very large sample of women. So I agree with you. I think that we have to get creative in terms of really the issue of engagement as part of our recruitment efforts.

Christina Chambers:

And the way that people want to be engaged and to obtain information, that we feel pretty confident you can get that way without having to probe and help somebody.

Lee Cohen:

Exactly. And also collaboratively working with IRBs as we sort of move to using digital platforms to conduct the sort of science that's really going to inform clinical care.

Geeta Swamy:

All right, thanks everyone. Thank you all for your presentation.

Christina Chambers:

Thank you too.

Geeta Swamy:

I think we're close to the [inaudible 01:27:10].

Christina Chambers:

Thank you.

Session 3

Megan Clowse:

All right. I hope everybody had some sustenance for lunch and we're going to dive into the afternoon. So now we're going to come up on session number three. This is going to be the FDA's considerations for construction, the Pregnancy Safety Study framework. We're going to be hearing some data that they have gathered over a recent time to show us what is already existing. So for session three, we're going

to hear from several FDA presenters on recent research that has informed their pregnancy safety framework, including drug utilization data and others that have helped form their key considerations.

So all of our speakers are within the Office of Safety and Epidemiology, and the Center for Drug Evaluation and Research or CDER. First we're going to have Wei Hua, Deputy Division Director in the Division of Epidemiology 1 within that office. Next is Adebola Ajao, who is a lead pharmacoepidemiologist within the Division of Epidemiology 2. Next is Aida Kuzucan, who's also a pharmacoepidemiologist with the Division of Epidemiology. And finally, José Hernández-Muñoz, a senior pharmacist and pharmacoepidemiologist with the Sentinel Core team, which is part of the regulatory science office within the Office of Safety and Epidemiology in the Center for Drug Evaluation and Research within the FDA. So first we have Wei.

Wei Hua:

Thank you, Megan.

Megan Clowse:

I think your mic should be on.

Wei Hua:

It's working.

Megan Clowse:

Yeah.

Wei Hua:

Yeah. All right. Good afternoon. My name is Wei Hua. I'm the Deputy Director of the Division of Epidemiology 1 in the Office of Pharmacovigilance and Epidemiology, office of Surveillance and Epidemiology CDER, FDA. Session three will discuss FDA's considerations for constructing a pregnancy safety study framework. We have a series of FDA presentations in this session, and I will start with an introduction to the topic. As mentioned by Leyla this morning, the general postapproval approach to assessing pregnancy safety include routine pharmacovigilance and the non-interventional or observational studies on pharmacovigilance, sorry. Yeah.

Routine pharmacovigilance refers to the review of spontaneous reports or case reports or case series from medical literature or other sources. Non-interventional studies include pregnancy registry studies, healthcare database studies, and the descriptive studies. Pregnancy registry studies are the prospective cohort studies with primary data collection. Healthcare database studies use electronic healthcare data such as EHR claims data. Descriptive studies may use primary data collection or electronic healthcare data, but different from the registry and the database studies, there's no comparator or sample size requirements. In parallel with routine pharmacovigilance, non-interventional studies are commonly used to generate postapproval pregnancy safety data to inform regulatory decision making. However, when and what non-interventional studies should be used and how they can be used efficiently remain a question. To address this question under PDUFA VII, FDA is committed to develop a consistent and transparent approach to help decide when and what postapproval pregnancy safety studies might optimally be used to obtain timely evidence of safety for regulatory decision making. Is this slide by slide?

The framework focuses on postapproval non-interventional studies to assess the safety of maternal exposure to drugs or biological products during pregnancy. Routine pharmacovigilance, clinical trials, studies on efficacy, paternal exposure or lactation, and operational issues are out of the scope of this framework. The framework also does not address labeling, benefit-risk assessment, or clinical practice. However, safety data generated from the studies under this framework in conjunction with other safety data such as routine pharmacovigilance, may inform regulatory decision making and the clinical practice.

So this is the language from the PDUFA VII commitment letter. To learn from the current state and to inform the development of a future framework, FDA is also committed to conducting two analyses. One is on how different types of post-market safety data, safety studies have been used by FDA and the industry. The second is to conduct a review of types of post-market pregnancy data that have been included in pregnancy labeling. So these analyses will be presented by Dr. Adebola Ajao and Dr. Aida Kuzucan in this session. After that, Dr. José J. Hernández-Muñoz will present a preliminary analysis of drug utilization data, and I will come back to present an overview of considerations to optimize the use of post-approval non-interventional pregnancy safety studies. So with that, I'll return it over to Dr. Ajao.

Adebola Ajao:

Good afternoon everyone. My name is Adebola Ajao and I'm an epidemiologist in the Office of Surveillance and Epidemiology within the Center for Drugs and Evaluation Research. Today, I'll be presenting our analysis of postapproval pregnancy safety studies associated with FDA approved products, drugs, and biological products. So here's, the FDA study team. This work is a collaboration between many individuals across multiple offices in the Center for Drug Evaluation and Research and the Center for Biologics Evaluation and Research. So this is the outline of my presentation today. I'll actually skip over that PDUFA VII commitment as this has already been introduced by Dr. Sahin and Dr. Hua. I will go over some background information on post-marketing requirement and commitments. Then I'll describe our analysis of postapproval pregnancy safety studies associated with FDA approved products. For this section specifically, I'll go over the study objectives, study design, data sources, study methods, the variables that we collected, and our study results. Then I would end with a summary of our analysis and some study limitations.

So postmarketing requirement or PMRs or post-marketing commitments or PMCs refer to studies and clinical trials that applicants conduct after approval to gather additional information about a product safety, efficacy, or optimal use. Specifically, PMRs are studies and clinical trials that are required... the applicants are required to conduct under statutory authority while PMCs are studies or clinical trials that an applicant agreed upon in writing with FDA. And those are reportable under 506B of the Food Drug and Cosmetics Act, the FDCA. Under section 505(o)(3) of the FDCA, postmarketing safety studies and clinical trials are required to assess a known serious risk related to the use of the drug, or assess signals of serious risk related to the use of the drug, or identify an unexpected serious risk when available data indicate the potential for a serious risk.

So onto our analysis of postmarketing pregnancy safety studies associated with FDA approved products, our study objective is to understand how different types of postmarketing pregnancy safety studies are being used by FDA. So over the next few slides, I would describe characteristics, status and impact of postmarketing pregnancy safety studies, assessing maternal fetal and infant outcomes. Our study is a cross-sectional descriptive analysis of human postmarketing pregnancy safety studies.

So for data sources, we query three data sources for pregnancy related postmarketing studies. Studies associated with drugs and biologic products approved by FDA as of May 31, 2022. The first data source we used was the FDA PMR/PMC database, which is an FDA internal system for PMR/PMC studies. For this data source, we included studies where the patient population is pregnant or population is pregnant

women and the study or trial include pregnant or pregnancy, or the PMR/PMC description mentions pregnant or pregnancy. For the second database, the FDA Office of Women's Health Pregnancy Registry, this is a database that includes open registries or pregnancy registries that are closed within one year of our data poll.

And the third database, the clinicaltrials.gov and this is a registry of clinical studies maintained by the National Institutes of Health. For this data source, particularly, we use the search criteria, pregnant, pregnancy, observational, and we limited to studies associated with a drug or biologic. Note that all of these three data sources are convenient samples and they do not represent the total universe of pregnancy safety studies.

So for our methods, we combine and deduplicated studies from the three data sources, and we included only studies initiated postapproval to monitor maternal, fetal and infant outcomes associated with exposure to FDA approved drugs or biologic products in pregnancy. We excluded studies that did not fit our inclusion criteria like animal studies, toxicology studies, pharmacokinetic studies, cross-reactivity studies, and lactation studies. For study variables, we collected study type, study goal, therapeutic class, study status, the year study was established, reasons for study termination, and if a labeling update occurred. So, this is our study search results and our units of analysis is our studies. So for both drugs and biologics, 310 pregnancy studies were identified from the FDA PMR/PMC database, 50 pregnancy registry studies from the FDA Office of Women's Health Database, and 60 pregnancy studies from clinicaltrials.gov.

After combining and deduplicating, 394 studies remain, of which 61 studies were excluded for reasons that I previously discussed. The final study that consisted of 333 pregnancies studies representing 270 products. For study type, of the 333 pregnancy studies, two-thirds or 63% were pregnancy exposure registry study or PER. As it's been clearly defined up here, but I'll repeat it, a pregnancy exposure registered study is a prospective observational study that collects exposure and pregnancy outcomes from women exposed to a product of interest shortly before or during pregnancy. Outcomes in exposed women are compared to outcomes in an internal control group of disease unexposed women or an external control group like the general population. So second we have, this was followed by descriptive pregnancy safety studies at 21%. And just to define that, I know it's been previously defined, but I would just say it again. Descriptive pregnancy safety study is a single arm study of exposed women with no comparator or sample size requirement. In our study protocol, enhanced pharmacovigilance and surveillance programs were actually classified under descriptive pregnancy safety studies.

Next, we have database studies with pre-specified outcome at about 15%. A database study with pre-specified outcome is a retrospective observational study with electronic health data, such as electronic health records or medical claims to assess the association between products exposure during pregnancy and major congenital malformations and other adverse pregnancy and infant outcomes. Last, we have randomized clinical trial at about 1%. So if you look at this data, it's no surprise that majority of the studies have pregnancy exposure studies because this is the study type primarily and traditionally required by FDA up until 2014. The FDA public meeting on postapproval pregnancy safety studies and subsequent guidance published in 2019, which expanded the scope of postapproval pregnancy studies to include complementary study designs like database studies. While descriptive safety studies are used when exposure in pregnancy is expected to be low or rare.

So for study goal, 99% of the 333 pregnancy studies were for signal detection. This is defined as studies designed to monitor pregnancies exposed to a medication for possible teratogenic effect when there's little to no prior human data identifying a specific signal. These studies are typically for hypothesis generation and may test one or multiple pre-specified outcomes at once. Then we looked at the top therapeutic class that the studies represent. The most common therapeutic class covered by the

pregnancy safety studies we reviewed is psychiatry at about 21%, followed by neurology at 18%, prophylactic vaccines and related biologic products at 14%, dermatology at 8%, metabolic at 7%, reproductive of genetics at 5% each, and rheumatology at 4%. There were about 59 other studies that account for 18% of studies covering 14 other therapeutic areas including oncology, hematology, and infectious diseases.

For study status, as of July 2023, when our study ended, of the 333 studies, 38 studies or 11% have been terminated, meaning the studies were ended before study completion. 65 studies or 20% have been completed, while 230 or 69%, which represents the majority of the studies, are still ongoing. So since majority of the pregnancy safety studies are ongoing, we wanted to better understand when the studies were established. To do this, we focused on 242 PMR/PMC studies shown in the red box for which we have reliable date for when the pregnancy study was established.

So this is a graphic of the year PMR/PMC studies were established. Just to orient you to this graph, the X axis represents the year the studies were established, while the Y axis represent the count of PMR/PMCs established per year, this data shows that majority, approximately 70%, of PMR/PMC studies were established in the last 10 years, with most studies established in the last five years. In 2022, we recorded a smaller number of studies established compared to previous five years due to our May 31 date or cutoff date. So based on this data, it's not surprising that majority of the pregnancy safety studies are still ongoing indicating that these studies have not had sufficient time to mature, be completed and inform any regulatory action.

So a review of the 38 studies that were ended before the studies were completed showed that 24 studies or 63% ended due to low enrollment, which is the low enrollment with the study designs. The study designs were deemed not feasible to obtain the needed data. Of the studies that were terminated for low enrollment, 58% were replaced by a new study type, typically descriptive pregnancy safety study. An analysis of these studies also showed that it took an average of eight years from study establishment to termination. A review of the 65 completed studies showed that majority, at 68%, were pregnancy exposure registry studies, followed by descriptive pregnancy safety studies at 18%, and database study with pre-specified outcome at 11%. This is a similar trend to what we observed in the overall study results.

For the 65 completed studies, 30 minutes... Sorry, I was looking at that. 30 studies or 46% resulted in a safety labeling update to the human data part of section 8 of the prescribing information. For studies that resulted in a safety labeling update, average time for study establishment to labeling update was 11 years, with a range of six to 18 years with majority of the time spent on protocol development and study conduct. This is not surprising as majority of completed studies, our pregnancy exposure registries a study type that's historically taken many years to complete due to the consenting process and the prospective enrollment nature of the study. At this time, we do not have sufficiently robust data to compare time to labeling across study designs. Thirty-five studies or 54% were completed and did not result in labeling update. Labeling is done on a case-by-case basis. We did not analyze reasons why labeling was not updated for these studies that were completed but did not result in labeling update. However, potential reasons may include recent study completion or need for corroboration of study results by additional studies.

So in summary, majority of pregnancy safety studies are pregnancy registry studies. Complementary studies like database studies are fewer as they are newer FDA requirements, and descriptive pregnancy safety studies are used when exposure in pregnancy is expected to be low. A small proportion of pregnancy safety studies have been completed. Majority of studies were established in the past five years and are ongoing. Approximately, half of the completed studies have resulted in safety labeling update. Average time from study requirement to labeling update is 11 years. Low utilization is a

recurring factor in some disease areas. As a small proportion of pregnancy safety studies have been terminated with majority of the terminated study deemed not feasible due to low enrollment of exposed pregnant women.

So our studies does have a few limitations. One is that the three data sources are convenient samples and do not represent the universe of pregnancy safety studies. So our results may have limited generalizability and majority of the studies reviewed were established in the last 10 years, but with peak really in the last five years. So most studies have not had sufficient time to accrue enough exposed pregnancies and the studies have not had time to mature and be completed to inform any regulatory action or even for us to even compare regulatory actions across study designs. I would like to acknowledge all the members of the FDA pregnancy safety working group that contributed to this work. Many of them are in this room. Thank you.

Aida Kuzucan:

Afternoon. Okay. My name is Aida Kuzucan, and I'm an epidemiologist at the FDA Center for Drugs, Division of Epidemiology. Today, I will tell you about an analysis we conducted in response to a commitment under PDUFA VII to describe the types of pregnancy data that have been included in pregnancy labeling since the Pregnancy and Lactation Labeling Rule, abbreviated here as PLLR, was passed in, was final in 2014. This study was a team effort with contributions from staff from the Division of Epidemiology and the Division of Pediatrics and Maternal Health. Here is an outline of today's presentation. During the next 20 minutes, I will tell you briefly about the analysis we conducted.

The primary objective of our analysis was to examine sources and characteristics of quantitative human pregnancy safety data included in the human data subheading in the pregnancy subsection of product labelings in formats consistent with the Pregnancy Lactation Labeling Rule again, which I will refer to as PLLR for the rest of the presentation. As one may recall, the PLLR removed the pregnancy letter categories A, B, C, D, X and instead requires narrative explanations of risk and supporting data in the pregnancy subsection of the label. Thus, PLLR-compliant labeling, which often references specific study results provides an opportunity to more efficiently identify the studies that informed labeling.

To achieve our objective, we conducted a cross-sectional descriptive study of a convenient sample of PLLR product labelings. We used standardized extraction sheets and created data sets in Excel to tabulate data and conduct our analysis. We used quality checkpoints to ensure accurate and consistent capture of data.

The next few slides will focus on the methods used in this analysis. The main inclusion criteria for this study were either approved or converted PLLR product labelings from July 2015 to December 2021. For the purposes of this study, we defined quantitative data as statements that contain numeric values, described the exposures or results, safety related statements from studies in the human data subheading in the pregnancy subsection that were more descriptive or qualitative were not captured. This means qualitative human safety statements that were potentially supported by quantitative studies were beyond the scope of this analysis. Pharmacokinetic studies were sometimes reported in human data subsections and often had numeric values. These studies were included in our analysis only if they supported a specific human pregnancy safety statement. Products with multiple formulations containing the same active moiety are counted once.

Here is the study flowchart. From 2015 to 2021, the FDA reviewed PLLR labelings for 1,795 product labelings. 811 included human safety data in the human data subheading and the pregnancy subsection. Five hundred ninety-five labelings had qualitative human safety data or PK data. Again, I repeat our analysis focused on a sample of pregnancy safety statements with numeric data. So narrowing our study scope to quantitative labelings was due to resource constraints. Two hundred sixteen PLLR product

labelings included quantitative information. Different product brand names containing the same active moieties were counted only once after deduplication based on active moiety. Our study sample consisted of 145 unique labelings.

On this slide, we list the product labeling and study characteristics that were captured in our study. For the purposes of this analysis, study type included six categories, pregnancy exposure registry, database study with a pre-specified outcome, database study with no pre-specified outcome, clinical trial, case report, case series, and other. A two-step study extraction process with two quality control checks was used to generate our dataset. In step one, we characterized product labelings and identified study type and data sources using standardized extraction sheets. Completed extraction sheets for the 145 product labelings were discussed at weekly study team meetings to ensure consensus. In step two, three epidemiologists use standardized definitions to identify study characteristics and data sources. Each extraction was then checked by another epidemiologist. Again, discrepancies were resolved through consensus. There were a lot of meetings. Descriptive statistics were used to describe key features and characteristics of studies and product labelings. This included frequencies and proportions. All analyses were performed using Excel.

Our main findings were as follows. Our sample included new and preexisting PLLR labelings between the years 2015 and 2021. From the sample of 145 product labelings with quantitative human safety data, we identified 177 unique studies. Product labelings had over 300 quantitative human safety data statements. A single study often supported multiple quantitative statements in multiple product labelings, so that's why there's only 177 here. For example, this occurred when the active moiety was an ingredient in multiple combination products. This bar graph shows the distribution of study types identified in our analysis. Each included product labeling could have one or more quantitative human safety statements supported by one or more studies. So the proportions here do not equal 100. Quantitative safety statements from a pregnancy exposure registry occurred in 46% of product labelings. Thirty-five percent had statements from database studies with a pre-specified outcome, 26% from a clinical trial, 10% from a case report or case series. And in 20% of study product labelings we found safety statements from a study type classified as other, which I'll describe further on the next slide.

Studies classified as other included pooled analyses, systematic literature reviews, meta-analysis, statistical or published guidelines and expert opinions. Here you see the distribution of therapeutic classes or organ systems for the 145 product labelings included in our analysis, the most common being the antiviral agents at 28%. This reflects the relatively large number of antiretroviral product labelings with quantitative pregnancy safety statements supported by the Antiretroviral Pregnancy Registry or APR that you guys heard about earlier today. One hundred seventy-seven unique studies were identified, supported the 145 product labelings. Study location is shown in this slide with most at 41%, including the US and other countries. Most studies, 86% included in our analysis were published or had publicly available reports.

Studies that were not published were generally from unpublished proprietary data submitted to the FDA often in a labeling supplement. Most studies, 81% in our analysis, had a study goal consistent with a signal detection study. Okay. In the next few slides, I'm going to present some examples of some commonly used study data sources that we identified in our analysis. Our first example is lamivudine. At the time of approval for the treatment of hepatitis B, the applicant agreed to a PMC for a pregnancy registry to collect information on maternal fetal outcomes. The PMC was fulfilled in 2004 using data from the Antiretroviral Pregnancy Registry and resulted in product labeling as shown in this slide. Overall, quantitative safety labeling for 38 product labelings came from the APR in our analysis, that's 28% of our total sampling. As described previously in a presentation by Dr. Jessica Albano, the Antiretroviral Pregnancy Registry is an international pregnancy exposure registry designed to monitor

prenatal antiretroviral medication exposures and to detect any potential increase in the risk of major birth defects. Interim reports of the APR available on a biannual basis.

This example is of clinical trial data. We identified 37 product labelings with clinical trial data in the human data subheading in the pregnancy subsection. In this slide, results of 10 randomized placebo-controlled trials of metronidazole conducted in 5,000 pregnant women with bacterial vaginosis are described. We identified 21 product labelings that use studies using data from a Scandinavian Birth Register. Denmark, Sweden, and Norway maintain comprehensive birth registers, which have been an important data source for pregnancy studies. For instance, Källén et al in Sweden published a study of SSRIs and persistent pulmonary hypertension of the newborn in 2008 using data from the Swedish medical birth register as shown on this slide. These data are included under the human data subheading and the pregnancy subsection for sertraline, a selective serotonin reuptake inhibitor approved for major depressive disorder in adults and obsessive-compulsive disorder in adults in pediatric patients six years and older.

Our final example uses Medicaid claims data. Each state's Medicaid agency collects enroll and collects data for persons enrolled in the Medicaid and the children's health insurance program or CHIP. These data are standardized and packaged into data sets, which can be used in observational studies. Findings from five studies using Medicaid data were described in 11 product labelings in our analysis. This study by Huybrechts, et al and colleagues is a propensity score stratified nested cohort analysis on major birth defects among 9,258 women exposed to risperidone in the Medicaid analytic extract data. And you're right, you're right there. I picked your paper.

Oh, that's great. Small circles. So, in summary, our analysis show that many different study types and data sources have supported quantitative human data statement in the pregnancy subsection in PLLR labeling. The most common human data study type described with numeric values in the human data subheading in the pregnancy subsection of PLLR labeling was the Pregnancy Exposure Registry at 46%. I listed each of them for you though. The Antiretroviral Pregnancy Registry was the most common source that we identified for the quantitative human data at 26%. Thank you very much.

José J. Hernández-Muñoz:

Okay. Good afternoon everyone. My name is José Hernández. I am a senior pharmacist and epidemiologist with the Sentinel Core team at the US Food and Drug Administration. The title of my presentation today is Preliminary Analysis of Product Utilization Data to inform the Development of the Pregnancy Safety Study Framework. For today's presentation, I will adhere to the following agenda items. Introduce the study objective, explain the product selection process, describe the story design and data used for the analysis, characterize product utilization during pregnancy, and present and discuss the study findings. The main objectives for this analysis were to describe the product selection process used to understand pregnancy exposure levels for assessing the feasibility of using electronic healthcare claims data for pregnancy safety studies. To characterize product utilization during pregnancy, among pregnancies that end in live births and to explore product characteristics that may be used to estimate exposure during pregnancy.

For the product selection process, the starting point was a lease of 249 products. The products in the lease came from studies identified for the analysis of post-approval pregnancy safety presented earlier in this session by Dr. Adebola Ajao. Product with pregnancy exposure ranging from zero to 2,500 during the 15-year query period were not including this analysis. The emphasis of this preliminary query was on product with medium and high exposures. Since low exposure products are not likely to be suitable for study in administrative healthcare claims data systems, which are the focus of the demonstration projects. However, a convenient sample of 28 products with pregnancy exposure less than 2,500 was

subsequently added to the analysis for representation of low exposure products to inform the framework development. In the end, a total of 72 products were included in this analysis. 44 products with pregnancy exposure greater or equal than 2,500 and 28 products with pregnancy exposure less than 2,500.

In this report, we identified pregnancies ending in live birth delivery among individuals between 10 to 54 years of age in the Sentinel Distributed Database and described medical product utilization during pregnancy for the selected 72 products mentioned in the previous slide. We distributed the query to six Sentinel data partners on August 4, 2023. These data partners included four national health insurance, Medicare, and Medicaid. The query period included data from January 1st, 2008 through January 31st, 2023. However, data contribution varied by data partner as shown in the following slide. For example, Medicaid contributed data to the query from January 1st, 2010 to December 31st, 2022, and Medicaid from January 1st, 2014 to December 31st, 2020. All other data partners started contributing data to the query on January 1st, 2008, but their contribution endpoint vary as illustrated in this figure.

A total of 7.3 million live birth were identified during the query period. This live birth were generated by 6.2 million unique individuals. The 25- to 34-year-old maternal age group generated 50% of all live birth deliveries during the query period. The highest proportion of live birth deliveries were captured from 2016 to 2020 as mentioned in the previous slide. This coincides with the addition of all-state Medicaid data to the sentinel distributed data.

Okay, so let's talk about the product exposure results. This chart illustrates the total number of exposed pregnancy for each of the 72 products during the 15-year query period. Each orange bar represents a unique product. For example, the product with the 851 pregnancy exposure describe a product that was used at any point during pregnancy and 851 of the 7.2 million pregnancies included in this analysis for the 15-year query period. A product where 851 pregnancy exposure were identified translating to one exposure for every 10,000 pregnancies during the query period. Even though we excluded from this preliminary analysis most products with low level of exposure during pregnancy, the overall exposure level were low. There were four products where the overall exposure was greater than 100,000 during the 15-year query period, and I should note that these four products have been in the market for more than 20 years as they were approved from 1985 to 2002.

Now let's look at the number of years the product contributed to the analysis. Products approved before 2008 contributed 15 years of data to the query. However, products approved on or after 2008 contributed from zero to up to 15 years of use data depending on their approval date. Three different levels of product use contribution were created for this analysis, zero to six years, seven to 13 years, or 14 to 15 years. The zero to six years group included products approved on or after January 1st, 2016, and there are 11 products included in this group. The seven to 13 years group included products approved from February 1st, 2009 to December 31st, 2015, and there are 11 products included in this group. And lastly, the 14 to 15 years group include products approved on or before January 31st, 2009, and there are 50 products included in this group or 69% of all the products. The next three slides will describe total product exposure during pregnancy by the number of years the products contributed to the query.

This figure illustrate the total number of exposed pregnancy for each of the 11 products contributing between zero to six years of data to the query period. As mentioned previously, all products included in this group were approved on or after 2016. The orange bar showed the total number of pregnancy exposure for each of the 11 products in this category during the analysis, and the blue dots, the number of years each specific product contributed to the analysis. Let me give you an example. There's a group right in the middle of your slide that contributed three years of data to the query, and during that period, that specific product contributed 67 pregnancy exposure for that specific product. The same

information that I just explained for this chart applied to the next two slides. Now coming back to this slide, overall, none of these products contributed more than 4,000 pregnancy exposure to the analysis. In fact, there was only one product associated with more than 700 pregnancy exposure during the 15-year query period.

This figure illustrate the total number of exposed pregnancy for each of the 11 products contributing between seven to 13 years of data to the query. As I mentioned previously, all products included in this group were approved from February 1st, 2009 to December 31st, 2015. The range of pregnancy exposure went from 38 to 23,047. No clear correlation was observed between number of pregnancy exposure and number of years product contributed data to the query. The highest exposure during pregnancy within this group was observed for a product that contributed nine years of data to the analysis and the lowest for a product that contributed eight years of data to the analysis. Of the five products with more than 1000 pregnancy exposure during the 15-year query period, three were approved in 2010, one in 2013, and one in 2015.

This figure illustrated the total number of exposed pregnancies for each of the 50 products contributing between 14 to 15 years of data to the query. As mentioned previously, all products included in this group were approved on or before January 31st, 2009. No clear correlation was observed between number of pregnancy exposures and number of years product contributed data to the query. The range of total pregnancy exposure went from 32, the lowest one, to 294,049, and 47 out of the 50 product contributed utilization data to the whole query period.

Now this figure illustrates the average number of exposed pregnancies by therapeutic class during the query period. The therapeutic classes with the highest average cumulative pregnancy exposure belongs to the sedative hypnotics and antidepressants. Sedative hypnotics are represented by only one product in this quarry and in the antidepressant by 15 different products. The therapeutic classes with an intermediate average community pregnancy exposure were immune suppressants, anti-infective antiretrovirals, dependency, ophthalmologic, antirheumatic, contraceptive, autonomically, cholinergics, antipsychotics, antimigraine, amphetamines, anticonvulsant and antidiabetics. In the therapeutic classes with the lowest average cumulative exposure were antidiabetic, antiobesity combination, antineoplastic, immunosuppressants, musculoskeletal, phospho inhibitors, asthma, dermatologic, antiasthma, autonomic, and immunological monoclonal, which in turn are represented by products that have been approved more recently as shown in the next slide.

This figure illustrates the average number of years of contribution to the analysis by products in each of the therapeutic classes. Only a few therapeutic classes had an average contribution to the product use analysis of less than 10 years. These are antidiabetic/antiobesity combination, asthma-dermatologic combination, dependency/ophthalmologic combination, antineoplastic/immunosuppressant combination, and musculoskeletal/phospho inhibitors. These are mostly the same therapeutic classes with the lowest average number of exposed pregnancies, suggesting that the observed low utilization for some therapeutic classes may be partially explained by their limited number of years contributing to the analysis.

These are our conclusions. Utilization of the 72 products from the convenient sample during pregnancy was low, especially among those approved after 2008. Eighty percent of the products included in this preliminary analysis had 10 or more years of utilization data in Sentinel to characterize their use during pregnancy. Hypnotics and antidepressant products show the highest utilization during pregnancy. The oldest sedative hypnotic included in this query was approved in 1992 and the oldest antidepressant was approved in 1961.

So what are the next steps? The product exposure characterization during pregnancy among live birth deliveries will be updated to include products that were excluded from this convenient sample. These

updates will inform or will be used to inform the framework development and demonstration projects implementation. Year of approval, disease, and product related factor will be further explored to observe patterns of product exposure during pregnancy and product exposure by trimester of pregnancy will be described as this will provide us with more precise point of exposure during pregnancy. And before I conclude the presentation, I would like to acknowledge the FDA Query Team for providing key insights and direction, the Sentinel Operations Center for collating study design and reporting, and the Sentinel System Data Partner for contributing data for this activity. Thanks for your attention.

Wei Hua:

So thank you José. So next I will provide an overview of considerations to optimize the use of postapproval non-interventional pregnancy safety studies. So as a reminder, in PDUFA VII FDA is committed to develop a consistent and transparent approach to help decide when and what postapproval pregnancy safety studies might optimal be used to obtain timely evidence of safety for regulatory decision making. Safety concerns may emerge from various sources, either theoretically based on biological plausibility or from animal data, clinical trials, routine pharmacovigilance, medical literature, similar drugs on the market, or other sources. Non-interventional studies are commonly used to address safety concerns. To select a study, we typically ask two questions. Is the study able to detect or evaluate a signal and how early the signal can be detected or evaluated? Four factors can be used to inform the selection of a study, including outcome of interest, study goal, study's technical capability, and the magnitude of drug exposure. In the next few slides, I will briefly discuss each of these factors.

So I'll start with outcome of interest and the study goal. Pregnancy safety involves a broad list of maternal, fetal, and infant outcomes. These outcomes may be related to drug exposure in different critical periods during pregnancy, and they may occur under different pathophysiology mechanisms. Therefore, for particular drug, the questions are what outcomes are relevant to drug toxicity and how much do we know about the risk based on available information? The data gap about the risk determines the study goal. If we already have adequate information indicating a signal, the purpose of the study could be signal evaluation to confirm the risk or to quantify the risk. If there's only some suspicion of a risk, depending on how much we know, a study can be established to detect a signal or evaluate a signal. In contrast, if there is a critical knowledge gap for a particular safety issue or population, the purpose of the study is mainly for signal detection, which can be followed by signal evaluation.

This slide shows the difference between signal detection and evaluation. Signal detection is typically for hypothesis generation to identify a signal. The outcome can be pre-specified or non-pre-specified, meaning the study has no targeted outcome, but may use data mining methods, such as TreeScan to evaluate many outcomes simultaneously. In general, most uncertainty due to bias is accepted for signal detection, different from signal detection, signal evaluation is a hypothesis testing study.

The outcome of interest is typically a pre-specified and a signaled risk. In signal evaluation, higher level of certainty is needed so that the study should have strong internal validity. Talking about signal, so one question is, what constitutes a meaningful signal in non-interventional studies. It should be noted that there's no universal threshold for a signal. While the chain of between missing a true signal and identifying too many false signals should be considered. It is indeed a scientific and a clinical decision. Depending on a study scenario, a signal can be defined based on exposed versus unexposed, observed versus expected. It may or may not require statistical testing, or it may consider a less strict type of one error to avoid missing a signal. A signal may also be defined based on one analysis or a series of analysis.

So moving on to the study's technical capability. Like all epidemiological studies, technical capability is the most important consideration for a study. This is because suboptimal study design, data, and the methods may impact the study's internal validity, introducing bias and the creating uncertainty of study findings. It should be noted that each study design element, such as study population, exposure, outcome covariates, if not adequately handled, can introduce bias. In session four, we will use outcome misclassification as an example and assume that is the only bias in a hypothetical study. But in real world, a study may have multiple biases, and each bias may have a different impact on study findings.

For pregnancy and safety studies, there are a number of longstanding technical issues. For example, how to accurately estimate gestation age, how to accurately ascertain live births and non-live birth outcomes. None of those issues are new. In fact, there is a wealth of math development and validation studies in medical literature. However, concerns remain for the validated methods, given the suboptimal validation approach and the results. In addition, comparisons across study types, such as comparing registry and database studies are incomplete. More complexity is that data and the methodological challenges differ by study purpose, study question, type of medication, outcome of interest, data source, and other characteristics of a study. While adequate evidence is needed to support a regulatory decision making, we acknowledge there is no perfect study and the necessary level of evidence depends on study goal, and it should be determined on a case-by-case basis.

So for example, in this table, we consider the tolerance of uncertainty based on study goal and effect size. If we are concerned about missing a small effect in signal evaluation, we might have less tolerance for uncertainty around the bias. In contrast, if we are concerned about confirming a large effect in signal evaluation or detecting a signal, we may be able to tolerate more uncertainty around the bias, given that we may favor a less controlled study that may be completed more quickly. So this slide shows the types of non-interventional studies that can be considered for postapproval pregnancy studies.

As mentioned earlier, a pregnancy registry study is a prospective cohort study as in primary data collection. In the regulatory setting, pregnancy registry studies typically have comparator and a sample size requirements. A healthcare database study with a prespecified outcome refers to a study using electronic healthcare data such as EHR claims data with additional data collection on medical chart review as needed. A healthcare therapy study without prespecified outcomes refers to a study using electronic healthcare data, but there are no targeted outcomes.

So one example is TreeScan, which is a data money method that scans a hierarchical tree of outcome codes, such as ICD 10 codes, to evaluate many outcomes simultaneously. In this context, descriptive studies refer to population-based studies using primary data collection or electronic healthcare data. They are different from review of case series in routine pharmacovigilance. Rather, these studies have a defined study population, but may not have a comparator, or there is a comparator, but the exposure is so rare, which does not allow statistical testing. In general, all four types of studies can be considered for signal detection. For signal evaluation, which requires higher level of certainty, registry studies and database studies with pre-specified outcomes are more commonly considered.

So moving on to the last factor on this list... Sorry, I need to pull a slide that I can see. Magnitude of drug exposure is another important factor that informs selection of study. This study shows the conceptual categories of magnitude of drug exposure in pregnancy from very rare to very common. What category a particular drug belongs to may be informed by the patient, product, and the treatment-related factors. And we would prefer to discuss in session four. On this slide, the concept is that when the exposure is very rare in pregnancy, signal detection is feasible using descriptive... signal detection is feasible using descriptive studies, signal evaluation is unlikely. When the exposure is very common in pregnancy, an adequately large exposed population can be quickly accrued postapproval.

Both a signal detection and a signal valuation may be conducted in electronic healthcare data. In this scenario, registry study involvement may also be more efficient. If drug exposure falls in the middle category, the likelihood of using electronic healthcare data increases compared to the very rare scenario. However, depending on the outcome of interest, sample size may still be a critical issue, especially for the outcomes of low prevalence. So this concludes the overview of four factors that we consider for selection of study. The next slide links the FDA's analysis of postapproval pregnancy safety studies, labeling, and the drug utilization to these proposed factors. The analysis results in this slide have been presented by Adebola, Aida and José. In the interest of time, I will not go through the results, but we'll only share our observations.

So first, the FDA's analyses indicate that a wide variety of approaches have been used to assess the safety of medication during pregnancy and informing drug labeling. Among different types of studies, pregnancy registry studies are used at the most and have primarily contributed to signal detection and informing labeling with a large proportion of those registry studies from a disease-based pregnancy registry. Also in the PLLR labeling analysis, 10% of product labeling came from case reports in the case series highlighting the importance of different postmarketing approaches. Second, slow enrollment in the data accrual seems to be a common issue and it can lead to long lag to labeling and the study termination or release.

Third, in the current analysis, drug utilization in pregnant individuals is low, especially for newly approved drugs. The pattern cannot be solely explained by the number of years on the market. Other factors such as those related to the patient and the product characteristics, and the treatment considerations will be further explored. So overall, these observations emphasize the need for a consistent approach to help determine the optimal use of postapproval pregnancy safety studies. Also, we need to better understand the potential gaps in decision making of use of those studies through demonstration projects. So this concludes my presentation. With that I hand over to Megan.

Megan Clowse:

Thank you. And that does sound like a lot of meetings and a lot of data, so thank you for the presentation. Before we jump into questions, I just want to summarize what I've heard so far today as well as in the context of my work as a reproductive rheumatologist for 20 years, taking care of women who are trying to make decisions every day. And I think it's always important to come back to the patients and the decision process that they're living with. So if we think about vaccines, I really mean it sounds great, you get it once or twice, but our medications for patients who have chronic diseases, these are pills that women have to decide multiple times a day to put into her body when she's pregnant. And any kind of level of concern or uncertainty makes it less likely that she's going to do that.

Similarly with our biologics, women are injecting themselves with a medicine. It's a big leap to be able to take when our culture really tells you that medicines are wrong. So it's really important. It's imperative that we actually really get this data out, that it is accurate as possible, that we don't get confused by things that are irrelevant, that we don't throw the baby out with the bath water in terms of medications, that can be really helpful because of small risks. And then we really help doctors and patients especially really balance who they are as a human being, what their body is able to do if their body is able to sustain a pregnancy safely, or if they need assistance from medicines. So I really think that there's a couple of things that I've heard today that are really important. One is that we really need a balance of studies.

Things take a really long time, but then suddenly things come out. And so if we think about, in rheumatology, for example, suddenly some rheumatologic medications were brought up as treatments for COVID. So suddenly there are new research coming out, and if there hadn't been a baseline of 20

years of data on our medications for some of them it would be confusing to everybody. But instead, because people have been slowly accruing data over the years, we're able to balance the information that we have. I also... I know this is a postmarketing discussion, but I just think it has to be said that premarketing is also important and that for some of our drugs, that's the time when there's actually the most pregnancies, or at least a sizable number of pregnancies. Those women have been followed. We know a lot about their disease, we know all of their con meds, we know a lot about them.

And so really making sure that we follow those patients really carefully so that when drugs are approved, we at least have a modicum of actual human data, I think is really important. And finally, that what's in the label is one thing, and then how doctors and patients interpret it as another. And helping all of those people really have a framework in terms of balancing the important risks versus the other risks versus who they are is I think essential and that that's something that we all really need to work towards. And with that, we're open for questions, saying that we only have a few minutes, but I'm open to questions, comments. There's somebody over there. You need the mic. Go there and then you're next.

Andrew March:

Hi, this is Andrew March from the National Academies, Dr. Ajao in your presentation, you mentioned that of the studies in your sample that were completed, about 50% resulted in labeling changes and it took about 11 years to complete those studies on average. Are there tools within FDA's disposal to improve accountability for getting those studies done faster and getting that data out to patients and clinicians more readily? Or are there additional authorities that FDA needs?

Adebola Ajao:

I take that

Megan Clowse:

Your mic is on mic.

Adebola Ajao:

Oh, yeah. I'll probably defer that to Dr. Sahin. But we really try within the postmarketing setting to meet the milestone dates that we set for these pregnancy studies. And if industries miss those milestones, we really encourage them to meet the milestones. But as far as what we can do to make it faster, I think that's part of what this workshop is about, really trying to understand what can we do better to make the studies go faster so we can get this information out to the patients quicker.

Speaker 7:

I think I'm little concerned [inaudible 01:08:51].

Adebola Ajao:

The question. Thank You.

Megan Clowse:

There's it. Here and then, also in the back.

Keele Wurst:

Mine's really quick. In your PLLR analysis, did you look at labels that had multiple sources of data?

Adebola Ajao:

Yes.

Keele Wurst:

And What was that like?

Aida Kuzucan:

So, in my graph that I showed, the proportions are over a hundred percent. And the reason why is because oftentimes, for example, one of the examples that I showed at the end, it had 10 randomized control trials there. So, that's 10 studies in our analysis. So, there was oftentimes a mix. There were some labels, for example, that only had five publications on case reports, case series. It was dependent on when the review or the labeling review happened and what available data was there. We are trying to put out what available information that there is in the label so that people can make an informed decision.

Megan Clowse:

No. Back there.

Sarah Elafros:

Hi, Sarah Elafros from Syneos Health. And I have a follow-on question to Dr. Wurst's question about the PLLR, I was surprised, and maybe it's related to the question about multiple data sources, that there were 145 unique quantitative labeling statements. Is that right?

Aida Kuzucan:

Yes. As opposed to what, it was too small?

Sarah Elafros:

Well, no, what I was surprised by was the statement about there were 177 studies, and maybe this is getting at this idea.

Aida Kuzucan:

Yeah, there's actually over 300 statements. Sometimes we had one study on a class of products that was reported in multiple labelings? To clarify the numbers, there was oftentimes a label that had multiple different study types and multiple studies quoted within that labeling. So, I'm sorry if that was confusing. I tried to clarify that.

Sarah Elafros:

No, no. That is helpful. I was just surprised by the statement that there were, the 26% of those labels contained APR data and followed by the Scandinavian databases and then by Medicaid. It seemed to me that I was expecting more studies even in those 145 quantitative labels.

Aida Kuzucan:

Do you mean database studies you were expecting more?

Sarah Elafros:

No, I was expecting more data sources-

Aida Kuzucan:

Unique studies.

Sarah Elafros:

... For those 145. And I'm wondering if some of that has to do with the fact that you limited, necessarily limited it to quantitative statements. So if there was a statement that said no increase in birth defect risk was observed, would that in database X or in pregnancy registry Y, that would not have been considered to be a quantitative statement. It didn't have a number. Is that right?

Aida Kuzucan:

Yeah, that was a limitation of our analysis. We chose to focus on the numbers because of resource constraints and because it was just simply, it was easier and quicker for us to identify. We started from a sample of 1800, so we had to pick a convenience sample. That was the way we made sure we had a piece that we knew we could finish in time. But yes, in the statement that you just stated where there was evidence from the literature without a numeric value in there, we would've considered that a qualitative statement, and that would've been out of the scope of this analysis.

Sarah Elafros:

That makes good sense. Thank you for clarifying, and thank you for a nice set of integrated presentations.

Speaker 9:

And I think we have time for one last one, just one moment.

Lee Cohen:

So Lee Cohen from Mass General. So just back to the PLLR. So the interval of observation obviously is not small, it's almost 10 years. And I was just curious if you could tell folks what the mechanism is. For the one example that you chose, for example, with sertraline with a reference to 2009, there were multiple, multiple studies that followed looking at persistent pulmonary hypertension in the newborn, including Dr. Huybrechts' seminal paper. And so I just, what's the mechanism for revising the information in the label? Because Dr. Clowse was talking about informing patients, but frankly what's in the label is sort of very old and was followed. So I'm just sort of curious what's to guide the clinician if the information in the label is frankly dated?

Aida Kuzucan:

So I'm going to say briefly that I am just in the Division of Epidemiology, and I think probably Dr. Sahin is a more appropriate person to answer this question. I'm going to ask her to respond.

Leyla Sahin:

So one of the questions I think that you're asking is there were a lot of studies that were published on sertraline and why were they not mentioned in this presentation? And I think that the examples that Aida presented were just some samples, right? I think it wasn't all comprehensive and all-inclusive. So I think that's one important point that may not have come across in the presentation. And then the other question is why are some of the product labelings out of date? And yes, that's a good question. And that I guess that responsibility falls on our industry partners. And it's the responsibility of every company to make sure that their labeling is up-to-date and reflects the most current and up-to-date information. And that whatever is included in labeling is not out of date, out of context, a misrepresentation, and so forth. So that is an important observation that, yes, we do want to ensure that our labelings are up-to-date. Thanks.

Megan Clowse:

All right, I think we're finished. I think we have a 10-minute break. Is that my understanding? And then we're going to come in and finally hear the framework. Okay. Good job.

Session 4

Geeta Swamy:

All right, thanks everyone. I also want to just thank everyone from our virtual space and for our virtual participants for providing questions through the Slido system. I know that we've not been able to get to all of your questions, but just as a reminder that those are all being captured and will be shared with our FDA colleagues for providing answers as well. Just to know we are trying to embed your questions into the discussion that's ongoing in the room as well. Thank you all for doing that. So please keep them coming as well.

I'm happy to be moderating the Session 4 today in which our colleagues from the FDA are going to be unveiling the highly anticipated, I know you're... We need like a drum roll thing, to talk about the Pregnancy Safety Study Framework. So in this session that's titled The Design of Pregnancy Safety Study, whoops, let me hit that slide, there we go, design of the Pregnancy Safety Study Framework. We're going to hear again from Wei Hua, who will be presenting the proposed framework, and learn also about some of the important study characteristics and factors that are essential to think about and be including in the framework.

Following from Wei's presentation, we're going to hear from additional individuals from the FDA who have contributed to the development of the framework that she'll be presenting. So you'll hear from Clara Kim, who's a supervisory mathematical statistician in the Division of Biometrics 7, Office of Biostatics in CDER. Leyla Sahin, again, who you've heard from several times today, but who is an OB/GYN, and Deputy Director for Safety in the Division of Pediatrics and Maternal Health at the Office of New Drugs in CDER. And Sara Eggers is Director for the Decision Support and Analysis [Staff] within the FDA center... Excuse me, for CDER, as well. So I'm going to turn the presentation over to Wei.

Wei Hua:

All right. Thank you very much and thank you for the discussion in session 3. In session 3 we discussed the individual factors that may be considered to inform the selection of studies. So in this session, session 4, we will share FDA's current thinking on how to integrate different factors in a Pregnancy Safety Study Framework. Approximately 5.5 million pregnancies occur each year in the US. Half of

pregnant individuals use at least one drug or biological product to treat medical conditions. So, typically, at the time of approval there are limited or no human data on the safety of product used during pregnancy. As a result for most products, human pregnancy safety data are collected postapproval.

The purpose of the framework is to develop a consistent and transparent approach to help decide when and what postapproval pregnancy safety studies might optimally be used to obtain timely evidence of safety for regulatory decision making. In the PDUFA VII commitment, the framework focuses on non-interventional or observational studies and this is in parallel with other safety surveillance approaches, for example, routine pharmacovigilance. Combined, all sources of safety data may inform product labeling, benefit-risk assessment and the clinical practice.

As a reminder, different types of non-interventional studies can be considered for postapproval pregnancy safety, including pregnancy registry studies, healthcare database studies with or without pre-specified outcomes, and the descriptive studies. The different types of studies and the studies of a same type may defer by data collection, data source, in study design and analysis methods. So what a given study can do and if one study is better than another should be determined on a case-by-case basis.

This is another reminder regarding the difference between signal detection and evaluation. Signal detection is for hypothesis generation to identify a risk, so more uncertainty may be accepted. Signal evaluation is for hypothesis testing to confirm a risk or to quantify risk. With that, a higher level of certainty is needed, so the study should have stronger internal validity. As discussed in session 3, four factors may be used to inform selection of a study. We will use these factors to construct the framework. The framework is outcome-agnostic. For a particular drug, the outcome of interest is determined by available information and the regulatory question.

This slide shows the preliminary framework to help determine the viable and optimal non-interventional pregnancy studies to meet regulatory needs for postapproval safety data. Study from available information, there are two pathways. In the upper pathway the regulatory question is defined based on what we know or what we don't know about a potential risk. For a particular outcome of interest, depending on the data gap, the study goal can be signal detection or evaluation in a desired timeframe.

To achieve the study goal, a study's technical capability is a critical consideration concerning the direction and extent of potential bias. Potential bias impacts both the study's internal validity and the sample size required to detect a statistically significant association. In the lower pathway, based on what we know about the patient population, the product, and the utilization of similar products on the market, we might be able to estimate the expected utilization of this product by pregnant individuals in US over time. Because this estimated magnitude of exposure reflects the overall pregnancy exposure in the US, we need another estimate. That is, of the total pregnancy exposures, how many can be captured by a given study?

Three arrows point to viable studies in this framework. A viable study is a study that can meet the study goal with necessary level of evidence within a desired timeframe. To determine the viability of a study, internal validity, minimum simple size, and the expected exposure should be considered. Subsequently, among of viable study options, an optimal study is selected, considering timeliness, resource requirement, patient burden, and other factors. The optimal study can be one study or a combination of studies combining different strengths.

To take a close look at the framework, we divided it into five steps. Step 1 is to determine the outcome of interest in the study goal based on available information. All right. As discussed in session 3, pregnancy safety involves a broad list of maternal, fetal, and infant outcomes with varying relevance to toxicity of a particular drug. Based on what we know or what we don't know, outcome of interest can be pre-specified or non-prespecified. At the time of approval, available information is typically from animal data and the clinical trials.

If a product has been on the market for some time, available information can also come from routine pharmacovigilance and the medical literature. If there are similar drugs on the market, risks observed for those drugs may also inform the potential risk for this drug. The data gap, for example, if it's a known risk or if there is only some suspicion of the risk or if we are concerned about an unknown or unexpected risk, determines whether the study is for signal detection or evaluation.

Step 2 is to assess a study's technical capabilities. In epidemiologic studies, internal validity is always a priority when evaluating the suitability of a study. Suboptimal study design data and methods can introduce bias. Since there's no perfect study, it is important to understand what's the potential bias a study may have, what is the impact of bias, and if the bias may change the interpretation of results. For a given study, the necessary level of evidence depends on what a study aims to achieve, for example, to identify a risk or to quantify risk. So how much uncertainty can be accepted should be determined on a case-by-case basis. In addition to internal validity, the direction and the extent of bias also affects the minimum sample size.

This slide shows the key parameters that we need to specify to estimate minimum sample size. The assumed true risk is specific to the product-outcome pair. The prevalence is specific to the outcome of interest. Power is the probability to detect an effect when there is a true effect. Exposed versus unexposed ratio can be 1:1, 1:2, 2:1 or other ratios, depending on how many exposed in the unexposed pregnancies we have in the defined study population.

Type 1 error is a probability of false positive when there's no true effect. So, by allowing a higher type 1 error, it is more likely to have false signals but less likely to miss a true signal. In the simulation, we also adjusted bias parameters to illustrate how the risk estimate may deviate from the truth and how the minimum simple size may vary in different scenarios of bias.

In this example, we assume a true ratio risk of 2, a 3% prevalence for the outcome of interest, 80% of power and a 1:1 ratio of exposed to unexposed. For illustration, we use outcome misclassification as an example of bias, which is reflected by the imperfect sensitivity and the specificity of outcome ascertainment. Each row represents a unique scenario in terms of the direction and extent of bias. When there is no bias, the projected ratio risk is the true risk in simulation. In the presence of bias, the risk estimate deviates from the truth in different directions to varying extents, depending on the interplay of bias parameters.

The minimum sample size also change accordingly. To highlight a few scenarios, in the top three rows, the risk estimate is biased to less than one, meaning the risk effect can no longer be observed. In these scenarios, the signal will be missed even if there is a large sample size. The bottom row illustrates another scenario where a twofold true risk is overestimated to 3.5. This inaccuracy still allows a signal to be detected but is concerning for signal evaluation if the bias is not accounted in the study design analysis.

In general, more bias is tolerable for signal detection and a more rigorous study is required for signal evaluation. However, as illustrated in this example, in the presence of bias, minimum sample size to detect a signal may be larger or smaller than the true sample size. More importantly, even if there is a large enough sample size, the signal may still be missed in certain scenarios. For example, where the direction of association is biased to less than one. Therefore, study internal validity is still important, even for signal detection.

Step 3 is to assess the magnitude of drug exposure and the study's capture of exposure. Magnitude of drug exposure is another important factor. Ideally, we would like to have a precise estimate of pregnancy exposures. In practice, this may not be possible, so instead of precise numbers, we use hypothetical categories to illustrate the idea. In this example, we have five categories, from category 1,

very rare, with less than 10 pregnancy exposures per year, to category 5, very common, with more than 10,000 each year.

The threshold for each category is hypothetical. When the magnitude of drug exposure is very rare, descriptive study may be the only option in this framework for signal detection and a signal evaluation is likely not possible. When the exposure is very common, there may be many possible study options for signal detection and evaluation. If the magnitude of drug exposure is somewhere in between, registry and the database studies are likely possible options for signal detection and/or evaluation, depending on other factors such as outcome of interest, study's technical capabilities, and a given study's capture of exposure.

As briefly discussed in session 3, estimating magnitude of drug exposure integrates multiple patient and treatment factors. The light blue circle represents the targeted patient population. The size of this population depends on the prevalence of disease or condition. The characteristics of patient population, in particular, age, sex, and whether the condition is common in pregnancy, inform how many of these patients may become pregnant. In the meantime, the product and the treatment factors inform how many of these patients may use this drug.

Considerations include potential market share of the product and the utilization of similar drugs on the market. If the drug is indicated for a pregnancy-related condition, the toxicity risk, for example, if there's pregnancy-related warning or contraindication labeling, or risk evaluation and the mitigation strategy, or REMS. We also want to highlight the individual treatment decision making, which considers benefits to patient during pregnancy, what is known about drug toxicity and the available options. Potential for inadvertent exposure in pregnancy also contributes to the total number of pregnancy exposures. Another factor should be considered is how long the product has been on the market.

This is an example of a hypothetical scenario to estimate the magnitude of drug exposure. In this scenario, we have a newly-approved drug for a serious condition, and this product is the 5th in class. Under the patient factors, there are 800,000 patients in the US. The condition predominantly affects females with age of onset commonly in the 20s and the 30s, and that the condition can flare during pregnancy. Under the treatment factors, there is no pregnancy-related warning or contraindication labeling or REMS. Although a relatively small market share is expected, given it's 5th in class, this drug is likely to be prescribed among pregnant individuals. Considering the need for treatment, drug benefit, and the lack of evidence of teratogenicity.

Integrating all these considerations, it may be enough to assume the exposure will be common in category 4 with approximately 1,000 to 10,000 pregnancy exposures per year, in the US. The magnitude of exposure estimated by patient and the treatment factors reflects a total pregnancy exposures in the US. A study's capture of exposure can be thought of as a fraction of the magnitude of drug exposure. This fraction primarily depends on how patients are enrolled or accrued in a potential study.

For example, for a pregnancy registry study, it relates to the scope of registry, the recruitment and retention strategies, patients' willingness to participate, and other factors. For a study using EHR or claims data, it concerns the size of the data source and its relevance to pregnancy. Time factors also inference the fraction that a study can capture, especially in the first few years of approval. For example, it may take time to establish a registry to start patient recruitment. For database studies, potential data lag should always be considered.

This is a simple illustration of calculation integrating the magnitude of drug exposure and the study's capture of exposure. See, the overall magnitude of drug exposure is estimated in category 4,that is between 1,000 and 10,000 per year. For a nationwide pregnancy registry study with successful recruitment and retention, the study capture is estimated at 10%, that is between 100 and 1,000 per year. Similarly, for a large-scale EHR and claims data source covering 40% of the US population, it is

reasonable to estimate the study capture at 40%. So, using this data source, the study may be able to capture 400 to 4,000 pregnancy exposures per year.

Step 4 is to determine study viability concerning if a given study can meet the study goal with the necessary level of evidence within a desire timeframe. The evaluation will be informed by a given study's internal variability, minimum sample size and the expected exposure. This is a hypothetical lookup table of projected risk estimates and the possible sample sizes based on various parameters. We will use the numbers from this table in the following example. This is a hypothetical example to illustrate the idea of integrating internal validity, sample size, and expected exposure to determine study viability.

In this example, the study goal is to detect a signal for Outcome X. Outcome X is a pre-specified outcome based on animal data. There are five studies in consideration. The first two studies are pregnancy registry studies. The difference is that Study 1 uses a perfectly designed case report form. While the case report form in Study 2 is designed with flaws. Studies 3 to 5 use the same healthcare database but different methods. In Study 3, outcome events are adjudicated by chart review. In Study 4, potential cases are identified by claims-based algorithm without a chart confirmation. Study 5 uses a method for non-prespecified outcomes based on a hierarchical tree of IC10 codes. The viability of study considers the likelihood of detecting a signal for Outcome X within 10 years of approval. For simplicity, we assume non-differential misclassification is the only bias in these studies. Given different methods, the sensitivity and the specificity of outcome ascertainment vary by study, introducing bias in different directions and to different experiences. As a result, the projected risk estimate varies by study.

In this example, although the risk estimates deviate from the truth, they're all above one. So the risk effect may still be observed for signal detection. Due to different direction and extent of bias, the minimum sample size also varies by study. Whether the sample size requirement can be achieved depends on the study's capture of exposure. Study 1 is a nationwide registry study with a well-planned recruitment and retention strategies. It expects to capture 10% of the total pregnancy exposures in the US. In this example, the annual pregnancy exposure in the US is estimated to be between 1,000 and 10,000.

So Study 1 is expected to capture 100 to 1,000 per year. Compared to Study 1, Study 2 is a smaller registry study and they may only capture 1% of the total exposure. Studies 3, 4, 5 use the same data source that covers 40% of the US population so their estimated capture of exposure is 40%, that is between 400 and the 4,000 per year. The expected 10-year exposure accounts for slow market penetration in the first few years of approval, the time needed to establish a registry study, and the potential data lag.

The likelihood of detecting a signal for Outcome X within 10 years of approval considers the study's internal validity, minimum sample size, and expected exposure. In this example, the potential bias does not change the interpretation of results for signal detection. So it is acceptable for all these studies based on the study goal. However, sample size is more likely to be achieved in Studies 1, 3, and 5, even if the exposure is captured toward the lower end of the range. So these three studies are likely to be viable studies. For study 2, the 10-year expected exposure is between 70 and 700, which is way below the required sample size, so Study 2 is not viable. For Study 4, sample size may or may not be achieved, depending on where in the range the exposed pregnancy can be captured. So Study 4 is possibly not viable, especially if a 5% type 1 error is used for signal detection.

This is a hypothetical example and it's only for simple illustration of concepts. In real study scenarios, we need to consider a more complex interplay of outcome of interest, various sources of bias, necessary level of evidence, and the expected exposure. It should be noted that the direction and the magnitude of bias affects the accuracy of risk estimate and the minimum sample size. In the presence of bias, larger than expected sample size may be required for a viable study. However, even with an adequate large

sample, risk estimate remains inaccurate and the signal may still be missed due to the direction of bias. Therefore, enhanced technical capabilities should be considered, depending on the impact of bias on study results and inference.

In addition, if the expected exposure is very rare, descriptive studies may be the only viable option in this framework for signal detection, and the signal evaluation is likely not feasible. If the expected exposure is not very rare, multiple viable options might be considered. The threshold for very rare may differ by scenario. It would be difficult to precisely estimate the expected exposure. Therefore, the refined categories are desired. One important concept not illustrated in this example is that the viable studies may be used along or in combination, either sequentially or simultaneously, to improve efficiency.

The final step of this framework is to select the optimal study from the viable options. For signal detection, one important question is, how early a signal can be detected? In this example, Studies 1, 3 and 5 may be able to detect the signal as early as Year 4, depending on where in the range the exposed pregnancies can be captured. Selection of optimal study considers timeliness, resource requirement, and other trade-offs. In this example, all three studies can detect a signal in a timely manner. Resource requirement is considered high for Study 1 as it needs to establish and maintain a large-scale registry. Both Studies 3 and a 5 use an existing data source. Study 3 conducts chart reviews so it may require more resources than Study 5. Although the perceived resource requirement for Study 5 is relatively low, one trade-off is that the method may not be efficient for the pre-specified Outcome X. Using the data mining method, it is likely to generate many false signals which will require additional effort to clean up. So, putting all this together, Study 3 is selected as the optimal study in this hypothetical example.

Next, I will walk through an example to summarize the thinking. For Drug A, animal data suggested some risk for Outcome X. There is no safety concern for any other outcomes from any other sources. With that, the regulatory question at the time of approval is if there is an increased risk of Outcome X associated with Drug A during pregnancy. Given the limited information, the study goal is signal detection, if detected, followed by a signal evaluation within 10 years of approval. A study using Database Z is being considered. However, this study may be subject to non-differentiating outcome misclassification with 80% sensitivity and a 90% specificity of outcome ascertainment according to a prior validation study.

Based on our assessment that the ratio and the magnitude of bias will not change the interpretation of results. So the potential bias is acceptable for signal detection. For signal evaluation, to quantify the risk, the study's technical capabilities need to be improved to account for bias. Based on biological plausibility and the clinical judgment, we assume a relative risk of 3 for the drug outcome pair, with an estimated 3% prevalence for Outcome X, and 80% desired power. 2,308 exposed pregnancies are required for signal detection, with a 20% type 1 error.

In the meantime, Drug A is indicated for a common condition that affects women of childbearing age. It's the 5th in class, and there's no pregnancy-related warning or contraindication labeling or REMS. Based on these factors and the utilization data of similar drugs on the market, the pregnancy exposure is estimated between 1,000 and 10,000 per year in the US population. Database Z covers 40% of the US population so it anticipates capturing 40% of total pregnancy exposures in the US. Integrating the study's internal validity, minimum sample size, and expected exposure, signal detection is anticipated in Year 4 with improved technical capabilities for more accuracy.

Signal evaluation to quantify the risk may be achieved in Year 5. With that, the study is considered a viable option. Comparing with other viable study options, this study can detect and evaluate the signal for Outcome X with the required level of evidence in a timely manner. The study can be efficiently operationalized using the existing resources. So this is selected as an optimal study.

This summarizes the FDA's thinking. Further development is needed to identify potential gaps in the preliminary thinking of the framework to determine how to estimate the magnitude of drug exposure, especially at the time of approval, and at the fraction that a potential study can capture. Also, we need to evaluate how these studies may perform similarly or differently in different scenarios. We hope to learn more from the demonstration projects, which will be discussed in session 5 tomorrow.

Based on the additional knowledge, the framework will be refined by the end of PDUFA 7. Finally, I'd like to acknowledge our colleagues from various offices in FDA CDER and CBER who made invaluable contributions to the development of this framework. All the information presented here is a product of team effort. Thank you.

Geeta Swamy:

All right. Wei, I think you get the opportunity to sit down because that's been a long time through all of that, and I think you probably saw people out in the audience going, "How do I interpret all that's in your brain that you were just conveying to us?" So I'm going to, I think, kick off some of the discussion, and I'm going to start with a few questions directed to our panelists and give Wei a moment to catch her breath as well. Leyla, I might ask you since you've been put on the spot a few times already today. That's a lot, and I think it's a lot that everyone's going to need to think through and absorb, but how do you think this framework is really going to then fit into our overall safety assessment for products as we go into this postapproval setting?

Leyla Sahin:

All right. Thank you for that question, Geeta. We've all thought about these issues. Like Wei said, it's been a real team effort across different offices and divisions, and so this is our first attempt to put this out there to you, our stakeholders, to get input from all of you. So I guess, in terms of where it all fits, the framework, I think, is one piece of the overall puzzle. The overall puzzle is drugs are approved, they get down to the market, and then we have to figure out how to get safety. And so this is one piece of the puzzle in terms of trying to figure out what are the ideal study methods or designs.

I think there's a few other considerations that we also have to keep in mind in the overall safety assessment of newly-approved products. I think one thing that I wanted to mention that I think hasn't been mentioned yet at this meeting is that industry has access to safety databases where anybody, any patient, consumer, anybody can report a pregnancy that has been exposed to their product.

And so we review these annual reports, and what's interesting to see is that we often see significantly more cases reported to sponsor safety databases, disproportionate in terms of then you look at the numbers of pregnant individuals who are being enrolled into a pregnancy registry or that you look at a database, a claims database, or we look at Sentinel and we just aren't seeing the exposures there. So I think we need to think about how to leverage that data. Although these adverse event reporting systems are set up for the public to report adverse events, pregnancy is a different kind of situation because folks are willing to report pregnancies that are occurring and so this is an ideal opportunity to collect data.

So we need to figure out a way to then get good follow-up data on those pregnancies that are being reported because they're not necessarily just adverse events, they're actual exposures of pregnancies that are occurring, and often they are reported during the pregnancy. And so this is a great opportunity to do really good prospective follow up and get those outcome data in those patients and try to leverage, see what we can do with that data.

Geeta Swamy:

Sara, do you want to add to that?

Sara Eggers:

Yeah, if I could add one thing, because I think it's been mentioned today, and we've talked a lot about this. We recognize that the individual patient benefit-risk has to take into account many factors, what the patient needs, what the benefit of the drug would be, what other treatment options are there and, importantly, what is the risk known about this drug? And so we just want to be, I think, very clear, as we've talked, that this is to fulfill that safety piece of the puzzle. I just want to put a nod out to the other comments that have been made about the importance of the overall patient benefit-risk.

Geeta Swamy:

So, picking up on that, then, do you see this framework, then, helping to provide a way to address some of that gap at the patient and provider level on what kind of information might be helpful or be known?

Sara Eggers:

I'll let others comment. With the same goals as we've talked about today, as all of them, all of the safety studies are intending to do, but I'll let my colleagues add anything to that.

Wei Hua:

Yeah, just to add to that, I think, as Sara said, thinking about the benefit-risk, we need data for risk, talking about the risk, and this framework really helps us to identify the optimum use of the studies to generate data for risk so we can consider in the ecosystem how to use those information.

Geeta Swamy:

There's a lot of, I think, assumptions that are going into this framework that are the ways to develop these strategies, the study designs, and so forth. So one of the things that is really important is about the numbers, meaning, well, there's a lot of numbers, but the numbers of expected. Expected cases, expected exposures, actual rates. How do you think that those things are going to come into play when a sponsor or a target group sees a condition as being really important, but maybe it's a rare condition, ones where they're very common, but people don't really see this as a significant concern. So there's going to be these sort of things that come into play when trying to decide what's the optimal design, even though that might not be what the going-in thought might be. So how do you think that's going to address?

Wei Hua:

I don't know if anyone want to comment first? I can go first. I think that's why this is so complicated and that's why we needed a framework to help us think this in a more structured way using a consistent approach. There are so many different numbers, different factors we need to consider at the same time. So we don't want to really focus on one aspect and forget the other one. We want to put everything together on a list so we can think on everything at the same time to evaluate their importance in different scenarios. I think it's really about how to put different thoughts, different important factors together.

Geeta Swamy:

I know there's going to be lots of discussion on this tomorrow, as well, but it strikes me as the starting-off point is there's approval of a drug or biological product, and then it isn't just go do a post-market evaluation. It's going to be discussion about what's maybe the best approach to doing the post-market evaluation. What is the right strategy to take? How can that be something that is discussed together? Is that conceptually what you all are thinking?

Wei Hua:

Yeah, yeah. Right. Yeah, I think, as Leyla mentioned in the morning, there are different approaches, not only the non-interventional studies, we also have pharmacovigilance. So the data come from all different approaches, but for non-interventional studies, because there are so many complicated factors, we just need to have this framework to help us to sort out the issues.

Leyla Sahin:

Are there questions for Clara? Other questions for Clara.

Geeta Swamy:

Okay, sorry. Yes, I'm sitting back here.

Speaker 5:

Just one fundamental question that I think might be easy to answer, and that is, in your example, you landed on one study type for a drug. Is that the goal that you're going to land on one study type for a drug or that you're going to end up with a group of studies because of all the different biases in different directions and so on?

Wei Hua:

Anyone want to comment first?

Leyla Sahin:

You can answer that one.

Wei Hua:

Okay, I can go first. It depends. I think the framework has enough flexibility for us to deal with different scenarios. And we do think about what is the best way when we consider all different factors at the same time. If it's more efficient to have multiple studies simultaneous or sequentially, we can go that way. Or if one study is adequate, then we don't have to use additional studies.

Geeta Swamy:

When do you envision this coming into the planning or the discussion? Meaning the decision about the right approach or the discussion with the sponsors? Has anyone thought through that yet?

Leyla Sahin:

Well, post-marketing requirements are issued at the time of new drug approval so ideally these types of discussions should be happening prior to those decisions, prior to drug approval. Yeah.

Geeta Swamy:

Okay. That was said ideally, right? I think, a question over there.

Evan Myers:

Hi, Evan Myers from Duke. A lot of these considerations are going to be the same for a given condition or class of drugs. Do you envision, every time a new drug is approved for a given indication, that the individual sponsor would have to go through this again? Or could there be kind of a library of considerations for a given class of drugs, for rheumatic conditions or hypertension, that could then be shared? Again, to improve consistency and comparability and all the things that we ultimately want.

Wei Hua:

Anyone want to answer?

Sara Eggers:

I can, and then I'll turn it to you, Wei.

Wei Hua:

Yeah, yeah. Go ahead.

Sara Eggers:

There are parts of the framework that are common. I would think they can be replicated, reproduced. But for a given study design's technical capabilities, that, I think where the biggest difference, the biggest customization's happening or the biggest difference that would have to be really thought through. I don't know, Wei.

Wei Hua:

Yeah, I agree. I completely agree. I think the framework, as we designed, it's not for a particular therapeutic area. It has the flexibility. So if additional factors should be really considered, we can definitely add that to the framework. So I think more thorough thoughts, maybe after the workshop, after we hear back from the stakeholders, we can think better how to refine the framework.

Jeff Roberts:

Jeff Roberts from Merck. First, I just want to applaud this work that was done, a very comprehensive set of data on all the studies we have in all these different contexts. And I guess I want to take it to the next step and say, have you started to think about whether we reached an optimal endpoint for each of these products in terms of the labeling that we got to for these studies that were done?

And I guess maybe to take that even a step further, have you thought about taking this new approach to this framework and applying it retroactively to some of the studies that have been done and saying, "Would this have worked better or different if we had this framework back when we were thinking about it then?"

Wei Hua:

Leyla, do you want to answer that?

Leyla Sahin:

All right. You're asking us to do a lot of work, Jeff. I think we have enough on our plate with these commitments with all these demonstration projects that we're going to have to complete by the end of fiscal year 2027 so I don't think we've thought about doing any kind of retroactive type assessment.

Geeta Swamy:

Well, I might pick up on Jeff's question though and say, let's say there's something that the label has been changed based on some data that's out there. It might be a question to say, was that the approach that you would've chosen if you'd gone through this framework? Just sort of as a hypothetical to say, take-

Jeff Roberts:

Are we satisfied with where we want to? Is it good enough?

Geeta Swamy:

Yeah, because I do think that I'm looking at this from the standpoint of a research investigator, and it's almost like when we go back to clinical trial design and go through the phases and understand. This is going to take people a while to figure out how to get the information and the inputs and understand what they even mean because you're talking about a lot of... The data's there, but how we put it into the terminology that you're using is going to take some learning.

So I do think sometimes using examples, real examples, where you can... I think your examples are helpful, but I think it will be even more impactful if you take it in a framework of this drug, this is where it went, here was the pathway, here's what it would've been if we were doing it through this model now. And that might be a way for people to grab that and then understand.

Wei Hua:

Okay, it's a good idea. Yeah.

Geeta Swamy:

Sorry, I turned this way so I know I'm biased to the side of the room.

Elsie Grace:

Hello?

Geeta Swamy:

Oh, sorry. I didn't mean to turn away.

Elsie Grace:

No worries. Hi, I'm Elsie Grace from Eli Lilly. I was just thinking about the framework and, thank you, it was a great presentation. I'm wondering how the Sentinel could be incorporated into the framework or the idea of the sufficiency in the PMRs. I don't know if there was any thought about that, particularly if some of this discussion is ideally supposed to be happening prior to approval?

Wei Hua:

I can address that at a very high level. At this time, we are only considering the science, so in the ecosystem, how the stakeholders can work together to generate the best data for decision making. So what stakeholder will do what, that will be the next step.

Elsie Grace:

Thank you.

Geeta Swamy:

I have a few questions out here. I think we have-

Speaker 9:

Hi. Just a quick question on how do you envision the framework translating into labeling? Leyla, you had mentioned spontaneous reporting. Is that sufficient to put in a label?

Leyla Sahin:

Well, everything of course comes down to the quality and the quantity of the data so... Oh, I thought there was somebody in the back who was making a comment.

Speaker 10:

No, no, I was just [inaudible 00:50:42].

Leyla Sahin:

Oh, okay. Sorry.

Geeta Swamy:

Okay, we'll cut in one second.

Leyla Sahin:

I think it's an assessment of what data you have, so what's available, and looking at the quality and the quantity of the data. We all understand the known limitations of pharmacovigilance data and sometimes that data can be messy, but sometimes that data can also be really informative. And I think we've had several examples mentioned today of how pharmacovigilance data has been useful. I can also think of another example of a cancer drug, trastuzumab, which is approved to treat breast cancer and clinicians who were treating pregnant individuals with breast cancer were observing that the amniotic fluid was going down, resulting in a condition called oligohydramnios, which was resulting in fetal death.

These were interesting observations. And then, to add on to that, treatment was withdrawn and that resulted in the amniotic fluid re-accumulating, so kind of like a challenge/de-challenge kind of situation. And then they restarted the drug and then the amniotic fluid went down again. These are based on pharmacovigilance, case reports that identified a signal, but the signal was so strong and it was actually with this challenge and de-challenge situation, it was actually confirming the signal as well. So I think you have to take every data source on its own and evaluate it for its own value.

Geeta Swamy:

Okay. Out there and one over that.

Speaker 10:

Okay, great. Thanks. So my question, it seems like from what I've seen just here shortly on the framework, a lot of it is dependent on the sample size estimates in the magnitude of the expected drug exposure, which we've heard earlier today that sometimes it's hard to predict and sometimes things don't pan out where we think they should. And even with some of the larger database studies, sometimes they still have small numbers. So how will that, I guess, allow for some flexibility in real-world situations if you go down the path of a particular study design and see that it's not really giving results or numbers of exposures that we expected it to in order to still come to a successful outcome?

Clara Kim:

Yeah, I'll take that one. I think that's very true that for us to get a pregnancy study that's successful, the actual numbers have to come out similarly to what we used in the framework or what we projected. I think someone mentioned it this morning about doing feasibility assessments. I think we can't predict the future. It's really important for the protocol to have plans to monitor exposure and capture rates and also have plans for feasibility assessment periodically.

I think Dr. Hua, in the slides, mentioned better prediction of exposure at the time of approval. I think that's something we should work on and hopefully, with this whole process, that we'll be able to better predict exposure. The framework is now incorporating this classification bias in the sample size requirements. So hopefully, with that, we'll be able to minimize the discrepancies and hopefully feasibility assessments, after initiation of the study, will help us get more successful pregnancy studies.

Speaker 12:

Okay, and this will be the last question over here.

Geeta Swamy:

Yep.

Dinci Pennap:

I'm Dinci Pennap from AstraZeneca. My question is for Wei. When you were presenting, you said something about 40% coverage, like 40% national coverage translates into 40% pregnancy coverage. Isn't that a stretch to assume that, if a data source covers 40% of the US population, you would expect to see 40% of the pregnancies, especially considering mother-infant linkage and considering case reviews? Because one of your slides assumes that case reviews, chart reviews, would give the same number of accounts as a database study, which I find that hard to follow. So maybe I'm missing your point.

Wei Hua:

There's no point there. It's a hypothetical number.

Dinci Pennap:

Hypothetical.

Wei Hua:

Yeah, it's hypothetical.

Geeta Swamy:

It's exemplary.

Wei Hua:

Yeah, yeah. Sorry for the confusion.

Geeta Swamy:

Okay. So I think we'll have to close the discussion tonight, but I just want to thank the panel for their contributions and I know we're going to have lots more discussion tomorrow, so you all better be ready hopefully for the other panelists and so forth. I just want to remind folks that you can continue to send questions in and also that there's discussion via the docket that is open and tied to the federal register notice. So we'll turn over to Marianne Hamilton Lopez to close us off, please. Okay.

Marianne Hamilton Lopez:

All right. Thank you again to all of our participants who contributed today. We heard a lot of exciting developments. Just as a brief recap, we started off the day with the FDA's opening presentation where the lack of pregnancy safety data in pregnant individuals was emphasized as a public health issue and that all stakeholders are needed to address this. In session 1, we heard from a variety of stakeholders, including those in industry, healthcare providers and a patient representative. We learned that oftentimes both providers and their pregnant patients must balance the risks and benefits of managing their disease and the safety of the pregnancy, which is a direct result of limited information.

With limited data, many patients feel torn between continuing medication while also ensuring the safety of their pregnancy in the face of limited data on medication. Creative solutions for collecting quality data are necessary to improve data availability and accuracy. In session 2, we heard from those who design and conduct pregnancy safety registries, the challenges and advantages of each type of registries were discussed. The Antiretroviral Pregnancy Registry was highlighted in its success, but it required a wide range of expertise and a firm commitment from various stakeholders.

Another key lesson in the success we learned from the CDC's COVID-19 registry is how critical it is to clearly define outcomes of interest, ideally narrow in scope, rather than open-ended, as well as plan specific analyses prior to implementation to guide the data collection. Then in session 3, we heard from our FDA staff about recent work in the postapproval pregnancy safety space. Their presentations concluded that pregnancy safety studies have not had a sufficient time to mature and inform the regulatory action, with only 46% of the studies informing labeling updates.

Next steps for studies examining product utilization data include further exploring years of approval and disease, and using these updates to inform the development of the framework. To optimize postapproval pregnancy safety studies we covered the importance of considering the outcomes of interest, study goals, studies' technical capabilities, and magnitude of drug exposure. Building on these learnings, session 4 covered how the Pregnancy Safety Study Framework operationalized these important considerations into a consistent and transparent approach in how we may be able to optimally obtain timely evidence of safety for regulatory decision-making.

As we conclude today, I'd like to remind and encourage, again, this audience to contribute to the discussion via the docket. It's currently open and the docket is seeking broader feedback from the community on this very framework that we're talking about. You can find the link to the docket via the federal registry notice for this public workshop, and it will be open till November 30. We're just going to

keep saying it. And with that, we will conclude day one of the workshop. We look forward to reconvening tomorrow at 10:00 AM virtually and in person here at the National Press Club.

I've been asked to remind you, if you came in person, please bring back your badge and we will be having tomorrow a live open public comment where individuals who have previously requested to provide an oral comment on postapproval pregnancy safety studies will have the opportunity to do so. If you're interested in participating in this open public comment period, please reach out to our staff at Duke Margolis. After the open public comment session, session 5, we'll discuss existing gaps in the design and performance of different pregnancy safety study types to better inform the development of the framework. We will learn about the expected demonstration projects that aim to inform future development of the framework that was presented today.

The last session tomorrow, session 6, will have targeted discussion with select stakeholders in the community to provide additional feedback on the FDA's Pregnancy Safety Framework. Again, thank you to all the speakers and panelists who shared their work and insights with us here today. We covered a lot of ground. And then, finally, thank you everyone who joined us virtually and in person. We look forward to seeing everyone tomorrow morning, and have a great afternoon.

Session 5

Gerrit Hamre:

Good morning, everyone. We are going to go ahead and start up. Day 2, second and final day of the Margolis FDA Workshop on Optimizing the Use of Postapproval Pregnancy Safety Studies. My name is Gerrit Hamre. I'm a research director here at the Duke-Margolis Center for Health Policy focusing on primarily the FDA Portfolio Projects that we have here. I very much enjoyed yesterday's opportunity to contribute and participate with all of you, the activity both here and virtually was really wonderful. As a reminder, we are convening this workshop under a cooperative agreement that we have with the US Food [and] Drug Administration. The conversation today, while supported through a cooperative agreement with FDA is not a federal advisory committee, so we will not be voting, making recommendations, et cetera. Also, as a reminder, this particular meeting is tied to a PDUFA VII FDA commitment to develop a framework to identify and meet the regulatory needs of designing optimal studies to obtain safety data in postapproval settings.

Now, we recognize that there are other important clinical needs for pregnant women, maternal, fetal and infant outcomes. We have discussed them yesterday, I imagine they will come up a bit today as well. But day 2 really does focus specifically on the draft framework that was discussed yesterday. At risk of an ad nauseum reminder, we invite you and sincerely encourage you to submit comments on the draft framework to the docket prior to its closing end of day November 30th, 2023. Also, we would like to remind you this is the first iteration of FDA's draft framework, hence the importance for your contributions as they can work on subsequent iterations that follow. To that end, thank you also for recognizing that FDA colleagues in some ways and sometimes can be limited in their ability to respond in detail, but certainly encourage as much active participation. All of your feedback given will help inform framework updates.

We began the workshop yesterday talking about a variety of those key challenges that informed the framework, but also the broader set of maternal, fetal and infant outcomes. Later in the afternoon, we learned about the efforts at the FDA to address these challenges that we've seen in the development of pregnancy safety study framework. Today, we'll build on those learnings. I'm looking forward to more future developments.

We'll begin the day today with an open public comment session where individuals who have previously requested to provide an oral comment on postapproval pregnant studies will have the opportunity to do so. We've got a couple folks here in the room and then we've got a couple folks that signed up virtually to participate. We'll jump into that in just a moment.

We'll then transition to session five of the workshop where we'll hear about the expected demonstration projects that are looking to fill existing gaps in the design and performance of different pregnancy safety studies. Also, used to inform the development of the framework and capabilities of Sentinel and BEST in responding to questions of pregnancy safety.

After a lunch break our last session will then highlight feedback on FDA's pregnancy safety framework. We've got a wonderful panel. We very much appreciate your comments and participation there. And then on our next ... Oh, I apologize, that workshop agenda should have been up there for that last piece. Give folks a moment there to look at it.

Statement of independence. We always like to remind folks that the Robert J Margolis Center for Health Policy is part of Duke University. As such, it honors the tradition of academic independence on the part of its faculty and scholars. Neither Duke nor the Margolis Center take partisan positions, but individual members are certainly free to speak their minds and express their opinions regarding important issues.

Also, just a reminder to everyone, once more, we will not be monitoring the Q&A function in Zoom since in-person attendees do not have it. We've been using the Q&A platform called Slido for live audience and questions. Please take a moment if you are here or virtually to scan the attached code, the attached QR code or navigate to slido.com. You can also go there and enter the code #PregSafe.

Please also, for folks virtually, continue to submit your questions. We received a bevy of excellent questions virtually yesterday and although you may not have heard people calling in virtually, that's because we were receiving them electronically sending them up to the moderators and the moderators were incorporating them when possible during the meeting. Thanks also for the same level of active participation that we had from you all yesterday. It was wonderful, both the dialogue during, in between sessions and after.

Lastly, I'm sure we'll still get a few messages about this today, happens every time. I'd like to remind you that meeting materials, including agenda, speaker bios are currently available on the Duke-Margolis event website. The recording of the workshop and copies of the presentations will be made available on our website. Generally happens just a few days after, certainly within a week's time.

All right, we are going to move on to the open public comment section. Bear with me if it's a little wonky for the first two virtual ones, we will make sure that they are heard loud and clear. I'm going to start with those virtual participants, so Sarah Obican, if you could unmute yourself, confirm that you can hear me-

Sarah Obican:

Yes.

Gerrit Hamre:

I can see you. I see you there. Good morning.

Sarah Obican:

Good morning.

Gerrit Hamre:

Would you so kindly introduce yourself? I'll confirm that we can hear you just fine and then we'd be very interested in hearing your public comment.

Sarah Obican:

Thank you. Good morning. Thank you to the organizers and the participants regarding such an excellent workshop. My name is Sarah Obican and I'm a maternal fetal medicine physician with training in reproductive toxicology and teratology. As any maternal fetal medicine practice, our team cares for pregnant patients with complex medical conditions as well, and I'm commenting this morning as the medical director of the Exposures Clinic of Mother to Baby Florida where we care for patients who have questions about medication exposure.

As we know from the literature, pregnant patients take on average in the US more than four medications during their pregnancy and therefore to ask for future research, some unfortunately difficult tasks, I would ask that our research projects help answer the issue of polypharmacy and how these medications interact with each other. Also in line with the principles of basic teratology, I would like to stress the importance of research focusing on the specificity of exposure in terms of timing during that pregnancy, the specificity of both dose and medication compound utilized when evaluating the very specific fetal and neonatal outcomes that we really want to know about.

Lastly, and I know this is not exactly in line with today's workshop, but I think really important, I would like to encourage all of us to help in the realm of education through our work in academia and government and in the pharmaceutical industry and any number of national and regional organizations such as Mother to Baby. I'd like us to focus our resources and our time in teaching the clinical providers and our patients alike about the risks and benefit of taking those medications and really the risks of chronic disease on that pregnancy. This is really how the work that you're doing today and how the future work that will come out of this meeting helps those patients we care for clinically every day. Thank you again for all that you do and the time that you have spent during this meeting.

Gerrit Hamre:

Sarah, thank you as well. Very much appreciate your participation and contribution. Next up we have our second and final virtual participant. Is it Al Romeo? Al, do you mind joining us, introducing yourself, and I'll confirm that we can hear you okay.

Al Romeo:

Thank you. Yes, thank you to the speakers and everyone behind the scenes that made this event happen. I'm Al Romeo, a teratogenic information specialist with Mother to Baby Utah and a member of the organization of Teratology Information Specialists. Now the 13 Mother to Baby affiliates across the us, when we talk to pregnant and breastfeeding individuals and their providers every day they have questions about different medications. Often they don't know where to find the information and even if they do find the information, they don't know how to interpret it. We rely upon the different pregnancy registries and the studies to provide future information on those many medications that have not been studied yet. We also rely upon those studies to provide more information about those medications where there's limited studies that don't provide clear and consistent information.

As we saw with COVID-19 vaccine in the example from Monday's, yesterday's presentation about rheumatoid arthritis, patients and providers often respond to limited or no information by stopping needed medications, even though avoiding those medications may be worse for the pregnancy.

Now, continued funding of the studies and expansion of those studies is needed to provide the research to answer those questions from patients and providers and avoid those adverse outcomes for pregnant

individuals and their children. So thank you for all that you do and appreciate all of that work. Thank you.

Gerrit Hamre:

Thank you so much as well. Glad that worked out actually pretty well with the virtual contributions. Two in-person volunteers to participate as well. First, I'm going to start with Katie Schubert. Katie, do you mind stepping up to this microphone and introducing yourself and sharing your public comment?

Katie Schubert:

Great, thank you. I am, oh, I'm not on. It's on now. Okay, there we go. Thank you. I'm Katie Schubert, I'm President and CEO of the Society for Women's Health Research. I'm really pleased to see everyone here today and I appreciated the opportunity to join virtually yesterday, so that was really great to be able to do that. I am here on behalf of SWHR, but also as a part of my own longstanding work to advocate for policy and regulatory action that will better include pregnant and lactating populations in clinical research broadly.

I remember attending the May 2014 meeting at FDA when I was representing the Society for Maternal Fetal Medicine and around that same time we founded the Coalition to Advance Maternal Therapeutics, which is still chaired by SMFM, but the administrative home is at SWHR. And so given all of that, we really appreciate all of the work that's gone into this topic. I do think it's a step in the right direction, just like that initial meeting back in 2014, and I hope that we can work together to move this potential framework forward more quickly so that the data and research needs are better integrated into the approval and post-market surveillance process.

There were a few things related to framing the issue generally that I wanted to just reiterate and note and really appreciate it being brought up yesterday. This shift in the frame of reference from not just focus on birth defects but also the maternal impacts, really critical. And also thinking about the end points and the outcomes and what are we really trying to achieve here. I think all of this is going to be really important to the intent of the framework, but also to support that shared decision making piece that was talked about yesterday as well.

To that point, one of the issues that was identified at the CAMT meeting initially was access to the data or information from those exposure registries. So back in 2014 you'd go online, you'd get a little note that said, please write to the drug sponsor and we will mail you a copy of the results. Just upon initial inspection yesterday, it's so much easier and better, so thank you. I think probably we have some of the internet to think for that, but also all of the work to think about how we can make sure that people have access to these things and we'll be much better off for that.

We do see these registries as one of a few critical ways to further this inclusion, and we've heard anecdotally from SWHR's working group on autoimmune that this is critically important. We did hear this yesterday as well, but I would argue that for a woman with something like lupus who has made a painful decision to never have children, we're probably far past time to be thinking about how we can make sure that if that is the path she so chooses, we can help to figure out how to make that happen for her. So on behalf of that working group, thank you again for doing that.

A couple of things on the draft framework, that consistent approach, the upfront more proactive approach versus reactive, that's really exciting. I think a really thoughtful and intentional approach to this is going to be critical. I also understand we're talking about pregnancy safety, but I would be remiss if I didn't say there's lots of other data points and elements we need to make sure we're looking at to incorporate back into, not just on the safety side but on efficacy and thinking about the maternal health crisis.

And when you have a framework that is final and we will be sure to do public comments by November 30th, we do welcome the opportunity to have FDA and others present to the coalition so that we can work with our corporate partners and our nonprofit partners to get the word out about this to better educate all of those members. So thank you again for allowing me to provide these public comments and we appreciate the opportunity to continue working with you.

Gerrit Hamre:

Thank you so much Katie. And then our final volunteer for the public comment session was Lee Cohen. Lee, thanks for joining us up here.

Lee Cohen:

Good morning. It's terrific that we're able to do this in a hybrid format with so many people joining us virtually, but I would be remiss if I didn't start by saying that it is terrific to be here in person and to see colleagues who many of us have not seen for frankly before the pandemic. And so I want to thank the organizers for putting together such an elegant hybrid meeting. My name is Dr. Lee Cohen and I'm the director of the Center for Women's Mental Health at Mass General Hospital in Boston and Professor of Psychiatry at Harvard Medical School. I'm also one of the principal investigators of the National Pregnancy Registry for Psychiatric Medications, which is based at MGH. The registry was established in proof of concept fashion back in 2008 on the shoulders of my colleagues, particularly Louis Holmes at the North American Anti-Epileptic Drug Registry, but with a focus at that time on new psychiatric medications that were coming to market and for which there were very sparse reproductive safety data.

Our initial work and work that continues was on the second generation antipsychotics such as aripiprazole, quetiapine, lurasidone, olanzapine, which were being used not only for psychotic illness but also as a mood stabilizer, an augmentation strategy for major depression and a wide variety of off-label uses such as anxiety, insomnia, and obsessive compulsive disorder. The registry has expanded to include study of newer antidepressants as well as sedative-hypnotics and psychostimulants. Over 3000 women have enrolled in the prospective registry, which includes an internal non-exposed control group to a medication being studied, but in women who've experienced psychiatric morbidity. Medical records are obtained after delivery to ascertain information across a host of obstetric, neonatal and neurobehavioral outcomes, but the primary outcome variable for the registry is risk for congenital malformations. Rate of procurement of medical records is 86% and retention of the cohort has been approximately 90% across the life of the registry.

Identified cases of presumed malformations are adjudicated by a blinded dysmorphologist and we've published extensively on the methods of the registry and results, which derive from various analysis of the registry data sets. The registry is funded with a multi manufacturer model outlined in the FDA guidance to industry on registries. We learned early on that in the case of second-generation antipsychotics, these medications were not teratogenic on the order of thalidomide. And then as the sample sizes grew, we learned that these medications as a class at least were not teratogenic on the order of sodium valproate. And now 10 to 12 years later, we can say with greater confidence, we can speak to the low likelihood of rates of congenital malformations that do not appear to be greater than what we see in the general population.

As a clinician scientist, and a reproductive psychiatrist specifically, what I've learned in over 35 years of taking care of psychiatrically ill patients before, during, and after pregnancy, and it so dovetails with the comments yesterday, is that no decision is risk-free and no decision is perfect. At the end of the day, if you ask me what I've learned in terms of taking care of this population is that at the end of the day, nothing is more important than maternal wellness.

So what is the ideal post approval surveillance methods to assess safety of medications used during pregnancy when there are 5.5 million births per year in the US and when 50% of pregnancies are unplanned and when use of medication during pregnancy is so common? Yesterday in the afternoon, a potential framework for surveillance was discussed. Indeed, we have more and more administrative databases of available with large numbers of pregnancies among birthing persons taking a variety of medications. But despite how attractive the possibilities are, given those large available data regarding a host of medications, what really is the quality of those data? Are they good enough to ascertain a signal of risk and are they good enough to use as a basis for what I tell patients in my clinic every Wednesday morning? I think the jury is frankly still out.

That being said, the path to meaningful data from a prospective medication registry is a long haul and small sample sizes, and at least initially wide confidence intervals, is a challenge. And so invariably an amalgam of methods will be probably ultimately used to delineate reproductive safety of medications as patient and clinician collaborate and as patients apply their own calculus as they weigh risk and benefit.

What did not get discussed yesterday directly? There was mention of the productivity of registries where there's a postmarketing requirement or a postmarketing commitment in place, but what about the majority of medications which are approved and where there is no PMR or PMC. Postmarketing surveillance for such medicines is a very substantial lift, absent any leverage to support such reproductive safety surveillance activities. And there was a question posed yesterday during one of the sessions to one of the panelists, and I would propose again that there is enormous variability across pharmaceutical manufacturers with respect to supporting the science that is described in this meeting. Many manufacturers may list an existing registry in the medication label as they are obliged to do so by FDA regulations, but these same companies will defer participation of support of registries as outlined in the FDA guidance into industry on registries. They don't really do it because they don't have to. These nuts and bolts issues of support of reproductive surveillance initiatives is really important and has to be incorporated into the formulation of any new postapproval surveillance framework.

I remain optimistic, frankly, that there is a path by which we can leverage the resources of FDA, academia, industry, healthcare systems, and advocacy groups to learn the most about the medications that birthing persons use to have the safest pregnancies possible while sustaining maternal wellbeing. Thank you.

Gerrit Hamre:

Thanks so much to all of our presenters, both in person and virtually. Now we're going to pass the microphone over to Dr. Evan Myers, who will be moderating the last two sessions of the day. For

Evan Myers:

Thanks, Gerrit. I'm Evan Myers. I'm a professor in the department of OB-GYN at Duke, an epidemiologist, core faculty member at Duke Margolis. I'm also Vice Chair of Duke's IRB for reproductive risks, so I review every protocol that both includes and excludes pregnancy. It's my pleasure to moderate these last two sessions.

In this session, we're going to explore the existing gaps in the design performance of different pregnancy safety study designs to better inform the development of the framework. Our presenters are Dr. Trish Bright, who's the Associate Director for FDA Sentinel System. She'll be sharing what we can expect in terms of the demonstration projects with the potential to expand our understanding of study design and data sources to further enhance the pregnancy safety framework.

Next will be Dr. Judy Maro, who's an Assistant Professor in the Department of Population Medicine at Harvard Medical School and Harvard Pilgrim Healthcare Institute. As the operations lead for the Sentinel Operations Center, she'll be sharing the role of the Sentinel System in monitoring for pregnancy safety.

And finally, Dr. Joanne Gruber is an epidemiologist in the Center for Biologics Evaluation and Research in the Office of Biostatistics and Pharmacovigilance in the CBER surveillance program at FDA. She'll be speaking on CBER's biologic effectiveness and safety or BEST Initiative and the surveillance of biologics in pregnancy. So with that, I'll turn it over to Dr. Bright for the first presentation.

Trish Bright:

Good morning and thank you for having me. So my name is Trish Bright. Let me see if I can get the clicker to work here. There we go. So my name is Trish Bright. I am the Sentinel Program Lead and I'm an FDA epidemiologist. I want to thank you all for coming here and we're going to be discussing the PDUFA VII Pregnancy Safety Demonstration Projects and how they might inform the Pregnancy Safety Framework by addressing knowledge gaps. I'll start by discussing the PDUFA VII Pregnancy Safety Commitments, then before I discuss the gaps in knowledge, we'll have an overview of what we already know. Then we'll discuss the demonstration projects "a" through "d" and how they can inform the Pregnancy Safety Framework, and you'll see what "a" through "d" are on the next slide. Then we'll discuss FDA's Active Risk and Analysis System or ARIA.

So you'll see on the left-hand side here is the Pregnancy Safety PDUFA VII commitment letter that some of my colleagues discussed yesterday. Today we'll discuss part (2) and part (3). So part (2) is incorporating feedback from part (1) and conducting five demonstration projects to address gaps in knowledge about the performance characteristics of different study designs. FDA will initiate the following demonstration projects, which may be modified as needed before September 30th, 2024. And you can see on the right-hand side there's (a) through (e). I'll be discussing (a) through (d) because my colleague will be discussing (e). And then part (3) is by September 30th 2027, based on the results of the demonstration projects in part (2), FDA will update the framework.

So what are the gaps in knowledge about performance characteristics of different study designs? We already know a few things about the data generated by different study designs. So this will be a bit of a review of what we discussed yesterday. Yesterday we heard people discussing registries and also discussing descriptive safety studies. These are the non-database studies. With the registries, we have systematic collection of pregnancy specific data in real time, and they offer a comparator, but it does require patients to enroll. The descriptive safety studies are without a comparator or predefined sample size, but they also involve systematic collection of pregnancy specific data in real time, but they also require patients to enroll.

So traditionally pregnancy registries have been the preliminary design or method to fulfill postapproval pregnancy studies. And Dr. Ajao discussed this a little bit yesterday when she showed us the historical trends. So why is that? Reading clinical narratives that include temporal sequences of events in the context of comparators helps clinicians to evaluate causality when there's a limitation imposed by small sample size. And data may become available sooner when each participant is enrolled. And we heard Dr. Shehan discussed this a little bit yesterday. Sorry, Sahin. And then could database studies also help inform safety assessments?

So again, a little bit more of a review. We discussed this yesterday. So now we're going to talk about database studies. We have signal detection and signal evaluation. A signal detection, we can have two options: no pre-specified outcome, like TreeScan, which offers broad coverage, non-specific confounding control and could be conducted at intervals as data accrues; with pre-specified outcomes, for example, sequential surveillance, rather than all outcomes, this approach involves a set of targeted

outcomes, and the power [to] detect is higher for targeted outcomes than for all outcomes. It has targeted confounding control and could be conducted at intervals as data accrues. For signal evaluation, we have a pre-specified outcome. Often we use an active comparator and we can use a new user design. Together this offers high internal validity and using real-world data to conduct longitudinal studies of medication safety and leveraging biostatistical techniques to mitigate bias and to conduct statistical testing of hypotheses. And this can be used to quantify rare events, and we may hear some more of this in the next presentation.

So what are we hoping to gain from the demonstration projects? The use-case demonstration projects will provide data to both inform and challenge our collective view of the strengths and limitations of the study designs for assessing pregnancy related outcomes in different contexts. Insights provided by the use cases can help us to update the proposed framework.

So in here I've grouped together project "a" and project "c" from the Pregnancy Safety Demonstration Projects as they were listed in the PDUFA VII commitment letter. Both of these projects involve pregnancy registries versus electronic healthcare database studies. Both involve exposures when the exposure is common. One approach will have signal detection, the other signal evaluation, and this will help us to understand the strengths and limitations of these two study designs and the study approach. Project "b" is a single arm safety study versus electronic healthcare database study. The approach is signal detection and the exposure will be rare, and this will help us to understand the strengths and limitations of these two study designs when the outcome is rare.

Project "d" will assess the performance and usefulness of major congenital malformations, or MCM, as a composite outcome using signal detection and signal evaluation when there's true risk for some, but not all specific malformations. This will help us to understand whether using MCM as the pregnancy-related outcome of interest is appropriate or compromises assessments in some context with a potential for dilution of effect. Assessing MCM is required in most of our pregnancy-related PMRs and MCM is relatively straightforward to studying claims, but understanding the performance characteristics of different methods and MCM algorithms is important for us to know.

So also in the PDUFA VII commitment letter, it allows for some modification of these proposed projects as the protocols are developed to help us better address the pregnancy framework gaps. So how might the findings of these four demonstration projects inform the proposed pregnancy safety framework? On the left-hand side, you're going to hear some of the context elements we heard yesterday, like the sample size, the market share, patterns of use, toxicity risk, severity of disease, et cetera. On the right-hand side are the study approaches that we just discussed. So we want to put these together, the context and the study approach, and determine in what context should a registry study, a database study, or both studies be required at approval. And I'm going to pause on this because this is an important question in what we're trying to address through this framework. In what context should a registry study, a database study, or both studies be required at approval?

We know the demonstration projects will tell us some things that help us with the proposed framework. For example, the demonstration projects will consider frequency of exposure. The demonstration projects will consider study design, non-database studies versus database studies. The demonstration projects will consider signal detection versus signal evaluation. The demonstration projects will consider whether MCM is an appropriate outcome. What else are we doing to address knowledge gaps? Well, the drug utilization information is still coming in and we'll include data on products not included in the current analysis. Data will also include further analyses such as based on year of approval or number of drugs in class, and we'll be developing approaches to estimate the magnitude of exposure.

So we're going to shift gears a little bit now and talk about the Sentinel System and FDA's Active Risk Identification and Analysis system, or ARIA, for those who may have less familiarity. The Sentinel

Initiative was launched in 2008 in response to the FDA Amendments Act, or FDAAA, in 2007, which mandated FDA to develop a post-market Active Risk Identification Analysis, or ARIA, system for medical products, incorporate data on at least 100 million patients by July 2012, from both public and private sources of healthcare data, have the capacity to both identify and evaluate safety concerns for medical products. On the lower right-hand side, you'll see some statutory provisions the FDA must work with. So FDAAA requires the FDA to determine whether ARIA is sufficient to assess a serious safety risk prior to requiring a sponsor to conduct a postmarket observational study of their medical product. And I've included here one of the guidance for industry that I find helpful. Now, on the lower right, you'll also see three purposes that were described yesterday in the presentations. Section 505(o)(3)(A) states that the postmarket studies and clinical trials may be required for any or all of these three purposes.

What I wanted to show you in the lower left-hand side here is a publication about how the Sentinel System was built out. It was built with the capabilities to address FDA's needs, and it started with a signal evaluation and worked its way down towards signal detection or signal generation. And the text in here is lifted from that publication in case anybody wants to go back and look at it.

I also wanted to show people that, on the upper right, you'll see a slide that looks familiar. This was presented yesterday and it describes how much do we know about the risk based on available information. And below you'll see also three categories that are akin to those purposes in the red box below, and you'll see them tied to different study goals including signal evaluation, signal detection. So in other words, the Sentinel System was designed to help FDA address the statutory needs. And we have the capabilities within the Sentinel System to address things like signal evaluation and signal detection, the very things we've been talking about in this workshop.

Now I want to talk a little bit about ARIA. ARIA uses a subset of Sentinel System's full capabilities to fulfill the FDAAA mandate to conduct active safety surveillance. ARIA is comprised of analytic tools and a common data model. The analytic tools are predefined, parametrized, and reusable programming or off-the-shelf programming, to enable faster safety surveillance in sentinel, in contrast to protocol-based assessments with full customized programming. The Sentinel Common Data Model is mostly electronic claims data transformed locally into the Sentinel Common Data Model format at each data partner and the data is kept with the data partner inside their firewall. Each data partner runs analyses using the data programs and sends the de-identified results to the Sentinel Operations Center. The Sentinel distributed database is a collection of harmonized data sets from many of the different data partners.

We're using the ARIA system to meet the FDAAA requirements for FDA to determine whether ARIA is sufficient to assess a serious safety risk. We keep metrics on our use of ARIA, including the challenges. On the next slide, I'll be showing a select set of metrics. This will be reflecting regulatory needs that were considered but did not use ARIA because there were challenges in capturing the outcome.

So this is the distribution of safety concerns insufficient for assessment in ARIA attributed to capture of health outcomes, and I know that the font is small, but if you look at the top row in red, you'll see that the majority of these were driven by pregnancy. And so often the outcomes as part of our pregnancy interests is studying MCM, and the demonstration projects will help us to better understand the performance characteristics of different methods and algorithms for MCM.

So in summary, the demonstration projects, other ongoing work, such as the drug utilization data, and importantly, feedback from this workshop, will provide insights to inform the proposed pregnancy safety framework. The demonstration projects will also help us to consider in what context a registry study, a database study, or both studies should be required at approval, but we don't expect one size to fit all. The demonstration projects will also help us to consider whether using MCM as a composite pregnancy-related outcome of interest is appropriate and what context it might be appropriate. So the next

presentation will provide more information about CDER's Sentinel System capabilities that could support the demonstration projects. So I want to thank you for listening.

Judith Maro:

Good morning, folks. Just checking that everybody can still hear me. I always have to lower the microphone at my height. So I'm going to pick up right where Dr. Bright left off. So talking about how we can use the Sentinel System to respond to FDA's regulatory needs in monitoring medication safety in pregnancy. And I want to say that this presentation is broken up into largely two parts. So the first part is to talk about and familiarize you with the data and the tools that have specifically been developed in response to what Trish just showed you, that finding of ARIA insufficiency. And then the second part of the presentation is to think about how those tools can be put to use in order to perform these demonstration projects. So I'm going to start out to say that, again, as Dr. Bright mentioned, the analytic tools were specifically designed to respond to FDA's regulatory needs in what we call those three prongs of FDAAA is the way that we refer to those. So those are the amendments act for monitoring the safety of medications in pregnant individuals.

So this is a chart that basically outlines the chart that Dr. Bright just picked up. It's just to say that, again, that top reason for why ARIA was deemed insufficient was the inability to adequately measure adverse pregnancy and fetal outcomes. And I want to say that the system has changed over time, so the launch of the Sentinel System officially is January 1st, 2016, and every time we sort of progress through time, we look at ways to invest in how to reduce ARIA insufficiencies by either adding data or tools. And so when we think about these inadequacies, the first thing that we think about is, which is the problem? Is the problem the data, or is the problem the tools, or both? And so the harder thing to fix, the longer leg to fix is the data. So we started there.

So this is a graphic that we have on the website and it shows over time how we have introduced enhancements to the Sentinel Common Data Model to fulfill FDA's regulatory needs. And so in 2018, again, starting in responses to these findings of ARIA insufficiency, we added a mother-infant linkage. And at that table we started off with commercial claims insurers who have access to protected health information for their members who can do linkage at a rate of approximately 70 to 80% in terms of the total deliveries, and then later we added Medicaid data. And in the Medicaid data that we added, we don't have access to that protected health information so the linkage rate is slightly lower.

So right now, this is a current state of affairs of live birth deliveries available in the Sentinel System. I'm focusing here on live birth deliveries as that was part of the insufficiencies, but we also have mechanisms to identify non-live birth deliveries. But with the live birth deliveries, total now, we have 13.5 million pregnancies with a live birth delivery over the years shown. You can see that there is a rise sort of in this middle area. This is where Medicaid data has been added in order to supplement the system, and you see that there's substantially higher numbers when Medicaid data has been added. All told, with those 13.5 million pregnancies, 10.8 million can be linked to an infant, so around 80%, which is the balance of the Medicaid averages and then the commercial claims averages.

But an important point here is that when we apply standard enrollment criteria, when we require enrollment during the entire pregnancy and a slight pre-pregnancy period of both medical and drug coverage, which is typically used to increase the internal validity of the epidemiological studies that have been discussed so that we can ascertain particular covariates of interest and things like that, we see a narrowing quite quickly. So we get to 7.3 million pregnancies after we put all those criteria into effect. So just something to consider.

So in addition to using that data, we turn our focus with ARIA insufficiency from the data to the tools. So how can we create tools that then can be used to access these data and to create evidence that can be

used for regulatory decision-making? And so consequent with the development of the mother-infant linkage table, we have developed a propensity score adjusted tools. And as Dr. Bright had mentioned, the focus first was on tools that dealt with targeted outcomes, right? So things for signal validation. And so this is just a little graphic to talk about, again, very similar to what Dr. Bright had mentioned. The goal here with the Sentinel System was to be able to address all of those three phases of signal management.

So from signal identification and detection, some descriptive safety studies, and then signal validation using a causal inference framework and common epidemiological designs, and then there's the ability to do that at a one-time study analysis, so that was what Dr. Bright was talking about with signal validation, and then to do them as data accrue. So those are over time with multiple hypothesis testing that's been generated. And just as she had said, just backing it up, these are tools that are parameterized at program execution. They've been pretested and quality controlled. They have a standard output.

And so once we build tools, we try to validate them. And so with the signal validation tool where we started with, we started with a study by Dr. Hernández-Díaz and Dr. Huybrechts who, for the folks who are virtual, are sitting right in front of me. So it's a study on topiramate and oral clefts, and we use these tools to do a conceptual replication of that study to ensure that we were getting the same types of answers. And as a standard procedure, because we are invested in the transparency of these tools, we do training on them after we develop them so that we can involve other people and collaborators in thinking about how to parameterize these tools, how to set up their own studies, how to use them when they have data that has been formatted into the Common Data Model. And this training was done in 2020.

So when we turn our attention from targeted outcomes now to untargeted outcomes, so this is where we're just, we call it scanning or data mining, we're scanning an area to see if there is any unusual occurrence of outcomes that may be associated with an exposure. We look at entire sections of the diagnostic tree of ICD-10 codes that are used in both billing and clinical data. And so we start out, so you see at the top of this graph, you start out with everything in what's known as the Q section of ICD-10. So this is everything related to congenital malformations, both major, minor, everything in that area. And then as we work down, you see that there is a narrowing by particular organ system. And then when you get to the very, very specific, so in this particular example at the bottom, you have spine disorders that are in particular regions of the spine. So you have the ability to do data mining and scan over this entire space from the very aggregate to the very, very specific.

But scanning takes power. So every time you look in one of these places, you're taking away from your overall study power to find something, and so there's a balance here between what you might want to do when you have a very specific hypothesis and you have a targeted outcome, in which case this kind of method is not really the most appropriate, versus when you don't know what you're looking for and you just want to be certain that you're scanning that entire space.

And so here, just again, sort of putting together the targeted outcome type studies and the untargeted outcome type studies, both of them identify a cohort, both of them create exposure episodes using medication dispensing. There was some discussion yesterday about some of the limitations for that, and part of the public comment this morning was about some of the limitations for that so I'm not going to go over that again. The main difference is really about the outcomes. So when we're doing a targeted outcome study, we're typically using a validated algorithm. When we're using a scanning process, we're not having a validated algorithm for everything we're scanning, we're just looking at the presence or absence of diagnostic codes. In both cases, propensity score methods are used to control for confounding, although in different ways because of the untargeted nature, there's sort of a different sort of context to what the propensity score means.

And then the different results that you get. When you have a targeted study, you're going to get an effect size and a confidence interval. When you have an untargeted study, what you're going to get is what we call an alert, which is what we say is pay more attention to what is going on here. There is an abnormal sort of a pattern of outcomes that are occurring that may or may not be associated with the exposure. They might be chance, they might be due to confounding, they might be a real thing. So what it's really telling you is, this is something I need to pay attention to. And so as we did before, when we validate, when we create a tool, we validate it. So we have two studies here that looked at one, what we think of as sort of a negative control, and others in a simulation study to show that what we are intending to do actually worked. And I should also say that a very similar study was performed by Dr. Huybrechts' group with a positive control demonstrating the same thing.

And so again, very similar pattern. When we develop a tool, we do the training. So in 2023 we did training on these tools and how they can be used by anybody who has data that's been formatted to the Common Data Model.

Judith Maro:

And this is, again, a commitment to transparency and also a commitment to allow people to comment on and use and tell us more about these tools. Back to ARIA Insufficiency. Now we've worked on investment over time in the data. We've worked on investment over time in the tools. Now what's still left? And so one of the things that's true about ARIA is the boundary of ARIA does not include medical record review, does not include chart validation. And so there is a concern about whether or not these outcomes are truly what we think they are, right? And so there have been multiple outcome validation studies in the ICD-10 period, particularly for terminal pregnancy outcomes, for live birth, for preterm birth, for stillbirth, for spontaneous abortion.

But there have been very few studies that have validated adverse infant outcomes in that ICD-10 space. There are a lot in the ICD-9 space, but not necessarily the ICD-10 space. And when I say a lot, that's relative. There are not a lot of validation studies overall. I don't want to give you the impression that there's a lot going around. With the current state of that investment in the Sentinel Analytic tools, how can we meet the needs of FDA in support of the pregnancy framework through these demonstration projects? This is just a little graphic, I'm not going to go through it because Dr. Bright did it so well of all of the commitments, A through D and E. But some of the key questions for the demonstration projects, and this is really echoing, I think, a lot of what I heard yesterday.

The goal is to generate real world data that is both accurate and timely and can be quickly converted, in this case via analytic methods or tools into real world evidence. And so I want to focus on these two features, the accuracy and the timeliness. And again, this is somewhat of a review. The registry based data, as we've been talking about, that uses primary research collection methods. And so we think that those outcomes might inherently have a higher level of accuracy because you're actually asking a person the question you want them to answer. Whereas in secondary use methods, you're just using the data that you have and there may be additional validation that needs to happen. Although yesterday we did hear from the CDC in their VSAFE program about if you're not really specific about what you're asking, you can still end up with registry data that's somewhat messy and difficult to parse. We also have timeliness concerns. How can we accrue this data as quickly as possible? And so we've heard about challenges with under enrollment and such.

Some of the key factors, I'm not going to belabor this, but just thinking about operational measures that we can incorporate into these demonstration projects, exposure sensitivity, exposure specificity, outcome sensitivity, and outcome specificity, particularly when there is a target outcome that is part of the process. One of the things I wanted to bring up to folks here is that it's important to really agree on

these outcome definitions even in a targeted analysis. This is a paper, it's from a group that does registries, and this is three different registries that looked at the same set of case data and said what they thought was a major congenital malformation. And you can see those three different numbers of prevalence are sort of all over the place. There's 46, there's 22, and there's seven. And these are different groups of people looking at the exact same data.

We also, in electronic healthcare data, if we don't well define our target, we can end up with prevalence from 2% to 20% depending on what we actually describe as a major congenital malformation. We really need to get our arms around a common definition that we all agree to so we all can be measuring the same thing. Some of the questions, again, bringing this back to the role of chart validation and how chart validation works in the system, and does it work differently if we're thinking about a Signal Detection Framework versus a Signal Evaluation Framework? And so in a Signal Detection Framework, is there a difference in terms of our tolerance for false positive versus false negative error with respect to the potential for chart validation? And I want to bring to you an example of something that we did as an experiment when we were doing a validation of lymphoma.

And so in this particular publication, we used something that's called a claims profile. This is a line list of all of the things, temporal sequence, of what happens to a person in their billing data. It's not the same as a chart, but it's actually really very comprehensive in many cases. And so we had clinician adjudicators look at these line lists and predict what they thought would be a true case of lymphoma. And so the two clinician adjudicators that we had tagged a number of, right there, the numbers were 87% where they got it right, where basically they predicted this would be a true case and then chart validation bore out that was a true case. And in the cases that they tagged but that weren't them, they were typically other related cancers. I'm not a clinician so I can't say what's related, but just that they were something that was in the neighborhood of lymphoma but was not lymphoma.

And so this might be a technique or an opportunity to do something because chart validation is really quite resource intensive and expensive. These claims profiles can be created quickly and they can be as a way to sort of screen things so that we're using our resources sort of in the most cost-effective and optimal way. Another thing that we talked about was how to actually treat major congenital malformations. Should we treat it like a composite where we look at 30 to 40 different things together and we look at them or do we create something like a modified tree? The tree that we've been working with in our targeted outcomes is much bigger than major congenital malformations. It includes many more codes and maybe we're spending time scanning there when we really don't want to. Maybe we could limit a tree to just the codes that comprise major congenital malformations and think about whether we're going to save power and how we're going to balance power and time to signal.

These are the types of questions that we're thinking about during these demonstration projects. And so some takeaways from the talk today. That Sentinel has the data and the tools available to perform the demonstration projects. That we're trying to explore and quantify material differences in the best approach. And we're paying particular attention to accuracy and timeliness in multiple approaches. But I would actually suggest that there's a third thing that we oftentimes don't consider because we think, oh, we're about the science, but this is addressed in the public comments today. And so I want to say something about it, which is that we have to think also about the cost-effectiveness of this. All of this requires a funding model and it requires a funding model that's sustainable and that creates the most evidence for the amount of resources that you have. And so thinking about how to do this in a cost-effective way is not really a very minor consideration. It's an important part of the overall picture so we can create a sustainable system.

The demonstration projects here are a framework that's designed to be generally useful. What we're trying to do is be robust over all of the possible situations, but we won't be able to sort of assess every

single thing, what I think of as corner cases, that are unique. We're trying to do something that's going to generalize. And the goal is to find conditions under which different approaches may be preferred. And as Dr. Bright mentioned, sometimes it may be one or the other or both. There is no expectation that a single approach is always and uniformly going to be preferable under all circumstances. And so with that, I'll turn it over to the next speaker.

Joann Gruber:

I think I used this mic. Okay, I'm going to... Good morning. My name is Joann Gruber. Let me see here. Let me get my slides. There we go. My name is Joann Gruber. I'm an epidemiologist in the Center for Biologics Evaluation and Research in the office of Biostatistics and Pharmacovigilance. Today I'll be speaking about CBER's Biologics Effectiveness and Safety or best initiative and the capabilities of the system to study safety of biologics used in pregnancy. Here's an outline of my presentation. I'll begin briefly talking about the demonstration project E in the PDUFA VII commitment letter. And then I'll move to introduce the Biologics Effectiveness and Safety or BEST initiative. And then finally, I'll talk about the capabilities of the BEST system to study the safety of biologics in pregnancy. As Dr. Bright, my CDER colleague has mentioned previously, CDER is taking on A through D in the demonstration projects, and CBER is in charge of the part E of the demonstration projects.

That is FDA's commitment to assess the performance of an algorithm using electronic health record or EHR and claims linked healthcare data for pregnancy related outcome or composite outcomes after use of vaccines in pregnant persons. Today I will discuss what CBER has done to build capabilities to fulfill this demonstration project, as well as what we will do in terms of studying safety of pregnancy and biologics. But before doing that, I like to briefly introduce the Biologics Effectiveness and Safety Initiative, or sometimes referred as the BEST Initiative, BEST System. I use these terms interchangeably. The FDA Amendments Act, which was introduced earlier of 2007, FDAAA 2007, mandated FDA build an active post-marketing safety surveillance system for FDA regulated products. In response to the mandate, FDA established the Sentinel Initiative. And so as mentioned previously, CDER uses the Sentinel system to monitor drug safety, and CBER uses the Biologics Effectiveness and Safety System, or the BEST System, to monitor the safety of biologics. That includes vaccines, blood components, and gene therapies and other biologic products.

I wanted to just briefly talk about the infrastructure of the BEST System. The BEST Initiative has access to a network of electronic health record or EHR databases, as well as large administrative claims databases. These databases are linked to immunization information systems. Those are IIS'es and they're local and state jurisdictions that collect information when people get vaccines through their public health departments. And so this was really important early in the COVID-19 pandemic where people were getting vaccinated outside where there weren't claims showing up in claims databases. That really has increased our capture of important vaccines. It was also important when we were studying the safety of vaccines used to prevent mpox, which were also distributed through our local and state health departments. These databases, in addition to having that linkage, we have a short data lag and adjustment for claims delay.

We're often looking at large amounts of vaccines given very quickly over a short period of time. And so we want the ability to study safety very quickly. And so we have a short data lag for that reason. For our databases, there is linkage to some medical charts that allows us to validate our claims. As was discussed previously, you're really seeing with the diagnosis code and how can we validate claims? And that is a real trade-off between resources because it is resource intensive. Our data are in a distributed data network as well. It's an expandable and adaptable common data. We use the OMOP common data model in the BEST System. BEST also has flexible analytic capabilities, and I found that that is really important when we're dealing with public health crises, like the COVID-19 pandemic has showed how

we have to be really adaptable and change to meeting priorities very quickly. I'd like to just quickly show the BEST data sources that we use just to highlight the diverse types of data. We have large administrative claims databases, electronic health record or EHR databases, and then link to EHR claims databases.

Next I'd like to discuss really what we're here about to talk about our capability to study safety of pregnancy and maternal fetal and infant outcomes. First I'd like to talk about the work that we've done for validating claims-based algorithms for identifying pregnancy outcomes in claims databases. Understanding the safety of biologics using pregnancy has been a real priority for CBER. And to conduct safety surveillance of biologics in pregnancy, BEST needs the capability to identify pregnancy outcomes using standard coding systems in the ICD-10 era, and as well as the ability to determine gestational age. As was discussed previously, that switch from ICD-9 to ICD-10 can really make you have to start all over again. And so this was some of the work that we had to do to do the validation. We decided to conduct a study so that we could know what we're seeing in our claims databases.

We developed claims-based algorithms to identify pregnancy outcomes and estimate gestational age in our administrative claims data. And then our second aim was to evaluate the performance of those claims-based algorithms by physician adjudication and link EHR charts. We did this study in the IBM Linked Claims-EMR database, and we studied persons age 12 to 55 years who are identified as female at the time of the outcome of interest, the pregnancy outcome, and who were continuously enrolled in medical benefits during the pregnancy episode and had the pregnancy outcome on or after August 1st, 2016. And that allowed those events to be in the ICD-10 era.

For aim one, again, we identified our pregnancy outcomes using our claims-based algorithms. And so those pregnancy outcomes included live births, still births, and spontaneous abortions, and we identified those using medical codes. And then we also used the best estimate of the pregnancy start date to then determine gestational age. Then for the validation, we did a sample of those pregnancy outcomes that were identified in our claims database. And then we used the structured EHR data and the global alignment of immunization safety assessment in pregnancy case definition to evaluate the performance of the claims-based algorithms. And then we estimated that percent agreement between those claims algorithms and our referent method and estimated our 95% confidence intervals. Here are the results of that study. The pregnancy outcomes are on the y-axis, and the x-axis shows that percent agreement between our claims algorithm and our referent method and the 95% confidence intervals. On the right side of the figure, you'll see the number of records that were reviewed for each of our pregnancy outcomes. Overall, you can see the claims-based algorithms did well at identifying our pregnancy outcomes with the range of the percent agreement from the claims algorithm and reference method being from 62 to 100%. These algorithms did particularly well at identifying full-term birth at 98% and spontaneous abortion at 100%.

Similarly, this figure shows the gestational age by our pregnancy outcome. Again, on the y-axis we have our pregnancy outcome. The x-axis again shows that percent agreements between the claims-based algorithm for gestational age and our referent method. And so the left panel shows that the agreement within seven days and the right panel shows the agreement within 14 days. You can see on the left panel, we had pretty good agreement when we allowed that seven-day window between our claims-based and referent method. With the range varying by outcome, it was 61 to 86%, and the highest agreement was for our full-term live births at 86%, and then our preterm live births at 82%. Moving to the right panel, when we allow that window to be within 14 days, we see a substantial increase in that agreement. That means most of our claims-based algorithm for gestational age within 14 days does a really good job at estimating the gestational age.

That range was 80 to 99%, and we had excellent agreement for full-term and preterm births at 99 and 93%, respectively. I went over that pretty quickly. There are a lot of really messy details in all of this work. If you'd like additional information, I encourage you to refer to our published protocol and our published report on the BEST Initiative website, and you can scan the QR code for a quick link there. And also we have published a manuscript summarizing the results as well. Before I go to the next thing I'd like to talk about, I just wanted to highlight some additional work that we've done to really get vaccine exposures also in pregnancy, building on the work that we've done to identify pregnancy outcomes and gestational ages. In this study, we looked at the vaccine exposures in pregnancy among privately and publicly insured women in the US. If you'd like additional information on that work, feel free to scan the QR code.

Next, I'd like to present the work that we've done to link mothers and infants on our claims databases. To study just the pregnancy outcome, you don't require that linkage, but if you want to know how does a vaccine that's given in pregnancy or a specific time during pregnancy, is there an impact on infant health later, we need that linkage. And so we set out to do that to link pregnant individuals to infants in our databases. Our aim was to link pregnant individuals with live deliveries to live borne infants in our claims databases.

I'll briefly touch on the methods. To do the linkage, we started with our claims databases, which were Carelon Research, CVS Health, and Optum. And we identified all live deliveries among females 12 to 55 years of age in our database who had a live delivery at least one year after our data start date, or August 1st, 2016 that allowed those episodes, again, to be in the ICD-10 era. And then next we identified liveborn infants who had a date of birth at least one year following our date of start date, or August 1st, 2016, whichever was later. And so then finally, we combined those records to see how many of these live deliveries linked to our live born infants. And to do that, we used a combination of subscriber ID and delivery and infant birthdates. The subscriber ID identifies the group of people on a given insurance plan, and individuals with the same subscriber ID are on the same plan and are usually part of the same family.

And so we require that the mother's subscriber ID match the infant subscriber ID as the first step. And then the second step, we also require that the mother's delivery date matched the infant birthdate. And so we allowed that match to be first an exact match. We required it exact, and then within three days and within seven days since there's some variability sometimes in those dates in our claims databases. In total, we identified, this is just the total deliveries by data source. And during the time periods listed, we identified over two million live deliveries, and that number varied by our data sources from over 1.2 million to just under 350,000 live deliveries. This figure shows our linkage rates for the mothers and infants, again, of all live deliveries, what percentage linked to a liveborn infant. And that is presented on the y-axis, that linkage percent. And on the x-axis, again, everybody had to match subscriber ID, and then we had that window that we allowed between the delivery date and the infant birthdate. Was it the same day, within three days or within seven days? Each of the colors represents our data partner. We saw that when we required a same day match of delivering infant birthday, we had pretty good linkage rates at 40 to 48%. But that increased quite dramatically when we allowed a little bit of give between those two like we see in our claims databases, those dates might be recorded at different times. We saw that linkage improved to about 70 to 81%. When we allowed that window to expand further, we didn't gain a lot. It was about 71 to 82%.

To just summarize, the BEST initiative is used by CBER to conduct post approval non-interventional safety studies of biologics, that includes our vaccines. And we've developed several capabilities within BEST to study the safety of biologics in pregnancy, including the ability to identify pregnancy outcomes and gestational ages using our claims-based algorithms in the ICD-10 era, and the ability to link live deliveries to liveborn infants in our claims databases. This, of course, all of this put together really

focuses on what we're here for, trying to determine and build evidence for safety of our medical products, including vaccines at the maternal, fetal, and infant stages. Our next steps will most likely be to evaluate the safety of a vaccine to fulfill the rest of the commitment with respect to pregnancy outcomes. I just wanted to acknowledge all of my CBER colleagues, some of whom are here, who have done a lot of work, and our FDA BEST partners who have done a tremendous amount of work to build this infrastructure. Thank you.

Evan Myers:

Okay. Excuse me. Now we're going to start the discussion, and I'm going to start with a question about gaps and challenges to secondary databases beyond just some of the methodologic ones, and in particular, Medicaid. When half of pregnancies are covered by Medicaid, they're obviously a very different population. They're at greater risk for a number of, and Judy already showed some of the issues with that data. What are some of the kind of broader methodologic questions in terms of generalizability, capturing appropriate populations and so forth?

Patricia Bright:

And I'll punt that to Judy.

Judith Maro:

A couple of things. One, Medicaid data are quite lagged. Typically, it's a two to three year wait depending on, so Medicaid, they release through their VRDC or res deck program. They release two releases every year, they're annual releases. And so the data, I think just from a perspective of monitoring, they're quite lagged relative to other, although there's a lot covered. The other thing that I highlighted was the enrollment criteria. A lot of individuals qualify for Medicaid as a function of getting pregnant, and so they don't actually start their coverage until they're already pregnant and may miss the most etiologically relevant window of their capture between three weeks and eight weeks, as was discussed yesterday. And so some of the challenge is just the missing data from when they aren't covered and you're not seeing what's happening to those individuals.

Also, we've done extensive amount of work looking at the differences in jurisdictions. There are 53 jurisdictions in Medicaid data, and each of those has a fee for service and a comprehensive managed care component to it, both of which the quality is different in those different data systems. And there are a lot of different rules. As Medicaid expands differently in different states, you have different coverage, you have different ability to follow, and you have different sort of access to prenatal care and pregnancy benefits. Those are all sorts of unique challenges of Medicaid as a data source, but I still think it's invaluable. So with all those limitations, I still think it's invaluable.

Joann Gruber:

I was just going to add on, too, because I do think that's one of the challenges we've had. It is important to cover as many pregnancies as you can and to get as representative of a population. So I was thinking the same thing about the in-bias that we have requiring when you start looking prior to pregnancy, through pregnancy, after pregnancy, it's a lot of enrollment for our fragmented system. And if people are only getting into Medicaid because of pregnancy, that's a challenge. And then one of the challenges we have is we're trying to do things as quickly as possible when a new vaccine comes out. And so to have that lag, it's really challenging for us when we're thinking about vaccines, particularly, to how to figure out how to include that population. Though we would really like to find ways to incorporate all as much comprehensive coverage as we can in terms of pregnancy.

Megan Clowse:

Yeah, this is Megan Clowse, rheumatology at Duke University. One of my main concerns that always comes into my head whenever I read a paper about using health claims data is whether the patient actually put the medicines in their body and when. Do you guys have an approach to validating that and so we have a better understanding or statistical approaches to manage that issue?

Joann Gruber:

Get vaccines.

Judith Maro:

Yeah, I guess I can start. So typically our base assumption is that if it's been dispensed, we consider that as a medication that's taken, but you can do various sensitivities to essentially test those assumptions and you can use quantitative bias analysis to try to overcome some of that. But in terms of validating, I mean really you need to do survey data. You need to send out surveys to folks and say, for example, I mean it would be an expensive program, but you could give people a mobile login at the pharmacy and ask them to log in and report if they've actually taken their medications or not. Now again, there's I guess concerns about people have about their private health information being misused, and so in the current information environments, stuff like that is very difficult to do effectively and correctly. So it's a challenge.

Patricia Bright:

Something else we can do is if somebody's taking a medication chronically, you can look and see, well, are they getting the subsequent refilled, the subsequent one after that refilled or not? Because if they get it once, even if it's chronic, but then they don't get it refilled again, that's a hint. But that only works with some kinds of medications and we could do a sensitivity analysis to compare those folks to other folks.

Evan Myers:

Here's a question from Slido. With increased use of claims data to support pregnancy safety studies, it seems that we should also be advocating for changes to medical diagnosis and prescription recording to help address some of the limitations that we face when using claims data. Has this been discussed and is there push back? And I assume that's trying to come up with standardizations for coding and practices across specialties and so forth.

Judith Maro:

I think in terms of the electronic health record and changing the components of it in order to ensure that you're capturing more information, one of the things that was brought up yesterday in the panel was whether the onus of that belongs, then, to the provider who is giving that. So if the provider is required to record that the patient reports that they're actually taking their medications as prescribed and things like that, that all of that additional recording, it just adds to the system. And there's a question about whether the value add, I think, is appropriate given the resources that it would take to do that or whether we fall back on techniques like quantitative bias analysis and sensitivity as a means to try to overcome that.

Elsie Grace:

Hi Elsie Grace from Eli Lilly. I have a question that kind of steps back to the earlier parts of the presentation and the framework. So I really look forward to seeing the results of the demonstration project. So thank you for all the work that's going to be going into that. I know that a lot of what we were talking about are claims study versus registry, and I know that we talked yesterday about the surveillance approach and I'm curious. I hear you when you say the demonstration projects aren't meant to test everything, but I'm curious about where the FDA's thoughts are on where that surveillance approach fits in and when that might be more considered.

Judith Maro:

I'm not the FDA, so I'll stay quiet.

Patricia Bright:

Yeah, I mean I know we use all data streams. Okay, so in this talk we've been focusing mostly on registries and database studies, but we heard yesterday that some very rich information comes in from industry. And so we have to think about all of it. And one of the points I made is not one size fits all. And so the question might be how do we get an efficient response and are there tricks we can use to prioritize certain medications to be followed, weighting different types of studies into that full feedback so we have a more comprehensive approach. I'm not sure we're there yet, and thinking about how the pharmacovigilance on the different type of products is going to complete itself in the end, but it's something we're definitely thinking about.

Speaker 7:

I'll turn it on for you. Keep talking.

Simone:

Simone [inaudible 01:18:09], AbbVie. Quick question on the framework as well, and I really look forward to learning more about the thoughtful projects being demonstration projects that will inform the framework. I have a general question in terms of the key elements for the framework and I would love discussions on that. At the time of approval, oftentimes some of that information, such as exposure of the drug or number of exposed pregnancies, is not known. So I'm wondering if you could comment on how that could affect, once the PMR is issued but things didn't pan out, the product's not being used as we thought it would be used, or women are getting pregnant, being exposed to the product, how they'll affect the discussions on the framework or approaches, the selection of approaches that would produce the most meaningful results to obtain informational pregnancy safety.

Patricia Bright:

It's a little bit of a tricky question because we're still focused on plan A and you're asking about plan B. But yeah, it's definitely real. I mean, we're trying to come up with a prediction model to estimate, and that's going to be part of the work that we're doing, but we know that it's going to be somewhat limited. We heard some representatives yesterday kind of saying, "Hey, we thought this would be used in a certain population and then it wasn't." So as every group gets together to kind of consider this, they're going to always have to think what's our fallback? What do we do if? And a lot of those discussions happen at the FDA and are internal, and so we're aware of it, but I think at this point, unfortunately, we can't sort of give you a pat answer.

Evan Myers:

But I would think, at least at the level by the time you go to approval, you'll have data from your trials of the proportion of women of reproductive age and who needed to use contraception during the trial. So that at least gives you a ballpark to start with.

Speaker 7:

We have one more question on the side.

Christine Olson:

I'm Christine Olsen, CDC. Thank you. This is really informative and I hope I didn't miss this earlier, but I think you focused on the congenital malformations and the pregnancy outcomes. I'm curious how well these systems work for something that's really important but really hard to study and that's longer term as development. I'm thinking particularly some medications may be more than others, and some of the psychiatric medications where that may be more of a concern, and I understand how difficult that is just from clinical diagnostics and the variability and how it's coded, but the coding systems seem like they might be the best option for these types of administrative systems for getting that information. But I don't know how well that performs or how much that's been looked at. Thank you.

Patricia Bright:

Yeah, I would say that is one of the more challenging outcomes that we have to look at. I have seen them in PMRs and teams have worked on those, but when you're looking at somebody, the developmental characteristics, two years, five years, 10 years, and to get that kind of follow up, it is hard. It's challenging. We don't have a lot that have gone down that track, but it is certainly an important outcome and we do need better ways to approach it.

Lee Cohen:

Hi Lee Cohen. A question and a comment. I agree so much, your comment about adherence and where that folds in, and I was wondering Dr. Maro, after your talk about is there any way to sort of think about a nested test where you could take a sampling of patients who are presumably taking a medicine and really, in the context of informed consent, actually going to see what to measure adherence so there's some way to factor adherence into the model.

It was just sort of... Because I agree with Dr. Clowse that having this discussion without really knowing whether the patient is taking the medicine is a challenge. It's a challenge to do that even to the patient sitting across from you in the office, and so let alone if you're looking at claims data. But I was so impressed with your, Dr. Maro, your presentation, and I was curious about your vision for trying to, in a way look at real data in relation to the claims data to see the extent to which it reflected reality or well at least what was in the chart.

And do you have a vision for scaling that? That's really my question because that would probably provide a level of confidence as we use those sorts of data that we just don't have at this point in time. And so I was sort of wondering your thoughts about that.

Judith Maro:

Sure. So a part of the Sentinel system that we didn't talk about today that was utilized previously, it was called the FDA Catalyst system, and the FDA catalyst system was a system that was not public health surveillance. So to be clear, the Sentinel system is public health surveillance and the Catalyst system was research, and because it was research, there was informed consent.

And so there were several demonstration projects, and the goal there was to use the claims-based information as sort of a patient finder. I'm going to find people and then I'm going to, through the commercial claims insurer, get permission to contact them or their provider and ask them some questions. And if you do that for a subsample, then you have an informed consent, you have the ability to do that. So all of that is possible. It just costs money. So it's all about resources and the funding model to be able to do that. But there are demonstrated projects that essentially showed you could do that. And so if you use that technique to find a subset of people to send them a survey, there was actually an approved survey device. It was called My Studies app.

You can still find at the FDA My Studies app where that survey device, a patient could log onto it. They could answer a short survey, and it would be connected to their longitudinal record. So you could look to see to incorporate those patient reported responses and see how that measures up relative to what you see in the record in terms of their longitudinality. And that is an optimal data collection system, but I'm not going to shy away from it. It's an expensive one. And so you just have to have the appropriate funding model in place to be able to do it.

Evan Myers:

Here's a question for Trish. Are there annual reporting requirements for ongoing pregnancy studies as part of a PMR?

Patricia Bright:

Yeah, so each PMR has milestones on it, and when they issue the PMR, they tell you how often you have to report. And so some do have annual reporting. I'm not sure that all do because I haven't been a reviewer for a while, but I used to look at a lot of annual reports.

Speaker 8:

They do.

Patricia Bright:

They do, okay,

Geeta Swamy:

So following from some of this... Geeta Swamy from Duke University, following from some of the discussion yesterday, I'm wondering about looking at the safety from the perspective of maternal health and outcomes. So when we focus on, for example, preterm birth, and if we think of that as a poor outcome for the infant, many times when individuals have medical conditions they're taking medications for, that preterm birth occurred because it's actually a treatment to improve the pregnancy outcome overall, avoiding stillbirth in the setting of severe preeclampsia and so forth. In addition, looking at things like preeclampsia, looking like hospital length of stay, readmissions. So are those factors that would be considered in a safety setting as well? Because if we focus, again, only on the infant outcomes, I don't think we're thinking about this from the perspective of the safety and why we're even talking about these medications in the first place.

Patricia Bright:

Completely agree that those are important. They're not required in all PMRs. Some things, MCM tends to be required in most PMRs. They're definitely required in some. And so the review teams are the ones

who sort of make the distinction. They look at the biological plausibility, they think about it, and they decide what outcomes will be in the PMRs and what won't be. But we have been doing sort of a lot of work outside of... We're focusing on MCM today, but there has been a lot of work in the pregnancy field in general, which includes the mothers and looking at the safety of the mothers and developing tools to address the needs of the mothers as well.

Speaker 9:

Going back to the feasibility question about what information we have at the time of marketing, could you talk a little bit about how the drug utilization studies might play a role here?

Patricia Bright:

Yeah, so one of the things that they're trying to do is, I think we heard yesterday, is we're thinking about how can we put together this framework and this model and trying to assess what use looks like. So for example, are those products that are later in class, do they tend to have less uptake? And if they have less uptake, are they going to have less sample size? If they have less sample size, we'd only be able to detect a risk of a certain level and that sort of thing. And so the drug utilization data will feed into that thinking, feed into the sample size and that sort of thing. I don't know if anybody else wants to add.

Judith Maro:

I guess I would just say that I think that what was outlined yesterday was the complementary nature of having the claims-based information to try to understand if the prevalence that it was being observed was actually because people really aren't using this product in this population. And so you could envision a potential outcome here where there are checkpoints or triggers that need to be met in order for certain things to be able to be completed, and you can continue to monitor that utilization as you go. So let's do a checkpoint at six months after approval, at one year after approval, and have delayed kick-in of requirements because that constitutes a new safety piece of information.

Evan Myers:

Are there gaps or challenges or things that we're still not going to know at the end of the demonstration projects? And so one that occurs to me is for many women with severe conditions, termination is part of the options. And because for a number of reasons that may not be in claims data, we won't be able to tell how many terminations there were, whether the termination was because of medical conditions or a congenital anomaly. Are there other things that we're still going to want to do? A demonstration project round two? Questions that we're need to address further?

Patricia Bright:

So I don't know that we'd be able to do demonstration projects because, like you said, there's some that are going to be outside of this kind of data. Another example, similar to what you were saying, is very early spontaneous abortions before it comes to medical attention where the woman just has the miscarriage at home. And so some of that won't be captured either. So I'm not sure if there's other areas like that where we could build out additional work, but we do have some flexibility in our demonstration project. So when the work group gets together to consider what to do and where the gaps are and how to use the resources wisely to get at that, they'll be thinking about some of those points.

Judith Maro:

I guess I would just add that I think we should really strongly consider our current political environment with respect to elective termination. Pregnancy registries, I think, are going to suffer from people who do not want to record some of that information in our current environment, and it's going to make it really challenging to be able to understand what's really happening there.

Katherine Wisner:

I wonder if we're thinking about different kinds of outcomes. So when I think about, say, individual patient and I want to understand what's the dose, what's the exposure, maybe plasma level and what's reproductive outcome, we have this really tight kind of toxicity model. And if we're looking at what is the reproductive safety on a population basis, it's a different kind of concept where you're really getting almost a parallel, like an effectiveness study, where you're integrating all of the different, did they take it? Did they not take it? What about all the other factors? So I think maybe we're trying to do something that's very precise, that's not very precise, but maybe more valid in terms of generalizability as far as what happens when you release a drug into a population. So I wonder what your thoughts are about that.

Joann Gruber:

Yeah, I definitely think population level is very different than individual levels. And at the individual level, you often don't know, but at the population level you're thinking, here's what we see. And I think that gets at a point that you had made yesterday or that was on the panel yesterday about absolute risk and relative risk. I think a lot of times we're looking at relative because partially there's a lot of we're thinking about, "Well, this is a decision, so you could do this and that. So relative to this, should I do that?" But overall, we're not contextualizing it, and I think that's something that we need to do in our studies. What's that absolute risk? And that is a population level phenomenon.

Patricia Bright:

I would just also say that's why we have our epidemiologists and clinicians work together.

Evan Myers:

And it also seemed just from a power, if you're powered to look for relatively rare outcomes, you're going to have pretty good precision to look for some of these more common maternal outcomes and help them.

Do you have any additional questions?

I think we've kind of covered the canned ones.

Gerrit Hamre:

Any other questions from either the audience or online? I'll give it just another moment. We did start a little bit early after the public comment, so if anything, we'll have a little bit of extra time for lunch.

Evan Myers:

Never a bad thing.

Judith Maro:

Agreed. Oh, I think there was a question.

Speaker 3:

Someone behind you.

Jessica Albano:

Hi, Jessica Albano, Syneos Health. I guess just to extend a little bit on, I guess, the question earlier around ability to accurately, then, in the changing political environment around pregnancy outcomes, thoughts from FDA on maybe relaxing. If we know gestational age at outcome and we know live birth versus non-live birth, to what specificity do we need to know the exact type of outcome and what kind of flexibility? I know the post-marketing requirements, they very explicitly state the specific pregnancy outcomes that have to be collected, but is there an opportunity for some reconsideration of how those are worded to allow more flexibility for both, whether it's claims or other database or pregnancy registries, to collect the necessary information without collecting it in such a sensitive way that might preclude being able to actually have accurate outcomes?

Patricia Bright:

Yeah, so I would say that there is flexibility. It was worded, but it was also there's a section there that say this can be modified as needed. And the reason is because when they wrote the commitment letter, it was so long ago, and they have to allow room for those who are doing the work to think about it, just like you're saying. This time, this place, what are the issues now? How has the science changed? And that sort of thing. And so I think that they're hearing what you're saying in your comments at this workshop and that that's going to help inform how they think about the demonstration projects and what factors and what outcomes they might look at. So this kind of feedback is good because it'll help make these demonstration projects stronger.

Joann Gruber:

I'd also just add and reiterate too that this is a distributed data network and privacy and protection of information is paramount. And in these systems, Sentinel System, VSD, the BEST system, we don't have access to identifiable data. And I think that's really important to remind everyone that it is behind a firewall, behind another firewall, behind another firewall. So I think that's something, when we're thinking about the electronic health record databases, that we can remind, too, that there's privacy.

Evan Myers:

Okay. Well, if there aren't any other questions, I think we'll take an early lunch and we'll see everybody back in this afternoon.

Gerrit Hamre:

We'll begin again promptly at 1:05.

Session 6

Evan Myers:

... experts in pharmacoepidemiology and pharmacovigilance in maternal and pediatric health who are going to provide feedback on the proposed pregnancy safety study framework along with two representatives from the FDA who will be able to respond and expand on the feedback.

And if you're like me and your postprandial short-term memory is not the best, I'm going to introduce each panelist right before they give their presentation rather than all at once. So we'll dive right in and start with Marie Teil, who's the Global Head, Women of Childbearing Age Program at UCB, biopharma.

Marie Teil:

Thank you. Thank you very much. Can you hear me? Yeah, thank you very much. I'm really delighted to be here. Thank you for the opportunity to have this discussion. And I want to say I've dedicated more than 10 years in pharma working on trying to close the data gap for women of childbearing age with chronic diseases. We're mostly in chronic diseases, so women who do not necessarily say have a disease of pregnancy actually, but may have uncontrolled disease during their pregnancies.

So we've been trying, like industry has been trying to do it all. We do pharmacovigilance, we look at enhanced pharmacovigilance. We try, especially in pregnancy, to look very, very carefully our cases. We have teratologists coming, reviewing these cases. We do prospective registry studies, we do claims database. We even went to do clinical trials like PK in pregnancy and placental transfer and milk transfer.

So a whole scope of studies that I think are very important for women, and we heard that from Mariah yesterday and we hear that all the time. So I really applaud the commitment of the Agency for this work. I know there's a tremendous amount of work behind it. It's really great to see that and we're looking forward to having a framework that we can use consistently and transparently for all our PMRs.

So as we're discussing this framework, there are a couple of key considerations that I would like to highlight. And the first one, so first of all, the privilege of coming at the end of the two days is that I don't think I have anything new to tell you, but what I'm going to do is try to summarize that or at least the key points that for industry are really important to consider in the framework.

The first one is the risk. How do we define risk for these people, pregnant women, who have various thresholds and various diseases? So we talk about major congenital malformation, we talk about teratologic effect of a drug. Absolutely! This is super important. But then we need to think about the risk to the child of a mother having uncontrolled disease during pregnancy.

And it has been said yesterday, it has been said again and again, but I think it's really crucial that we think about how do we integrate in this framework, this risk of the fetus- to the fetus- of having a mother with uncontrolled disease. And it can be huge, especially in certain diseases. So I think we really need to think about it and how we do that.

The second point is about sample size, not surprisingly. Sample size is a challenge for all of us, I think. And we heard that again. And when we talk about magnitude of exposure, of course the size of the population of interest is important, but we need to think when we talk about percentage of exposure is that of something that has been happening again and again practically and that we hear from physician is that women would stop their treatment at positive pregnancy test, especially for unplanned pregnancy. But unfortunately still now in plan, but definitely for unplanned. So the fact is we're not going to have huge sample size. It takes years. We saw the Sentinel data, we saw that. So how are we integrating this small sample size in signal detection and in signal evaluation for this particular framework? And how are we interpreting this data? Because the interpretation is actually what matters to patients and physicians.

And lastly, very much in the same vein is because the data are sparse, it's not one study with one data source that is going to help us. We need various data sources. All these studies are complimentary or could be complimentary. And I think we need to consider with this framework addressing multiple data sets and really looking at what we have available today. Leyla, you talked about pharmacovigilance data

yesterday. Yes, there's a lot of data in pharmacovigilance database and we do enhanced pharmacovigilance. So how can this data be used to enrich the data form for these studies? How can we use data from clinical trials? We have some data. They are not large because we still exclude pregnant women from our studies, but we have some data and we plan now to actually do protocol where we continue following these women during pregnancy, if they get pregnant during the clinical trial. So let's think about pulling all this data together. Sometimes there are PK data that are available. Shouldn't we include them in the framework and get a better sense of the context for these patients?

So that's pretty much my three points. I have a lot more I could share, maybe during the discussion, but at least as a start and I want to leave some space for everyone. So again, thank you. I think we need to keep in mind while we're doing that, we're doing that for the patients, for the physicians that need to have this discussion. And so we need to inform them with the right information at the right time. And I would argue the right time is yesterday almost. So let's think that way. And I think by having this forum, we are already making huge progress in that direction. So thank you.

Evan Myers:

Thank you Marie. Next we have Sonia Hernández-Díaz, who's a professor of epidemiology at the Harvard T.H. Chan School of Public Health, where she serves as Director of the Pharmacoepidemiology and Real World Evidence Program. Sonia?

Sonia Hernández-Díaz:

Yeah, thank you. I would like to thank the FDA for creating this forum of collaboration and also I applaud the consistent systematic framework that is being proposed while being flexible and dynamic and having the demonstration projects to kind of pilot test it. So I would like to give a little history from my perspective.

Around 20 plus years ago Dr. Allen Mitchell and others that are here in the audience, Leyla, Tina was here yesterday, Sara Ephross. They proposed a framework, a proactive surveillance that will include all the different sources of data, including first the registries proposed as the first line of defense against Thalidomide that will enroll 5500 exposed pregnancies as soon as possible after approval and they will detect major teratogens. That was the goal, and knowing that it is kind of impossible to power a pregnancy registry to identify small effects on specific malformations that happen in one in 1000 to one in 10000 births.

And another advantage of the registries is the quality of data. Having direct patient information on whether the person took the medication and also on outcomes and the quality and the range of outcomes that one can obtain from direct patient or direct to clinician information. But then the surveillance proposal included the databases that will identify more moderate teratogens or you can also consider effects on other outcomes.

And that as data accumulated, they were going to narrow the boundaries around the risks. So the intention was not to test safe or not, but to narrow what we know so that we can rule out a strong risk. And then it was also proposed that if a signal is identified, other designs like case control studies could follow up and look at the associations with specific birth defects.

So that was 20 years ago. I think some things have changed now in the sense that the timing of when to identify the signals is not that clearly different between the registries and the current databases that as we heard this morning, had huge sample sizes and some of them have timely access to the data. So the investigators in the past had to wait for the pregnancies to be exposed for most outcomes, nine or so months for you to observe the outcomes and that's only perinatal outcomes.

And then if the data was recorded in administrative records, the investigators had to wait say three years to have access. Now for some databases that data lag time is kind of six months. So the benefits in time between registries and databases is no longer one of the major differences. However, the difference in quality and outcomes remain. So I think when comparing the two sources, I will focus on that aspect mostly.

Having said that, small numbers is a challenge for everybody. We saw yesterday that around 60% of PMRs are not completed because they cannot enroll in sufficient numbers. And of those completed, we have information 2 to 14 years later. So if there are no exposed pregnancies out there, neither registries nor databases can find exposed pregnancies and if they are exposed pregnancies, we cannot wait 14 years. So I think that's where this framework is going to help operationalize in a systematic way. And because of that, I will like to start with giving two recommendations and then we will have plenty of time to discuss them.

One, I would recommend to the framework for it to include internal dynamic rules that allow for an adaptive design, a priori of what would you do if you find this or if you don't find that, if you accumulate example of expose or not.

And another recommendation would be for the demonstration projects to be realistic and use either examples from the literature or from the PMR to use examples that are more connected with reality because except for things like the vaccines in the pandemic, we do not have 1000 to 10000 exposed. We have more like 10 to 500, in that range. So that's my initial recommendation and I'm looking forward to the discussion.

Evan Myers:

Thank you, Sonia. Next we have Krista Huybrechts, who's an Associate Professor of Medicine and Epidemiology at Harvard Medical School in Harvard T.H. Chan School of Public Health, an Adjunct Associate Professor at Boston University School of Public Health. She's co-founder and co-director of the Harvard program on perinatal and pediatric pharmacoepidemiology.

Krista Huybrechts:

Thank you. Good afternoon, and thank you very much to the organizers for the opportunity to join this panel today. As my colleagues have done, I will share some reflections on the draft framework based on a review of the framework and then also informed by the discussions and questions of the past day and a half. I will shift a little bit more to the specifics on the framework itself. And I'd like to start by joining again my colleagues in thanking the FDA team for basically the thoroughness with which they're sort of approaching what is a really challenging question, but one that is in desperate need for an answer.

So first, as has been mentioned a number of times during this workshop, I believe it's really important that we keep in mind that as a guiding principle that none of these studies is perfect. Each has its advantages and disadvantages, and as Sonia mentioned, each was initially conceived of with a different purpose in mind. So in that sense, I really don't see them as alternatives to each other, but more as complimentary approaches when trying to answer questions about drug safety in pregnancy. And I think each of them can really shed a different light on the question that we have.

Secondly, there's been quite some discussion about the limitations of some of these approaches in the context of misclassification of exposure, misclassification of the outcome and confounding. And it has or the implications in terms of study size have sort of been highlighted. But what this sort of suggests is that the size of the study or what it suggests is that it's sort of a framework of statistical significance testing because no matter how large your studies, if you have a systematic bias, that bias is going to be there and your point estimate is going to be off.

And I'm not convinced that we really want to go towards statistical significance testing, but the goal is really to get as good an estimate as possible of the magnitude of the effect and the uncertainty that we have around that effect. So we are very interested in what is that upper limit of that 95% confidence interval because that will allow us to rule out strong teratogens.

So rather than really focusing on the study size and therefore on the statistical significance, one component that I think is really important is to think about other tools that we have available. And it was mentioned this morning as well, we do have quantitative bias analysis that allow us to assess what the potential impact on that point estimate, on that confidence interval is of potential systematic errors regardless of which approach that we're using.

Then in the current draft framework, there's sort of two important driving factors. So one is the minimum required sample size and the other one is the number of expected exposed pregnancies. But I believe that these characteristics are really not fixed for a given drug, but are sort of a characteristic of the drug outcome diet that we're studying. And I just want to sort of explain a little bit what I mean here is, as we all know, the minimum required sample size is not only driven by the relative risk or the risk difference, but by the prevalence of the outcome.

And most of the studies we're going to be doing are going to be looking at a whole range of different outcomes. So in that sense, it's a little bit of a hard concept to sort of wrap your head around. It's like what will be the minimum required sample size for a given drug? And then related to that, depending on the outcome we're studying, the ecologically relevant time window will be different.

So again, when thinking about what is the number of exposed pregnancies, well it will depend on the outcome because it depends on the ecologically relevant time window. So for those two reasons, I'm really not sure when we're trying to answer the question, what is the expected study size that we need - it's a characteristic of the drug, but it's really that drug-outcome association.

Closely related to that, like another driving factor in the framework is the outcomes of interest. And it's been mentioned a number of times. We don't know based on randomized control trials what the signals are. We know that animal studies don't really translate well to human situations. And then drug structure and function don't really give us a lot of information. So although sometimes based on the preclinical data we might have some concerns and some signal that we want to evaluate further, I think in most cases I wonder don't we really want to evaluate a whole range of potential outcomes.

So again, I think in most situations at the time of drug approval, you're really faced with a situation that you're interested in evaluating this whole range of potential outcomes. Then coming back to the validity considerations, I think we have learned a lot in the last decade or so in terms of the importance of confounding, especially confounding by the underlying indication and the factors associated with it.

And we know that different data sources have different strengths and weaknesses in terms of the types of information, confounding information that they have available. But one thing, and that sort of goes back to the comment that has been made a couple of times is like we're dealing with a very limited sample size in the first couple of years. So regardless of how rich the confounding information is that we might have available in those first couple of years, I think we sort of have to accept that there's only so much we can do in terms of adjusting for confounding.

We really need large numbers in order to be able to take advantage of those rich information that we have available. And that brings me to the last point that I wanted to touch on in my opening remarks. I think a formal decision treat type framework such as the one that we've been discussing here is incredibly useful. One point that I wanted to make is it seems like we are really focusing this whole process on trying to identify what the optimal study is. And I'm wondering whether that is really a question that has an answer.

Like, is there really one optimal study or should we really take advantage of the different types of study approaches? And one component that seems to be missing in that process is some, and I think Sonia was referring to that, some pre-specified decision points. It's a dynamic process and early on when we're going to have very few pregnancies, the best we're going to be able to do is sort of monitor use, start describing what we sort of see happening in these pregnancies, whether it's based on a registry, whether it's based on a database.

It would be really important then to get some consensus as to how many exposed pregnancies do we feel is sufficient for a given outcome to then start moving towards looking at an association, looking at causal inference. Accepting that initially is going to be with relatively limited confounding control. As we have more exposed pregnancies, we can become more strict in terms of adjustment for confounding.

Similar for some of the approaches that were discussed this morning when we're using more of a scanning approach to sort of try and detect signals. We know from the simulation studies that have been done that we really need a large number of exposed women often in the thousands. Not a single drug is going to have that many exposed pregnancies initially. So what is the time point?

Can we pre-specify how many exposed pregnancies that we need before we start using these approaches? And perhaps if we can sort of focus on these dynamic decision criteria that we can start generating the evidence as the information accumulates over time, as it accumulates in real time while capitalizing on the strength of each of these different approaches. So I'm going to leave it here for my opening remarks.

Evan Myers:

Thank you, Krista. And finally we have Janet Hardy, who is currently an independent consultant but has extensive experience in academia, government and industry as a perinatal pharmacoepidemiologist, including serving as Executive Director and Head of Pharmacoepidemiology at Biohaven Pharmaceuticals and Pfizer. Janet.

Janet R Hardy:

Good afternoon everyone, and thank you. I'm deeply grateful for the invitation to join this panel and to be part of this important workshop. And I just want to say that my statements reflect my own opinions. And I appreciate the effort that's gone into organizing this workshop and to the efforts that the FDA has put into analyzing the data that was presented to us and developing this important framework. We do need to advance this area.

And so efforts to put together a framework are really valuable. I'm in agreement fully with everything that's been said so far. I can't dispute any of that. I'm a firm believer in the contribution of multiple study types and contributing to a body of evidence and registries and databases are complimentary of each other. I can't emphasize that enough as my colleagues have already. And I also want to address the presentations so far today that have focused on databases because we do have an enormous amount of information that come from databases, whether it's individually, Sentinel. They too have their limitations.

And I want to emphasize the importance of using them for the right question. They can be incredibly strong and helpful with the right type of question. It's not just a blanket study question that can be addressed and have them be relevant. So I think it's really right question, right study and complimentary studies. And so addressing some of their limitations, I want to backtrack and say that in recent years, recent decades, so much effort and progress has been made in developing very elegant algorithms and analytics strategies to use databases.

However, as already indicated, they are still limited by the data that goes into them and the data that is missing. And some of that missing data reflects our underserved and minority communities and that should be recognized. So thinking of the study question that best address their strengths is important. I also want to just make a point about looking at databases and expecting prevalence of exposure to pop up as being very meaningful.

The FDA sends PMRs and PMCs that are based by indication, and that's how we define our groups, our cohorts by exposed and unexposed. And I think it's important to recognize that depending on the indication, so going back to again the question that is asked, not every prenatal visit and not every historical visit in the pregnant person's record has a check mark encoded visit for that indication. So you may not be able to identify your populations that are truly representative of prevalence.

It's not that easy. So it may also indicate that there are not that many people in the database that you're going to find, and that's one of the points that is worth noting. So it may not reflect all of your exposed people. I also think that it's very important to ask the right question of databases based on the type of exposure you have. So these are oversimplified, but I think they are relevant examples. I think we also have a limited type of outcome that can be asked of these databases. That has also been mentioned.

Something that hasn't been mentioned but was addressed in the presentations was the value of chart review to validate outcome studies. It is very true that these are a valuable tool, but in practice, let's not forget that chart reviews can be requested. It's not often that you see 100% requests come back and that you get all the information that you want. In fact, many times the yield can be quite low and at tremendous cost. So what you have in terms of generalizability is worth reflecting on.

Next point, single outcome studies. I think some of my colleagues already know my feelings about this. It was mentioned earlier. I would like to see further rationale for this being part of the framework. As it currently stands, I believe this strategy may be taking us a little bit backwards in time, and I'm concerned about focusing on findings from a single source if that's where it's coming from, perhaps such as an animal study.

If I understood correctly, we know that animal models may not always represent the humans, but if we are to focus on a single outcome study, I'm concerned about loss of precious time and having a broader array of outcomes, having them staggered in time if need be because we rarely conduct clinical trials pre-approval for involving pregnant patients. And isn't our purpose still active surveillance, in fact enhanced active surveillance? Which leads me to want to refocus us on the important questions of the patient.

A person who becomes pregnant is generally, or contemplating pregnancy, is generally wanting to know, is this drug, molecule or vaccine going to be safe to take in pregnancy? They're going to be really interested if you have a piece of information that says, well, this drug caused for example, oral facial clefts, that's really nice to know, but I think they're really interested in the broad array of outcomes.

And so keeping that in mind when you're looking at a reduced or limited or even single outcome. And I do have an alternative proposition for consideration and that is to look at innovative and efficient methodology that maximizes what our current understanding of complimentary studies are, including the use of digital technologies and have our registries and databases be more inclusive of underserved populations. Let's reach out to them and be better at what we're doing right now. Have them support each other in terms of validation.

And importantly, as has been mentioned already for the proposed studies, but in our existing studies have built in milestones that include a feasibility assessment at let's say year three and several years there later thereafter, having agreements with the FDA to say, if we're not meeting our targets, let's stop. Have a discussion and say either we re-strategize or refocus our target and look again in the next

couple of years or three years. And if it's not working, there needs to be an agreement to either have a decision made or have some stopping rules.

And lastly, I want to make a proposal, since we're talking about adding PK studies, I think it would be appropriate to ask for the world and think about going beyond PDUFA and including generics in some of our studies. Those are women, pregnant people experience pregnancies on generics as well and those would enhance sample size and the ability to study. So those should be considered. I'll turn it back to general discussion.

Evan Myers:

Thank you, Janet. Now we have Lynne Yao who's the Director of the Division of Pediatric and Maternal Health in the Office of New Drugs at FDA and Bob Ball who's the Deputy Director of the Office of Surveillance and Epidemiology. And Lynne and Bob, I don't know if you want to comment on some of the things you've heard just now, things broader from the past couple of days.

Robert Ball:

So I have one question that I would like to ask the panelists. Picking up on a point that Dr. Hardy just made about using electronic healthcare databases in a focused way because one of the ... we've heard very broadly about all the many issues that are involved in assessing and using drug safety information in pregnancy. And you heard from Dr. Maro this morning about the Sentinel program and from Dr. Bright before about the ARIA program. And so, one of our somewhat narrow questions in the context of all this other large issues is how do we optimize the use of the ARIA system for pregnancy safety assessment? And so I think you heard from Dr. Hua yesterday and some of the other FDA presenters that the vast majority of PMRs that are issued are for signal identification.

And so, one of the questions we're really trying to address is can we use the ARIA system and some of the tools like the TreeScan methodology in a particular way, let's say focusing on a certain number of major congenital malformations to at least contribute some useful information to this whole array of questions that are in front of us.

Krista Huybrechts:

I'm happy to take the first attempt.

Marie Teil:

Go ahead.

Krista Huybrechts:

So I think the data can obviously be used. I mean if you're talking about, I mean you made reference to the approach such as TreeScan, I think the challenge remains like the number of exposed pregnancies if you want to avoid running an analysis and ending up with no signals that are all false negative signals because you just didn't have the power. So that is one thing I think we need to be very cognizant of. I think traditionally we had been more focused on false positives because there was this concern if you don't have a predefined hypothesis, you're screening across all of these potential outcomes, you're going to have all these false positives. But I think what we've come to realize is it's really the false negatives that are probably more of a concern than the false positives. So that is one thing. I think there's definitely a minimum requirement in terms of number of exposed in order to get any useful information out of an approach like that.

The other issue I think is important to mention is in a massive data source such as the Sentinel, you're basically using a lot of the information on pregnant individuals that you have available in the U.S. And if you're using it for signal identification or signal detection, the next step will then be signal evaluation. And one of the question that brings up is then what do you use for signal evaluation? Because you sort of used all of your available pregnancies already. And in the context of some of the work that we've done, which is completely outside of the FDA work is like we've reached out when we did a study in the U.S. and found a signal reached out, for example, to our colleagues in the Scandinavian countries to see could they replicate it and do they get the same signal? Or for another study that we're actually in the midst of doing right now where we're looking at antipsychotics, we found a potential signal that had not formally been identified and we reached out to Dr. Lee Cohen to see whether in the context of the registry, perhaps they found to say, but I think you're going to need that complimentary data source if you're using such massive data source for signal identification in order to confirm this thing. So I think those are two important challenges when using such an approach.

Evan Myers:

Anybody else have anything?

Lynne Yao:

I did want to make sure I heard correctly. So just a question of clarification. The framework that we proposed I think was intended to come up with an optimal study design at time X, if you will. And what I'm hearing from I think all of you and the panel is that time X is not necessarily the optimal time. So there's the optimal study for an optimal time and what's not included in the framework yet is this idea of repeating the process or at least not relying on a single time point to make a final decision about what study might be the best. I just want to make sure that I've heard that correctly.

Janet R Hardy:

If I may just make a point, I would agree with you actually. I think these are long studies. It takes a long time once the drug hits the market to penetrate, the population has to understand the uptake in pregnancy and it's a dynamic process, I think that can't be emphasized enough. Sonia made the point earlier, as did my other colleagues, and I think the point about putting milestones in protocols. Situations change, other competitor drugs come on the markets, other alternatives come out. We have to be flexible and going in, you just never know what the uptake will be. And women stop their drugs was the opposing point.

Sonia Hernández-Díaz:

Yeah, I totally agree with the dynamic and how the framework can have more of a decision tree approach, including if there is signal detection and there are signals, then what do you do afterwards? And I think under the same discussion, the decisions about safety or for the label that are made at point A might be different five years later or 10 years later. And that's okay, that's not a mistake.

You have to do some recommendations with the information you have at approval and those might be different from five years or 10 years afterwards, not because there was an error, but because we gather more information. But what to do at each of the times, I agree it's a dynamic process that it would be great if you could have in the framework some allowance for that planning of what to do with what.

Janet R Hardy:

And isn't that how science evolves as more information comes along?

Krista Huybrechts:

But maybe just to clarify one point, I think a lot of these decisions or these criteria can be known or can be predefined at the start when you sort of have to make your decision, but the process itself would have to remain flexible, right? It's not that you have no idea what's going to happen in five years and therefore you need to completely reevaluate, but you can sort of now set some criteria that you then evaluate in five years to decide which of the two roads you'll continue on.

Marie Teil:

Yeah, I agree. I think it would be really important to have this a priori set up of milestones but not going blind into it, really knowing where you are going and depending on what you get, adapting your framework and making it more flexible.

Krista Huybrechts:

Definitely.

Evan Myers:

I'd like to follow up on the point that I think has been made by all the panelists and I think throughout the two days is that the decision involves many considerations other than the risk of major congenital malformations and that there are formal mathematical ways to weigh the trade-offs and use them for sample size.

As Krista said, what degree of certainty do you need to do? What kind of process should there be to kind of think about what those outcomes should be for a given drug or class of drug? Who should be providing the input to the sponsors to FDA? Is it clinicians? Is it patients? Should there be a formal process for soliciting that input and when should that occur in the approval process? Sonia.

Sonia Hernández-Díaz:

I can start the discussion. I think there was a discussion for vaccines some years ago, and the guidelines have a minimum set common set of outcomes that are if you use the lower hanging fruit because of their significance in morbidity and mortality like major congenital malformations to either frequency like preterm or NICU admission. So there can be a minimum set of common outcomes I will propose. And then depending on the drug maybe because if we are aware of some toxicity or physiopathologically there is expected outcome, then ad hoc add some of additional outcomes of interest for each particular drug.

Evan Myers:

And should benefit be maternal or neonatal benefit be included in that as well?

Sonia Hernández-Díaz:

Are you asking me?

Evan Myers:

Well just I'm throwing that out.

Sonia Hernández-Díaz:

I will allow my clinician friends to respond to that one.

Evan Myers:

Marie.

Marie Teil:

So yes, I think they should indeed. And I would say that a way to perhaps do it is to look by therapeutic areas. We have a pretty good idea by therapeutic area, by disease and clinicians have a good idea about the natural history of the disease, of course with the current standard of care.

And so we know quite a bit about these risks and having that as a baseline, having that as a common way. Now the whole challenge is who's doing that and how are we going to set that up? Because I think this is not a one pharma thing, it's a whole disease area.

And something we can think about then maybe as an option is to think of the equivalent of what Europe has done with IMI in a public private partnership, where actually the companies are working with the public to make a consortium and have this type with pre-specified disease area and that are prioritized and then starting looking consistently at that. So that could be a way to think about it.

But having them as a baseline for all therapeutic areas would maybe be a way. And we see the antiretroviral registry, which has a lot of information because it's done at that level.

Evan Myers:

Any audience questions? Garrett? Dr. Cohen has a question.

Lee Cohen:

Thanks. First of all, thank you for such an extraordinary series of comments that sort of fold science into the practical aspects of delivering care. And so this is a question for anybody on the panel, and Dr. Yao, it really goes to your point about should pharmacovigilance sort of be dynamic.

And so I agree, but I sort of wonder what do you do then, and I pitched this to my colleagues from Boston when you then publish in New England about the relationship between lamotrigine and clefts. When you have a positive finding, how do you walk it back when time then proves that to be frankly not as great a concern as was published? And what do patients and colleagues do?

And Krista, really goes to your comment that in a way false negatives implies that we just haven't gone far enough sometimes because we're underpowered. But I worry because false negatives imply that we have to sort of move forward and get larger samples. It's false positives that ends up on CNN and petrifies patients. And so I don't know the answer to that question, but I would pitch it to the panel.

Sonia Hernández-Díaz:

Since I think you were referring to your Boston colleague, the lamotrigine finding was neither us nor in the New England. But it's a good point. And when we had then the topiramate and clefts, what we did was to try to replicate it simultaneously in the registry data, in the database and in the case control study. And in that occasion it was replicated.

But we find that we are in that situation, as you know all the time, what do you do when you have few exposed and one or two malformations that make it into a 10 fold increased risk, but with huge confidence intervals? And how to responsibly react to that because when you find it, you do not know if it is true or not. You know it was false only a few years later. So I think that's an important question. And

even for the HIV pregnancy registry that has been going on for many, many years, they publish everything every six months.

So how to make it public without alarming necessarily the users. I think that's a great question and I don't have the complete answer. I think that I agree, we have a responsibility of not going to CNN before doing our best and going first to our peers, to the scientific community, to regulatory agencies like, hey, we have that. We didn't expect it, and try to be as responsible as possible without hiding information. So I think that that can be part of the framework. When do you release data? I think that might be a good.

Janet R Hardy:

I think that's a phenomenal point and it's been a concern for years and probably won't go away because we have greater access to information for everyone now and access to information by people who won't necessarily understand or see or take the time to read the caveats and the confidence intervals around that piece of information and may react in fear or litigation.

And at the worst case, the drug is litigated off the market. It's a tremendously challenging question. And as data accrues, if you see a change and a signal goes away, that would be optimal. But when to release early information or not to release information, I don't know if there's really an easy answer to that, but it's an important one.

Evan Myers:

Krista.

Krista Huybrechts:

Yes. I just want to make sure to clarify one point that I mentioned, my intention was not to give the message that if you don't pick up any signals, it must mean that you're underpowered, right? My point was like we know that if you're sort of scanning across hundreds of different outcomes and you have to adjust for the multiplicity of testing a few hundred patients, you're guaranteed not to pick up the signal. So that's sort of in that context.

It's not like no signal means you must be underpowered and you need to keep looking. That's not the message that I wanted to give. And then in terms of what do you do when you have a finding and it gets published? I mean that's sort of what I was alluding to earlier.

I think it's really important if it's the first time that we identify this signal that we try and replicate it before it sort of goes out to the public, either in another data source of the same kind or ideally in a different type of data source, but at least one other study where you see the same signal gives you some reassurance that it could still be confounding, it always could, but it gives you more reassurance that maybe something real is going on versus you really are dealing with a random error and a chance finding.

And then in terms of, I think, and I've had many conversations with a lot of you here, I think a big component that we're still dealing with is risk communication. Yesterday we had the discussion even there's different qualities of studies that are out there and we sort of rely on the peer review process to sort of distinguish between what are good and what are not so great studies.

But in essence, I mean it's still the providers and the patients who have to sort through all of this literature, which is often very conflicting. And I think that whole risk communication piece and how do we communicate to providers, to patients what is potentially a real risk versus what is just noise in the literature is an enormous task ahead of us, I think still. But they're also focusing on absolute risk and risk

differences, I think that was mentioned at some point this morning is also important to put into context any potential large relative risks that we might be seeing in an individual study.

Evan Myers:

Marie.

Marie Teil:

So my point was exactly that on risk communication, not only when, but how do we communicate that risk? I think we have a tendency to really put it out very quickly because we want that to happen very dramatically because we think it may be a risk. I think everyone, all of us have a different threshold for risk, and this is where the context of the need of the patients need to be considered as well.

And I think the sad part about it is that when we communicate like that without the context around it, patients tend to not get into registry, not continue the treatment, which is absolutely fair. But then we have less data, so how do we validate? So it's kind of this vicious circle that we cannot get out of. So it would be fantastic, of course, if we find the magic recipe to do that. Unfortunately we leave it a little bit to the physicians and to the clinicians to have this discussion with the patients. But having some rules and some a priori decisions on that would be really, really helpful, I think.

Evan Myers:

Jean. Jean. No, Jean. Oh, Lynne. I'm sorry.

Lynne Yao:

No, no problem. Actually, I wanted to ask another clarifying question that is sort of a follow on to Lee's question. We have talked about an optimal study, that of the studies that are possible or feasible, what is optimal? And what I'm hearing, what I've heard from the four of you is that there are pluses and minuses, pros and cons to all of the studies we've suggested, the database and the registry and that they should be thought of as complimentary.

So my question is, when you move beyond the signal detection or signal identification to the signal evaluation, it sounds like there's a need, what I heard was replication, what I heard was kind of this confirmation that almost you're going to need more than one type of one set of one confirmatory additional set of data or additional methodology to really understand that signal to confirm it or to evaluate it fully. Did I get that right? Is that what I'm hearing?

Evan Myers:

Sonia.

Sonia Hernández-Díaz:

I think that in general, yes, but that there might be situations where you might not need to. For example, if we go back to thalidomide, if you had this framework at that time and you started a registry, after 50 exposed, you will know that there is something going on and then you will of course not try to replicate that. But that's exceptions, thankfully. So I think that in general, yes, but there may be situations where you might take a different route in the middle of the path.

Lynne Yao:

Sure, sure.

Robert Ball:

Can I just ask a follow-up to that? So one of the shifts that we're proposing in the framework is from kind of a, I call it almost a shotgun approach, to saying, well, let's quantitatively assess the trade-offs between quality of data, size of data available, timeliness of likely achieving a completed study.

And it seems like I'm hearing that the panelists don't think that that's likely to work. And so some of this has to do with, let me just give a concrete example. Let's suppose you were able to do a registry study, but it took 10 years to get an answer, but in a claims database, you could do it in seven years, but it was just a signal identification study.

But the registry study, because it has better data, would at the same time 10 years, also be able to do an evaluation, not just the signal detection. So you could imagine that the next phase of the claims data signal identification would be some kind of assessment, maybe it could be in another database or in another system that would do the evaluation. And I guess the question is, is that sequential approach likely to be better in any circumstances than what I would call the shotgun approach?

Evan Myers:

Krista.

Krista Huybrechts:

So I just wanted to make one comment in terms of the pros and cons of the different approaches. When you said, for example, with the registry, what if you can use it for signal identification and because the quality of the data is better, then you can also go on to signal evaluation. But I think the reality is given no matter how successful that a given registry is, if you want to go down to specific malformations, I don't know whether other than maybe the antiretroviral registry where you have sort of all of the different drugs contributing to the same registry, I'm not sure it's realistic to expect that no matter how good the data are that you'll ever be in the situation that you'll have enough to be able to start looking at these individual malformations to give you an example.

So I think that's where sort of they're not truly alternatives in the sense that maybe in that sense, a database study can allow you to look there. You bring up like, well, can you ever go to signal evaluation there? And I think to a large extent, it also really depends on how is the study designed and not so much the data source itself. And I think maybe there are some places in the framework where we would benefit from a little bit more of a distinction between there's the data source with pros and cons, but then there's also the design and the analytic approach with pros and cons.

And I agree, if you're just going to look for diagnostic codes, it's not going to be a very valid outcome. But if you're looking for multiple diagnostic codes on different days, some procedure codes, if you have validated in a subset of your sample, those outcomes, I don't think it's necessarily the case that you can say we can't do signal evaluation in the context of a database.

For some outcomes, yes, for others, not. The same with, we're now starting to focus more on also looking at non-life birth outcomes. I mean, the data are messy and you're going to have conflicting outcomes. Sometimes you see a still birth followed by a live birth three weeks later, but you can either use the data as is and then you have bad quality data, or you can sort of look or work very carefully through some hierarchical algorithms that give you a lot more reassurance. So I really think we would benefit from distinguishing data source both with strengths and weaknesses and design and analysis, both with strengths and weaknesses.

Evan Myers:

Megan first.

Megan Clowse:

So I'm glad I'm talking right now because actually your question sort of is related to what I was going to say. And that is, now that I've been in this field and taking care of patients for 20 years, what I've seen is that medications sort of go through an evolution over probably a 10 to 20 year period of use in pregnancy.

So it starts off, nobody purposefully uses a new medicine in pregnancy and what you have are a few accidental pregnancies, but some of those women have been exposed to the medicine, some of them haven't because accidental, according to the doctor, is often not accidental to the patient. And she actually knew that she was trying to get pregnant and stop the medicine.

So there's those sorts of confusions, but initially there's very little use. Then there's sort of an accrual of some data that maybe isn't terribly worrisome. And so then maybe some people start using it and then maybe some more start using it. And then maybe another signal comes out that is worrisome. A signal that comes out that's worrisome will cut way down, at least in my field, the use of a medication at which point doing any kind of pregnancy registry is going to be exceptionally challenging.

And so, I'm a little concerned with the idea of sort of moving forward with only database for a while until there's something that we need to study because you lose those 10 years or five years or whatever of accidental pregnancies, planned pregnancies, and you can't go back and replicate that data. Sure, you can do a database study later because they're all sort of just floating in there, but you can't really do like a, is the patient taking the drug? What week did she take the drug? Teratology, physical exams, that sort of thing. You can't generate that 10 years down the line.

Kathy Wisner:

Thank you. I wanted to get back to this idea about risk communication because I think there are ways that we can really mitigate the negative effects of publications that come out with adverse effects related to exposure. And I've seen this used much more effectively, I would say in the last five years. And that is, if a major publication comes out, there's an editorial link to it.

An example is for our JAMA psychiatry board, there was a recent publication of serial MRIs in children exposed to SSRIs in pregnancy. And the investigators found changes in the brain related to exposure to SSRIs, but also to depression. And what happened was there was a very good, well-thought-out editorial published with it that ended with the comment, these findings should not change current clinical practice.

And I saw the media picking up the editorial comments as well. So I think we have a responsibility to be publishing along with these kinds of papers, a more thoughtful understanding, the whole picture type of response. And I think that is really an effective way to at least mitigate some of the difficulties.

Evan Myers:

And there's a follow-up question related to that is how should that be communicated in the label? I don't know if the panelists have any suggestions.

Lynne Yao:

Wait, somebody used it yesterday, so I'll use it again. It depends.

Speaker 11:

It's the right answer.

Robert Ball:

Yeah, I was just going to say, not about the labeling per se, but FDA has many different ways of communicating risks, not just labeling. So Drug Safety Communication, Dear Doctor Letters. There's a whole suite of tools. I think one of the things that FDA typically does when there's a signal that's of concern, but there hasn't been time to fully evaluate it is looks through all the available evidence at that moment and then kind of assembles one of these communications, whatever, targets it to the public, but also to professional societies, whatever the most appropriate thing is.

So I think the FDA does have a very robust risk communication strategy in this light, but there does need to be a broader approach to communicating these kinds of risks.

Lynne Yao:

Can I also say that we will, and I think it's been mentioned by Leyla yesterday, we will get on the phone with our colleagues internationally, through the cluster, to ask what they've seen, what they've heard. I think that's a really important point that we want to make sure that we've checked it out with our regulatory colleagues around the globe, and then also consider what effect that kind of communication would have, not just in the U.S. but in other parts of the world too. So we would definitely talk to our colleagues too as part of trying to figure out what the communication should be.

Jessica Albano:

Hi there everyone, Jessica Albano, Syneos Health. Love the attention that the APR is getting. But I do want to just add on to something I spoke about yesterday and just kind of bring it to the attention again. But the importance of the advisory committee is really important, and it can help with this communication and determining of risk and how and when to report things.

We've got a fabulous advisory committee that includes members from CDC, NIH, FDA as well as practicing physicians who are both reporting and not reporting to the registry. So they're on the ground, they're the ones counseling the patients and have a wealth of information and experiences that they can relay.

And we also have a patient advocate who can ask the question of how do we communicate this? If I'm sitting with a patient in lay language, I need you to distill this down to me in a way that I can communicate it so that we can think through those things. And whether it's a registry or a database, that's really an invaluable perspective to have those multiple viewpoints giving input and direction because certainly as the investigators, the operational side, that's not our place. And it's not just the place of the sponsoring companies either. They really do need that direction and input.

Sara Ephross:

Thank you, Sara Ephross, Syneos Health. I think that the discussion that's being had now about sequential versus parallel approaches is really important because I think we've made really good strides over the last couple of decades, going from a sequential framework where there was a pregnancy registry with a potential signal, then we looked at databases, then if the potential signal held in the database, then we went to a case control approach.

And I think that only elongated the time to either support or refute the original signal or potential signal that was identified in the pregnancy registry. So I think whatever the eventual framework is, it's at least

to my mind, going to need to allow for in parallel evaluation, recognizing the strengths and limitations of each of these kinds of studies rather than maybe taking us backwards to a more sequential process.

But as we've heard, and so many of us in this room know it depends. So there needs to be flexibility because at least the ones that I've seen, none of these questions could necessarily follow a predefined, predetermined, non-question dependent approach. Thank you.

Robert Ball:

Can I just ask a question, which is somewhat in follow up to that? So yesterday there was some discussion of a national registry. And there's many different sources of data. So there's Sentinel, there's BEST just within FDA. The CDC has its systems. The data holders themselves can do their own studies or companies or others can use those data sources.

But one of the questions that keeps coming up, not just in pregnancy safety, but it feels like very particular here, is what does the ideal system look like and how do we all make it sustainable? So it's not just like a piecemeal FDA does a PMR, a company does a database study, FDA does something in ARIA. How do we build that sustainable system that combines claims, EHRs, registries, pharmacovigilance, all these different data sources?

Sonia Hernández-Díaz:

If we can dream, if we could connect everybody, and that mean if we didn't have confidentiality issues, that would be ideal because we could identify the exposed, but them help the follow-up by linking them with electronic health records and claims at the same time. I don't think that's possible right now, but if you ask what would be the ideal situation would be to combine all of the above.

Robert Ball:

So you think that the biggest barrier right now is just the privacy and exchange of information across systems?

Sonia Hernández-Díaz:

Well, I would not call privacy a barrier, but because we have to care about it, that will not allow linking and therefore improving the quality of the claims, but also expanding the identification of exposed and the follow-up of those that enroll in registries. So if all the systems could communicate, I think that would facilitate our lives for any of the studies.

Evan Myers:

Janet.

Janet R Hardy:

I think that there's two types of studies actually that you're talking about having everybody involved in one database and one registry. Did I catch that?

Sonia Hernández-Díaz:

I was actually even linking those two ideal studies.

Janet R Hardy:

So in any case, thank you. I think it also needs some agreement and interest from sponsors because they are the developers of our therapeutics, and it takes a common protocol in many cases. And we have very motivated sponsors, and we have also a recognition that this takes money. And I think that comes back to the FDA to have the regulatory authority to be asking for these types of studies. The cogs of the wheel move slowly, but there's multiple parties to move forward with this type of structure.

Evan Myers:

How big a challenge is variability in how things are coded or if we're using EHR, how they're documented? I mean, is there additional work besides just if you could just link everything electronically, how similar is the data set from the Mayo Clinic versus Kaiser versus some of these other ones? Would there be some additional work in terms of consensus on common data elements, common definitions, and so forth that would have to be done?

Sonia Hernández-Díaz:

Sure. Sure. Sentinel, for example, has this common data model. So again, that's a pie in the sky I was dreaming of doing. I'm not saying that that can be done tomorrow, but ideally if everybody was under the same kind of healthcare structure electronic health record database and that could be linked somehow to the like nested pregnancy registries or the other way around I think then, the two main designs or data sources that we discussed, they could reinforce each other to the point of being one if you wish at the end and it would require some data.

Krista Huybrechts:

And while, I mean, I won't argue that standardization is not good, but I think sometimes you can benefit from having the differences between the different systems, right? If one sort of has better capture of some kind of information, even though it's sort of only a smaller subset of patients that you have, maybe you can test some of your assumptions or your concerns in the main data in the smaller subset.

I mean, sometimes we do studies where we link a subset to the electronic medical record system like in the local Boston hospitals then, but at least to try and get a better sense, we're concerned about confounding by X, is it confirmed in this subset or not? So I think sometimes standardization is good, but sometimes we can take advantage of potential differences in terms of richness of information and differences in coding.

Amy Ramanadham:

Hi, this is Amy Ramanadham. I'm FDA. Quick question from a virtual attendee, and I think it's a good question directed towards our non-FDA panelists. What do you think about communicating equivocal information in drug labeling? If some data suggests risks and some do not and a causal association is unclear, should we put that in labeling or wait? I think there's a lot of people listening that would be interested in that response.

Krista Huybrechts:

That's risky now to say something. As a non FDA non-regulatory person, I wonder whether there is information. I mean there's a difference between absence of any evidence versus there is some evidence and it's clearly not pointing towards a strong risk side. So I'm not sure being very ambivalent to say some yes, some no, we really don't know.

But providing some information that yes, there are some studies that have been done and at least no clear very strong risks have been identified. Maybe more modest increases can be excluded. It seems to me that for providers and for patients that there is value in that information. How exactly to frame it, that's a different point. But I think even if the information is more ambivalent, I do think there seems to be value in it.

Janet R Hardy:

Would it be worthwhile in bringing back the, it depends? But also I thoroughly agree with what you've just said, but also the number of studies that are contributing to that ambivalence and quality studies, I think also bears keeping in mind.

Lee Cohen:

But Lynne, share with us, correct me if I'm wrong with following the establishment of the PLLR, we have a precedent because there are many labels that include language, which highlights the fact that there may be data pointing in one direction and other data that didn't support that. And so that already exists.

I think it's a fantastic question from whoever shared that virtually because that's reality. And I think sharing that ambiguity with our colleagues, providers is a good thing that the answer may not be nailed down at that exact moment. I think it's actually, it's a wise thing. Dr. Ball, I think your point was just so critical.

I think the greatest challenge is firewalls and when folks have tried to do what you just were describing, managing firewalls can be very challenging. And actually when those firewalls are removed, we pitch it to very seasoned data managers to frankly navigate the very real issues of compatibility of different data sources.

But those have sort of been managed more than sometimes the amount of time that goes by while we're trying to navigate agreements in firewalls between various stakeholders in the space.

Elsie Grace:

Hi, I'm Elsie Grace from Eli Lilly again. I have a question about this idea of taking a sequential approach. And I've been wondering how we might think about this within the context of a specific study design. So for example, a registry. If you think about, I assume this is how everyone does it, but we develop a full protocol where we talk about we're going to do these propensity score mass methods and we put a lot of resources into developing these really big thorough protocols when we have no idea what the uptake will be like.

And also, I'm wondering about the idea of can we consider when it might make sense to enroll a comparator group? Because I know for example, we want to be careful about temporality and when one group is studied versus another, but maybe especially from a physician or patient, does a woman participating in a registry as a comparator and if that registry never reaches a specific number to be able to do those comparisons, are we using those patients' time well? Are we using the physician's resources well? So I just wanted to think about the idea of not even between studies and having a sequential approach, but even within a study design, does it make sense for us to think about what an appropriate sequential approach might be? And that might end up that there's protocol version A and SAP version A, and then depending on what happens, maybe there's another iteration. So just wanted to bring that up for discussion.

Evan Myers:

Sonia.

Sonia Hernández-Díaz:

Yeah, I think that's what we were referring to with the dynamic adaptable framework that will have milestones and checks every year, every half a year to revisit even the statistical protocol as you said. No, I agree with you. You have five and you have to develop your propensity scores. So I totally agree with having that in place within the framework as being a live system with a priori rules, but that have that flexibility at certain points.

Janet R Hardy:

I think though that to begin with is there should be a robust full protocol, assuming that we can move forward and then hitting milestones where probably if it's not working, we're going to peel off and say, well, we can't do this. But to revise and add things to the protocol as you go along is ... I like the idea of starting with a full protocol and then saying, okay, this isn't going to happen, and as we reach our milestones.

That would be my preference. An interesting thought about, I appreciate your thoughts on whether we're using comparators time and HCPs time well. What came to mind for me was clinical trials. We have control groups in clinical trials, not everything works out and sometimes trials are stopped. Sometimes they simply, they're ineffective and everything stops. Do we ask the same question at that time of the comparators? Is there-

Elsie Grace:

[inaudible 01:17:18].

Janet R Hardy:

Yes, but there's-

Elsie Grace:

[inaudible 01:17:25] periods are presumably fairly short [inaudible 01:17:29].

Janet R Hardy:

They're still agreeing to participate in a study and provide their time and information at the beginning. I think at the beginning of the intent is hopeful and that it will move forward. I don't know, that's just a personal opinion on the fly.

Evan Myers:

I think we're going to need to stop because unfortunately this is a great discussion. I'd like to thank the panelists and the FDA representatives, and I know I'm sure Gerrit is going to, again, encourage everyone to submit questions to the docket and to continue the conversation. Thanks.

Gerrit Hamre:

Thank you. Before brief closing remarks and yes, thank you, Evan, one final reminder. Sincere thanks to the virtual participants. We're grateful where we've been able to incorporate your suggestions live. But know also that all submitted questions are forwarded to the FDA for their review and consideration.

They are sincerely grateful and this will contribute to consideration for framework revisions. Thanks also to all of our participants who contributed today. One thing that stood out throughout, folks that have contributed, you've devoted your careers to these issues. The wisdom that you bring is evident and it also highlights how incredibly challenging, how extraordinary some of these challenges are.

So thank you for your career efforts and please do continue them. Yesterday we heard from diverse stakeholders representing patients, researchers, providers, industry, and others on important considerations on the proposed pregnancy safety framework and also associated pharmacological challenges and decision-making during pregnancy generally.

Topics are highly complex, obviously requiring cross-sector collaboration to answer questions about the safety of pharmaceuticals during pregnancy. We also heard a lot about the need for multi-sourcing of data when possible. The FDA developed their framework with all of this in mind, an effort to optimize study, design data, study design and data, ultimately translating it into real world clinical practice.

Today we heard about upcoming demonstration projects and how they will inform the proposed framework as well as inform us in what context a registry study, database study, both, additional sources also should be required at time of approval. Themes on key factors contributing to the accuracy and timeliness of pregnancy safety data included exposure and outcome sensitivity and specificity, as well as gaps, the FDA is looking to fill via the demonstration projects.

Also, anticipation of gaps that will still remain. Additionally, capturing targeted outcomes and considering data beyond claims were discussed multiple times. Our last session of the day concluded with a robust discussion on the application of the framework and provided a range of recommendations for consideration in the future development of the framework presented in this public workshop. So many great inputs and dialogue involved from all in that session. Thank you so much for wrapping the day with such candor and thoughtfulness.

For Evan's reminder, we'd like to remind audience members, this is the first iteration of FDA's draft framework. Further work will occur via those demonstration projects just referenced leading to an updated framework. The community's inputs and insights are highly encouraged, necessary, and welcomed. Please do so by the end of November 30th if submitting to the public docket. You'll find that docket via the federal register notice of this public workshop.

Again, all of this material, slides, presentations, agenda, et cetera, will post in just a matter of days. Before we go, we'd like to take a moment to thank everyone involved in putting together this excellent event, first of all, and in a recurring fashion, our speakers and panelists, moderators, et cetera, very grateful. Some months of work that led into it, sometimes years.

Grateful for that too. I also like to thank specifically our colleagues on the FDA planning team, including Vicky Chan, Amy Ramanadham, Wei Hua, Leyla Sahin, Lynne Yao, and many, many more. Both your presentations today and all the work leading up to it, we really are quite grateful for it. Of course, I always like to do this last, the great team here at Duke, our moderators, Megan Clowse, Geeta Swamy, Evan Myers, Marianne Hamilton Lopez, and also our project staff, including Maryam Nafie, Dure Kim, Hannah Vitello, Luke Durocher, Nancy Allen LaPointe, and Kate Tsiandoulas. Thank you.

Finally, thank you to everybody who attended in person, echoing what so many other people said, delightful to see you. This is my first in person in almost four years. Probably done a couple dozen prior to that. Meant a lot to me. Saw a lot of old friends too. With that, have a good afternoon. Hope to catch you all again soon. Thank you all.