Zika Virus (ZIKV) Individual Participant Data (IPD) Consortium

---

Annual Meeting Report

26-27 January, 2021
Online

Table of Contents

Background

Despite the high levels of research investment in Zika virus (ZIKV) numerous outstanding questions remain related to the short-and-longer-term effects of fetal exposure during pregnancy. In February of 2017, ZIKV researchers from around the world came together to develop a path forward for leveraging existing participant-level data to better address these questions.

Following the 2019 Consortium meeting in Campinas, Brazil, the Consortium met on January 26 and 27, 2021 online via Zoom to update members on the progress.

The three sessions were organized as follows:

- Session I: Review Consortium progress;
- Session II: Review working groups' progress;
- Session III: Discuss the preliminary plans for the statistical analysis and harmonization pilot

DAY 1

Introduction

Day 1 began with welcome remarks by Drs. Nathalie Broutet of WHO and Consortium Co-Chair, Ricardo Ximenes. Dr. Broutet presented the agendas for Days 1 and 2, with a brief description of each session, and she stressed the need for discussions around the way forward and finalization of the IPD-MA timeline for 2021.

SESSION I: REVIEW AND UPDATES OF THE ZIKV IPD-MA CONSORTIUM

The goal of the ZIKV IPD-MA Consortium is to improve our understanding of the effect of ZIKV infection during pregnancy on neurocognitive development.

The Four Objectives of the IPD-MA are as follows:

1. Estimate the absolute and relative risks of fetal infection; miscarriage (<20 weeks), fetal loss (≥20 weeks gestation), microcephaly and other manifestations of CZS and later developmental delays.

2. Identify factors that modify women's risk of adverse ZIKV-related fetal, infant and child outcomes and infants' risk of infection.

3. Use information on the relative importance of different effect measure modifiers identified in Objective 2 to decompose the total effect of ZIKV infection during pregnancy on adverse fetal, infant and child outcomes into (1) the direct effect of ZIKV; (2) the indirect effect of ZIKV as mediated by the effect measure modifier of interest (e.g., DENV, CHIKV or STORCH pathogens) and (3) the effect of the interaction between ZIKV and the mediator of interest.

4. Develop and validate a risk prediction tool to identify pregnant women at a high risk of an adverse ZIKV-related outcome and to inform couples, healthcare providers and/or resource mobilization.

In the first session, members of the WHO project management team presented an overview of the goals and timeline of the IPD-MA since the Campinas meeting in December 2019, as well as an update on progress made since then. Funding for ZIKV IPD-MA activities from 2016-2021 have been provided by DFiD/Wellcome Trust, USAID, and WHO HRP Alliance. One area of focus throughout the meeting was the impact of COVID-19 on researchers, and how to continue to adjust IPD-MA project deadlines to reflect delays at the study level.

As of January 2021, 67 cohort sites, representing longitudinal data from 22 countries or territories on 30-40 thousand pregnant women and their children, will be participating in the IPD-MA. Fifty-seven of those studies have returned a signed LoA for data sharing and authorship, and 16 sites have used OpenClinica to upload their de-identified, participant-level data to the HRP e-archive system.

Five working groups have been meeting since March 2020 or earlier, including the Social Sciences Working Group (SSWG), the Outcomes Working Group (OWG), the Exposures Working Group (EWG), the Harmonization Working Group (HWG), and the Statistical Analysis Working Group (SWG). Substantial progress has been made since Campinas on project and data management objectives, the collection of metadata and the harmonization and pilot curation of metadata, and on a multi-site qualitative study. Further, the past year saw steps taken towards the development of the statistical analysis, as well as the dissemination of two commentaries and related publications from the SWG.

In the area of project and data management, 2020 saw the **WHO team** work on:

- Study recruitment (ongoing since 2016)
- Development and refinement of the data upload/management system in order to support data behind the WHO firewall
- Collection of required study documentation
- Establishment of communications with investigators about outliers, differences between shared datasets, and published data

The **Harmonization Working Group:**

- Developed and refined a harmonization system and harmonization SOPs
- Completed a variable-by-variable alignment of data dictionaries with the 20-20-20 key variable list
- Developed template harmonization code tailored for each of 16 datasets in Stata and R
- Initiated study-by-study data checks for the pilot harmonization
- Developed the master codebook.

Additionally, the metadata survey was developed, disseminated, and analyzed, metadata tables were created, and work is ongoing on a metadata publication.

The **Social Science Working Group**:

- Updated their qualitative methods in order to adapt to COVID-19
- Conducted in-depth interviews with 3 different groups of women at 7 sites in 3 countries Developed a codebook
- Trained on qualitative analysis software

Finally, the **Statistical Analysis Working Group**, which began meeting in October 2020, developed novel statistical approaches to addressing measurement error in the context of IPD-MA and began working with the pilot data.

Because of COVID-19 response, some changes were made to the IPD-MA timeline, as well as a brief review of participation status.

Discussion and Q&A, Session I

Session I included an interactive session facilitated by Lauren Maxwell, in which questions were posted at www.menti.com in order to allow participants to give immediate feedback.

Question 1: Why does the initiative continue to be important to you?

Responses included the desire to pool data across studies to better understand risk in a unique cohort, to understand differences in findings by setting, to fill a gap in qualitative literature in Zika, and to continue the global effort to fight the virus.

Question 2: What would you like to see from the ZIKV IPD-MA in 2021?
Most respondents stressed their desire to continue to work collaboratively on this effort, and to begin to see some preliminary results and progress towards publications.

Question 3: Beyond our 4 objectives for the IPD-MA, what questions would you hope to answer with the harmonized dataset?
Responses to this question were limited to the 34 studies that agreed to contribute data beyond the scope of the 4 objectives, and focused primarily on the desire to learn more about long-term adverse outcomes and understand the full spectrum of congenital Zika syndrome (CZS). Additionally, other respondents stressed the benefits to understanding the international collaborative process, particularly within the context of a global pandemic.

Drs. Maxwell and Ximenes facilitated a discussion of the following topics of interest to Consortium members:

- Contributions expected from different researchers and promptness of responses required to adhere to current timeline, and more discussion of the timeline itself
- Flexibility and balance required to collect data from researchers who are on the frontlines of the ongoing Covid-19 pandemic
- Support needs from investigators to ensure a successful harmonization process
- Challenges of incomplete datasets
- Expectations of consistent meetings between different working groups and statistical analysis team, particularly monthly meetings between SWG, OWG, and EWG
- A goal for 2021 is having a dataset ready to build production models and start publishing

SESSION II: UPDATES FROM THE WORKING GROUPS

Outcomes Working Group

Presented by Sarah Mulkey, George Washington University/Children's National Hospital

This session summarized the past year's work from the Outcomes Working Group. The group discussion from the Campinas meeting was recapped. At the Campinas meeting, the group defined the TOR activities and deliverables, which included agreeing upon outcome variables, particularly for the initial variable mapping and harmonization, and reviewing the outcomes portions of the metadata survey and the master codebook. Upcoming activities will include providing expert opinion to refine the statistical analysis plan and to review preliminary statistical findings and provide feedback.
Nine group meetings were held between March 2020 and September 2020 in which the groups goals were accomplished as follows:

March 26, 2020 – May 12, 2020:

- Group examined metadata survey
- Discussions on challenges for mapping variables from all studies to one common codebook (master codebook)
- Discussions on heterogeneity, variable measurement and definitions
- Considered more subtle clinical findings
- Discussed use of different sets of criteria, example of prenatal/postnatal microcephaly in context of Zika
- Reviewed and finalized pregnant woman and outcomes women tab in 20-20-20 key variable list

July 21, 2020 – September 15, 2020

- How much variability between studies as far variable measurement? Importance of capturing different levels of specificity
- Moving from short key variable list to the longer variable list that will be the IPD master codebook
- Discussed metadata table shell and upcoming metadata publication
- Reviewed fetal tab in master codebook as a group
- Group asked to review infant/child variables in master codebook

Exposures Working Group

Presented by Mabel Carabali, McGill University

This session described the activities of the Exposures Working Group from April 2020 to August 2020, which included a pause between June and July. The working group had 7 meeting/calls with 4 to 10 participants and had ongoing interactions via email. While most deliverables were successfully completed in 2020, the final analysis plans and metadata manuscript preparation are activities ongoing into 2021.

The Terms of Reference for the Exposures Working Group included the following activities:

- Identification of study-level variables that most closely correspond to 20 key exposure/infection variables
- Documenting and implementing harmonization decisions based on working group feedback
- Developing, finalizing, and disseminating the metadata survey
- Drafting, refining, and finalizing the master codebook by selecting variables and definitions

With the following expected deliverables:

- List of 20 key exposure/infection variables (became ~60vars)
- Key composite definitions (e.g., ZIKV+, ZIKV-, ZIKV indeterminate, etc.)
- Final exposure/infection section of metadata survey
- Final exposure/infection section of master codebook
- Final analysis plan (in collaboration with statistical analysis and harmonization teams)

The expected deliverables for 2021 and beyond are to review and finalize metadata publication and tables in February 2021; from February 2021 on, close collaborations with SWG to interpret metadata and harmonized participant-level data from pilot harmonization; from July 2021 on, expert guidance for final harmonization and analysis phase.

Social Science Working Group

Presented by Edna Acosta Pérez, University of Puerto Rico

This session began with an overview of the ZIKV IPD-MA qualitative sub-study, which has used in-depth interviews to learn more about women who were pregnant during the ZIKV epidemic, regardless of whether they had an adverse ZIKV-related outcome. The women, who were primarily located in Brazil and Puerto Rico, shared their values, preferences, perceptions, and opinions about how they learned about their ZIKV status and their possibility of an adverse pregnancy, as well as their perceptions about the IPD-MA itself. The study underwent modification due to the COVID-19 pandemic, including changing from focus groups to individual interviews and excluding currently pregnant women. The objectives of the study-which were also expanded due to the COVID-19 pandemic-are to understand:

1. How women want to be informed of the probability of adverse ZIKV-related fetal, infant, or child outcomes
2. Community perceptions of acceptability of using de-identified, participant level data for the ZIKV IPD-MA
3. Challenges associated with having an infant or child who has experienced a ZIKV-related adverse outcome
4. How women feel about the use of de-identified, participant level data from surveillance systems and cohort studies for meta-analyses for risk prediction
5. Use of WhatsApp groups for risk communication duringZIKV
6. The dual burden of ZIKV and COVID-19

To date, 36 interviews have been conducted, 25 have been transcribed, and coding is in process. The SSWG has been meeting weekly-and sometimes twice a week-since April 2020 and have been using this time to collaborate on codebook development and to learn the analytic software.

Some of the preliminary findings are as follows:

- Women have a good general knowledge of ZIKA virus prevention and transmission, although almost no women mention sexual act as a possible source of transmission
- In general, perceptions of risk are related to communications with their healthcare providers
- Women understand the importance of collecting de-identified participant-level data for performing analyses
- Women explained their preferences for how to receive their ZIKV screening/ultrasound results

The timeline, despite some challenges, is for the study to be completed by the end of 2021.

Harmonization Working Group and Metadata

Presented by Lauren Maxwell, WHO; Mabel Carabali, McGill University, Brooke Levis, Keele University

This session began with an overview of best harmonization practices as outlined by Maelstrom Research Group, which conducted a training with the HWG in Fall 2020, and which has guided the work of the HWG since they began meeting in April 2020. As recommended by Maelstrom, the HWG started with a clear protocol and then moved to study selection, they are working to define the master variable list in two steps, with the first being a pilot harmonization to identify challenges needing to be addressed before the final harmonization, they are clearly documenting the harmonization process in order to retrace and change steps if needed, and they will disseminate harmonization products as well as metadata products.

As mentioned in the presentations by the OWG and EWG, harmonization started with the 20-20- 20 variable list in order to define 60 key variables, then that list was expanded to 235 key variables with agreement with the working groups. This will be expanded further to become the master codebook, which will have 1372 variables as part of the main harmonization. Over the past year, the harmonization group has been aligning the study-level variables in the pilot datasets to the key variables, and then coding the variable alignments and using that code as template harmonization code.

The most important lesson the HWG has learned through their activities is that harmonization is a very complex process, especially when dealing with an emerging pathogen such as Zika. There have been several challenges confronted by the group:

- Working through the process in four languages
- Challenges with mapping because not all codebooks reflected what was in the actual datasets
- Codebooks were that were updated after initial harmonization was done, necessitating multiple rounds of harmonization
- Missingness, which was partially resolved by setting variables using metadata information
- Very time-consuming data cleaning

The metadata survey was distributed to all sites and is necessary to gather details from investigators on in-study heterogeneity found in study characteristics, variables measured, and standards used for classification. The survey contains more than 2000 variables, which will ultimately be reduced to 50-60 used for analyses. Thus far, 52 sites have at least partially completed their surveys, and the harmonization team strongly encourages all sites to complete their surveys. The presenter for this section, Brooke Levis, also described the process by which investigators can click through to the metadata spreadsheets using links provided in the presentation, and then use the comment section to add missing information or variables that were not included in earlier iterations of the survey. The spreadsheet contains a tab for each table and a row for each study.
She stressed the importance of providing the most up-to-date information as possible in order to ensure completeness of the data.

Summary of main metadata characteristics:

- Most data contributors are from the Americas, most participants came from hospitals/healthcare centers, most were pregnant, HIV status of study participants are about half positive and half negative, there is wide variation in how enrollment status relates to ZIKV, DENV & CHIKV testing status
- Tables are available for ascertainment of ZIKV status, immunoassays and PRNT, developmental screening/assessment tools (most not measured by studies), ascertainment of microcephaly and CZS (how studies categorized it or their standard for use varied widely), demographic/socioeconomic/workplace/behavioral factors (whether or not this information was collected varied widely by category), pregnancy outcomes, and maternal STORCH pathogen exposure

Discussion and Q&A, Session II

For the final discussion of Day 1, Ricardo Ximenes asked for input from each of the working groups about the type of activities they would like to see going forward. As harmonization will be ongoing and there will be substantial interaction between harmonization and the other working groups, how will the working groups best interact with these processes and each other during analysis,

interpretation, and publication, and what kind of support is needed?

Sarah Mulkey from the Outcomes Working Group:

- More overlap with other working groups, especially between the OWG and the EWG
- Subgroups within the larger working group, for example prenatal and postnatal outcomes groups
- Smaller groups would encourage participation and the ability to collaborate more easily between meetings, as they would allow for more scheduling flexibility

Mabel Carabali from the Exposures Working Group:

- Both the OWG and the EWG should work more closely with the Statistical Analysis Working Group, which has biostats expertise but needs the contributions of experts in subject matter areas
- Agrees that subgroups would be helpful

Edna Acosta Perez from the Social Science Working Group:

- Highlighted their outputs and path forward on the qualitative study, as the work of the SSWG is parallel but independent of the other working groups.

The discussion continued with a comment from Thomas Jaenisch, a Consortium Co-chair, who was interested in input from PIs regarding timelines. Thomas offered the example of ZIKAlliance, which is running behind schedule because of the need to reconduct laboratory testing and he suggested that a questionnaire could be sent to all groups regarding their timelines for completion with data, in case other sites are experiencing delays. Additionally, he shared that his group is looking at different types of tests (such as seroconversion), and that it would be interesting if groups could compare different testing procedures used, since many have not provided data on lab testing, and that perhaps a repository could be created for this information. This type of question is also related to the metadata, which needs more participation after undergoing a lot of refinement with inputs from the working groups. For example, diagnostic and quality control sections are important to determine type of infection because of the overlapping Dengue and Zika virus epidemics.

One question posed to the group is whether test data should be submitted. Initially test data was submitted, with the idea that it would inform harmonization, but so many changes occurred before submission of full data that the entire harmonization process was having to be repeated. The option of submitting test data might work for sites that can't share data by June of this year, because harmonization is so time consuming. This discussion circles back to the topic of changing timelines and how the project can accommodate groups that will not have their data ready by deadlines, and to the idea of a questionnaire to get a better sense of individual study timelines.

As a final question for the session, Nathalie Broutet asked how the review and checks of variable name translation occurs since numerous languages are used? Mabel Carabali shared the process in more detail, since she speaks multiple languages and worked on the 20-20-20 list with the working groups, as well as the master codebook which was based on 8 initial codebooks. Because she is an infectious disease physician and epidemiologist, she was able to translate and work with the medical terminology, and investigators and data managers themselves have provided clarity and have interacted extensively with harmonization team. She stressed that the group has been recording their methodologies and processes, including documenting how consensus was reached, for the benefit of future publications.

Day 2

Introduction and Review of Topics from Day 1

Ricardo Ximenes offered some opening remarks, in which he stressed both the impressive level of work that has been done thus far, as well as the challenges and amount of work that lies ahead.

Caron Kim followed these remarks with a recap of the discussions that were held on Day 1, including the overall progress of the project to this point, as well as some more specific details about working group progress, a summary of discussion topics, and the overall IPD-MA timeline.

INTRODUCTORY SESSION: REVIEW OF PRINCIPLES ON GOVERNANCE AND DATA SHARING

Presented by Rob Terry, WHO

This session discussed sharing health data for improved impact and good practice in governance arrangements to share individual participant data. The Ebola Data Platform was used as an example of how a data sharing platform might be successfully implemented. When sharing IPD without standing consent, emerging good practice is defined as follows:

- Create a comprehensive governance package (Purpose, Ethics, System, Technical, Legal, Economics
- Sharing IPD needs managed access, not free and open access
- Engage ERC/IRBs and stakeholders early in the process
- Develop a clear research agenda
- Develop a generic research protocol for ethics approval
- Data deposit agreement
- Data access agreement
- Credible data access committee
- Current good practice examples – open access to all agreements on IDDO websites

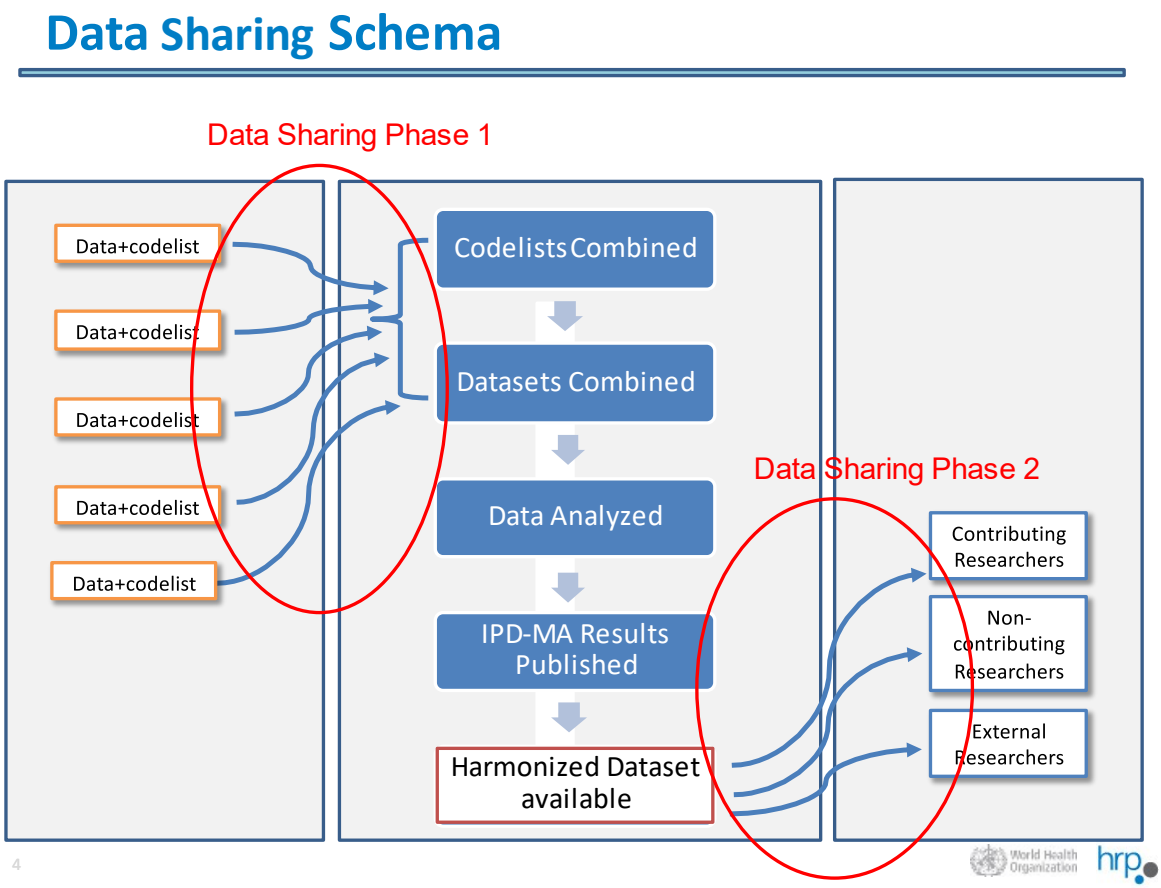
A transparent governance framework was discussed using the EDP example. The technical and managed components of the framework were defined, with the key component being a means to define uses of data, as well as have a managed and collaborative method to make IPD available to researchers other than initial PIs. The session concluded with a Q & A session with the following topics addressed:

- Capacity building: Any question being considered must give a plan as to how they will engage with researchers from the affected countries. Those researchers must be given the opportunity to either refuse, join that research team, or act as a technical advisor
- Discussed the implications of having a unique cohort of pregnant women and how the data could be used to answer other questions. To address concerns about sharing this data, boundaries around data sharing is one suggested solution by way of declaring the types of questions the data is thought to be most appropriately used for. The DAC would start by determining if a proposed question meets the agreed upon criteria
- It was reiterated that data sharing is voluntary, and protocols only apply to studies that agreed to share data
- Discussed the advantages of contracting with an existing supplier as opposed to building a platform from scratch
- PI participation in Steering Committee was discussed. Current protocol is to only contact the PI if they request to be contacted, to avoid a delay in the process
- The key roles and differences of each committee within the governance framework were discussed

SESSION III: STATS DISCUSSIONS

Overview of Analysis of Pilot Dataset

Presented by Soe Soe Thwin, WHO



This session began with a review of the data sharing schematic introduced at Campinas a year ago. This established the covenant structure and the data sharing agreement and determined whether to share analytics and to which groups. An overview of the IPD-MA process was given, with a skeletal structure of the process as follows:

- Agreements and approvals in place
- Develop Data sharing platform
- Datasets and data dictionary are shared
- Data dictionaries are combined
- Datasets are combined
- Analysis dataset extracted and analyzed
- Publication of IPD-MA results

The current study status was given as of January 26, 2021. To date, of 67 total cohorts, 16 cohorts have signed LoA, shared Codebook, and shared full dataset. A total of 57 cohorts have signed an LoA, and have not yet shared their Codebook and/or full dataset. Of the 10 remaining cohorts who have not signed an LoA, 8 have not shared Codebook or dataset. The initial cut-off for inclusion of cohorts in the IPD-MA was June 2020, at that time total cohorts were 33. The initial cut off for baseline data upload for inclusion in the IPD-MA was end of 2020. The baseline data is through delivery and post 6 weeks of delivery. The updated cut off for baseline data is end of Q1 due to COVID-19. The tasks identified at Campinas were shared. Two noted updates were increase from 190 variables from 10 datasets to 237 variables from 16 datasets for the preliminary report after analysis at the end of 2020, and the Consortium meeting to present analytic result, originally

planned for April 2021, will be pushed out due to updates of the timeline. There was a review of the electronic data platforms, which include the HRP e-ARCHIVE system with a single sign-on, OpenClinica, the department administered web-based platform to upload datasets and codebooks, HRP Dropbox with administrator access for documents and code lists, and HeiBOX (a GDPR compliant cloud sharing folder system with two factor authentication, protected by the Heidelberg University firewall) with administrator access for analytic datasets.

An overview of the IPD-MA analytic process through 2022 was presented, with the key focus for 2021 being the following:

- Statistics Plan Finalized March 2021
- Datasets combined June 2021
- Descriptive Analysis September 2021
- IPD-MA Analysis December 2021
- Share Follow up data ongoing 2021 through 2022
- IPD-MA Analysis with Follow-Up data ongoing July 2021 through 2022

A comprehensive list of the total 67 studies/cohorts and data sharing status was shared.

Descriptive Statistics

Presented by Anneke Damen, Utrecht University

This session began with a recap of analytic objectives:

1. **Estimate the absolute and relative risks**, for set of outcomes, of fetal infection; miscarriage (<20 weeks gestation), fetal loss (≥20 weeks gestation), microcephaly, and other manifestations of CZS and later developmental delays for women who do and do not experience ZIKV infection during pregnancy
2. **Identify factors that modify women's risk** of adverse ZIKV-related fetal, infant, and child outcomes and infants' risk of infection (e.g., gestational age at time of infection, clinical or subclinical illness, concurrent or prior arbovirus exposure, other congenital infections, and posited effect measure modifiers).
3. Use information on the relative importance of different effect measure modifiers identified in Objective 2 to **decompose the total effect of ZIKV infection** during pregnancy on adverse fetal, infant, and child outcomes into 1) the direct effect of ZIKV; 2) the indirect effect of ZIKV as mediated by the effect measure modifier of interest (e.g., DENV, CHIKV, or STORCH pathogens); and 3) the effect of the interaction between ZIKV and the mediator of interest.
4. **Develop and validate a risk prediction tool** to identify pregnant women at a high risk of an adverse ZIKV-related outcome and to inform couples planning a pregnancy, healthcare providers, and/or resource mobilization (e.g., vector control strategies; antenatal care; open access to contraception).

Variables of interest, both exposures and outcomes, were discussed. Primary and secondary outcomes were outlined; it was noted that at least 30 outcomes are included. Metadata was presented. There were 15 studies in the dataset displayed in this presentation, although to date 16 studies are in the entire pilot dataset.

Pilot dataset descriptives

- 15 studies in the pilot dataset presented here (there are 16 studies in total, one is not included in these descriptives) – she presented list of cohorts and number of records for each

Study design

- 12 cohort studies
- 2 Active surveillance systems
- 1 study had no meta-data available

Source of data (some studies have a combination of sources)

- 8 recruited from hospitals or healthcare centers
- 6 community-based
- 3 recruited travelers

Population

- 11 included pregnant women
- 1 women of reproductive age
- 2 other women
- 2 studies only enrolled ZIKV PCR pos (+) pregnant women
- 1 study enrolled women with ZIKV PCR pos (+) or ZIKV probable infants
- 5 studies had other criteria
- 6 studies are missing

Methods for determining ZIKV status

- 5 used a clinical case definition
- 13 used molecular diagnosis (e.g. RT-PCR)
- 9 used immunoassay
- 2 used rapid diagnostic test

The steps for obtaining objectives were also outlined and discussed:

Step 1: Data Exploration

- Histograms of continuous variables to check outliers
- Frequency tables of categorical and dichotomous variables
- Explore sporadic missing values and systematic missing values

Samples of the data organized by variables of interest were presented. There were indicators of systematic missing values. Some of the pilot datasets require more data cleaning, checking, and recording.

Step 2: Multiple Imputation

- Impute simultaneously for all objectives
- Multilevel multiple imputation – account for clustering in data within different studies
- Include all relevant exposures, outcomes, confounders, effect modifiers
- Studies with systematically missings will be excluded for the specific outcome

Step 3: Calculate absolute risks of outcomes and relative risks for pregnancies with and without ZIKA exposure (currently in early stages)

- All primary outcomes and secondary outcomes with sufficient information
- Absolute risks separately for both exposures (maternal and fetal ZIKV infection)
- Relative risks of pregnancies with ZIKV divided by without ZIKV exposure
- Forest plots of study-level estimates of absolute and relative risks
- Combine with two-stage meta-analysis

The session concluded with an opportunity for participants to ask questions. The focus of the questions and resulting discussion was on the missing values in the datasets. Some participants sent direct messages about studies that were reported during the presentation as having missing data that should not have had missing data. In response, it was reiterated that studies have already been queried and the Harmonization Work Group will continue reaching out to studies to ensure accuracy. Some revised codebooks and datasets have already been received that may account for data that appears to be missing.

Objectives and Analysis Plan

Presented by Harlan Campbell, University of British Columbia

Two different statistical analysis plans will be followed:

- The primary plan, which will take care of primary analysis and looking at the outcomes of interest
- The secondary plan for quantifying measurement error, which is a sensitivity analysis for the data, to see what results look like if corrected for measurement error

The motivation for the primary analysis plan is 2-fold, to understand overall risks and also to identify specific risk factors for specific adverse outcomes. Despite all of the studies that have looked at this before, the estimates of the risk for specific outcomes has not been determined. It would be possible to get the simplest estimate by looking at the number of a certain adverse outcome divided by the number of pregnant women with ZIKV. For absolute risk, however, there are so many cohorts and each cohort has systematic differences, so the team will estimate cohort-specific probabilities and then combine through a 2-stage meta-analysis. Estimates of increased risk will be obtained by fitting a logistic model and getting cohort or study level estimates in a 2-stage meta-analysis. It can be difficult to get an estimate for increased risk if there is too much variation or a lot of missingness, but IPD allows all data to be combined into a 1-stage meta-analysis and allows for the incorporation of confounding information into the meta-analysis, which gives improved estimates for the increased risk. The analysis plan includes a whole table of positive confounders, how much can be included depends on the completeness of the data. Objective 2 of the IPD-MA is identifying factors that modify a woman's risk of adverse outcomes, so interaction effects will also be included to look at how they might impact the relationship between ZIKV and adverse outcomes.

The secondary analysis plan focuses on measurement error and is motivated by the concern that there is not just random error, but many sources of systematic error. The IPD-MA protocol lists five factors that could contribute to systematic error: selection bias, confounding bias, missing data, publication bias, and measurement bias. The discussion today focused on measurement bias, which is bias caused by measurement error (continuous variable) or misclassification (binary or categorical variable). This is frequently considered a study limitation in epidemiology, and it is very rare that studies correct for it, but the goal is to do so in the ZIKV IPD-MA, and to go beyond a sensitivity analysis for measurement bias. By using expert opinions from each study and the working groups to inform parameters, it will be possible to use Bayesian models to adjust for measurement error and give a better understanding of uncertainty. This will be done with all the different outcomes of the exposures of interest at the subject level and cohort level, as well as confounders and possibly effect modifiers.

Questions concerned the plan to deal with missingness of confounders and effect modifiers, and how imputation might be used if there is not much differentiation in some variables, such as SES in countries where Zika is only found where there is poverty.

Dr. Campbell explained that for the first question regarding cohort-level missing variables, the analysis will include random effects, for example SES missing for one study, which would impact the estimate. This problem would be taken care of by cohort-specific random effects.
In terms of discriminatory ability, there would be no concern about bias from confounding, since those disparities wouldn't exist in the sample.

A final question asked whether the dataset could account for all the types of serological assays, and how nuance in the data could be captured if this is not included.

In this case, the issue will be helped a lot by the metadata tables, because the EWG helped inform what was included in the tables so that they would reflect measurement error in the exposure. It would also be possible to have one analysis that restricts itself to data from studies that have good ascertainment of a particular exposure and apply those models to others. The master codebook did try to include more variables that might address definition of exposure.

Missing Data and Heterogeneity

Presented by Thomas Debray, Utrecht University

This discussion began with an overview of the plan to deal with missing data and heterogeneity.

There are two types of missing data: sporadic missingness, which occurs within studies with individual values, and systematic missingness, which happens in IPD when variables are not present for all participants. In order to deal with this, first the investigators should verify if missing data are truly missing and, if so, what reason there is for the missingness. For the ZIKV IPD-MA, the harmonization team has tried to recover some of these missing values in this way. Occasionally cohort studies employ eligibility criteria up front, so they do not measure a particular variable – it looks like it is missing but it is just that all participants are positive for the same exposure, so a determination can be made that it is not actually missing. Additionally, sometimes missing variables can be calculated from other variables that are present. In these ways, the harmonization process can help deal with missingness, and even some measurement error.

After harmonization, the statistical analysis team can start working on the analysis plan and the imputation, and it is important to understand why data is missing and distinguish between different causes:

- Is the reason for missingness the same for all participants? In this case, it is dependent on the individuals themselves, which is the easiest problem to resolve, but also the least common
- A more common scenario is when the missing data is related to a characteristic of the participant that was not observed, so the data is dependent on observed information
- Sometimes the missing data is dependent on unobserved information or the missing data itself, for example the differences in how people report their exposure depending on the attitude towards exposure
- It is not ideal to simply discard participants with missing data, this leads to data reduction and loss of statistical power, and also introduces bias because often people with missing values differ from people with complete data.

The IPD-MA will use multiple imputations, borrowing strength from variables related to the missing variables, which leads to less bias and more appropriate standard errors. This requires that the analysis be performed multiple times to pool corresponding results. This will be a joint modelling imputation, which assumes that all predictor and outcome variables describe a multivariate joint distribution. The distribution will then be characterized by estimating the mean and covariance, despite any incomplete data, and individual observations will be used to create tailored distributions. In epidemiology research distributions are often skewed, so adjustments will be made to account for complexities. There are more complexities in an IPD-MA because some studies are small that the computations cannot be made in the same way, so we must borrow information across studies and combine that information while performing imputations and accounting for heterogeneity between studies. With the multilevel joint modelling, each study has own distribution, and each study's means and covariances are different, but related. An important note is that heterogeneity must be accounted for in these imputations and enough studies must be included. Additionally, the imputation model should account for individual and study level covariates and use metadata to define the structure of the model. Ultimately, a plan will be constructed with pilot data as an example of this imputation, then the plan will be refined and extended as more datasets become available.

CLOSING SESSION: THE PATH FORWARD

Discussion and Q&A on Revised Timeline and Goals for 2021

Nathalie Broutet led the final discussions for this year's meeting. She started by observing that while the online format for meetings was not ideal, it did have some practical benefits because it allows a larger group to meet. Additionally, it makes it possible to meet more frequently, and she suggested another meeting in July of this year, which met with approval from the group.

Much of the remaining discussion addressed timeline changes made necessary by the COVID-19 pandemic. By consensus, the decision was made to extend the cut-off date for sharing data up to 6 weeks postpartum from end of March 2021 to end of June 2021. Numerous participants responded positively to this idea, and none expressed any concerns, so the new deadline was adopted.

For working groups outputs, it was suggested that working groups will continue to work on the harmonization process through the end of 2021, rather than the June. Both the OWG and the EWG agreed to extend their work further, and acknowledged the importance of their groups to the harmonization process. Additionally, they confirmed an interest in interacting more with each other and with the SWG, perhaps by joining in the SWG meetings, which are ongoing.

The statistical analysis plan will stay at finalization Q1 but getting the full analytic dataset will have to expand beyond Q2.

Finally, it was agreed that a questionnaire will be sent out to get updated information on dates when studies can submit data.

The meeting concluded with an acknowledgement of the amount of work that has gone into this very complex process, and with thanks to all for their contributions under very challenging circumstances in the past year.

Acknowledgements

Annexes

AGENDA

Consortium Co-chair and Meeting Chair: Ricardo Ximenes
Consortium Co-chair: Thomas Jaenisch

Meeting Objectives:

- Review prior agreements and objectives, including timeline
- Review progress of the different working groups over last year since the meeting in Campinas
- Based on the results of the pilot of the IPD-MA, set objectives and timeline for thefinal phase

DAY 1: 26 January 2021		
Time CET	**Agenda Item**	***Facilitators*/Presenters**
3:00-3:15	Welcome and Introductions	Nathalie Broutet & Ricardo Ximenes
3:15-4:00	**SESSION I: REVIEW AND UPDATES OF THE ZIKV IPD-MA CONSORTIUM**	
3:15-3:30	Review of goals/objectives/timeline	Caron Kim
3:30-3:50	Update on Consortium Progress since Campinas	Lauren Maxwell
3:50-4:00	Q&A/Discussion	*Lauren Maxwell*
4:00-5:30	**SESSION II: UPDATES FROM WORKING GROUPS**	
4:00-4:15	Outcomes Working Group	Sarah Mulkey
4:15-4:30	Exposures Working Group	Mabel Carabali
4:30-4:45	Social Sciences Working Group	Edna Acosta Pérez
4:45-5:00	Metadata/Harmonization Working Group	Brooke Levis, Lauren Maxwell & Mabel Carabali
5:00-6:00	Discussion: Lessons learned from assembling a pilot dataset, and the way forward for working groups	*Ricardo Ximenes*

DAY 2: 27 January 2021		
Time	**Agenda Item**	***Facilitators*/Presenters**
3:00-3:15	Review of Topics from Day 1	Caron Kim
3:15-3:45	Review of Principles on Governance and Data Sharing	Rob Terry
3:45-6:00	**Session III: STATS DISCUSSIONS**	
3:45-4:00	Overview of Analysis of Pilot Dataset	Soe Soe Thwin & Lauren Maxwell
4:00-4:30	Descriptive Statistics	Anneke Damen
4:30-5:00	Objectives and Analysis Plan	Harlan Campbell
5:00-5:30	Missing data and Heterogeneity	Thomas Debray
5:30-6:00	Path Forward: Revised Timeline & Goals for 2021/ Discussion	Ricardo Ximenes, Caron Kim, Soe Soe Thwin, Nathalie Broutet, & Lauren Maxwell

LIST OF PARTICIPANTS

Country	Cohort/PI	Name of Participant
Brazil	Brasil	Patricia Brasil
	Costa	Federico Costa
		Mitermayer Reis
		Pablo Aguilar
	Cunha/Prata	Antonio Alves da Cunha
	Duarte Passos	Ana Bertozzi
	Hofer	Cristina Hofer
	Joao	Maria Teixeira
	Ko	Albert Ko
		Cynthia Rodamilans
		Deolinda Scalabrin
	Marques	Ernesto Marques
	Martinez Espinosa	Flor Martinez Espinosa
	Moreira	Elisabeth Moreira
		Luis Galvao
	Mussi	Marisa Mussi
		Conrado Milani Coutinho
	Nogueira	Mauricio Lacerda Nogueira
	Segurado/Silva	Vivian Avelino da Silva
	Silva	Antonio Moura da Silva
	Siqueira	Isadora Siqueira
	Turchi	Marilia Turchi
		Luiza Emylce Pela Rosado Schmaltz
	Ximenes	Ricardo Ximenes
		Celina Turchi Martelli
		Thalia Velho Barreto
		Democrito Miranda Filho
Colombia	Arias	Stacey Schultz-Cherry
		Dayana Pino
	Becerra	Carlos Becerra
	Gilboa	Suzanne Gilboa
	Lopez-Medina	Eduardo Lopez-Medina
		Isabel Hurtado
	Mercado Reyes	Carlos Marin Correa
	Mulkey/Debiasi (also USA)	Sarah Mulkey
		Roberta Debiasi
	Sanz Cortes	Magdalena Sanz Cortes
	Villar	Luis Villar
		Maria Consuelo Miranda Montoya
Ecuador	Soria	Carmen Soria
		Mary Regato Arrata
French Guiana	Cabie	Rebecca Grant
		Anna Funk
	Pomar	Leo Pomar
Grenada	LaBeaud	Desiree LaBeaud
		Amrik Kang
		Calum Macpherson
Guatemala	Figueroa	Lester Figueroa
Honduras	Alger	Jackeline Alger
		Pierre Buekens
		Maria Luisa Cafferata
Kenya	Widdowson	Peninah Munyua
Mexico	Alpuche	Celia Alpuche
	Borja	Victor Borja
		Cesar Gonzalez
		Concepcion Grajales
		Jose Esteban Munoz Medina
Peru	Gotuzzo	Eduardo Gotuzzo
		Rodrigo Cachay
	Ochoa	Cesar Ugarte-Gil
Singapore	Chan	Derrick Chan
		Yee-Sin Leo
		Tun Linn Thein
Spain	Bardaji	Azucena Bardaji
		Elena Marban
	Soriano-Arandes	Toni Soriano-Arandes

Suriname	Juliana	Natanael Holband
Trinidad and Tobago	Sohan	Karen Sohan
USA & Territories	Crawford (Virginia)	Katherine Crawford
	DeWilde (US Virgin Islands)	Leah DeWilde
	Garcia-Saavedra (New Mexico)	Luigi Garcia-Saavedra
	Hinojosa (Texas)	Steven Hinojosa
	Piccardi (Maryland)	Monika Piccardi
	Rettler (Utah)	Hannah Rettler
	Valencia Prado (Puerto Rico)	Miguel Valencia Prado
		Camille Delgado Lopez
Venezuela	Tami	Adriana Tami

Organization or Group	Name of Participant
Consortium Co-chair and ZikAlliance	Thomas Jaenisch
ZikAlliance	Sarah Bethencourt
PAHO	Freddy Perez
	Bremen DeMucio
WHO	Nathalie Broutet
	Vanessa Cavallera
	Myriam Cortes
	Edna Kara
	Caron Kim
	Ornella Lincetto
	Katherine Littler
	Lauren Maxwell
	Ingrid Rabe
	Diana Rojas Alvarez
	Janet Sayers
	Ute Stroher
	Rob Terry
	Soe Soe Thwin
	Maria Van Kerkhove
	Annelies Wilder-Smith
NIH	Walla Dempsey
	John Moye
MERG	Elizabeth Brickley
CDC	Cindy Moore
	Van Tong
Exposures, Outcomes, Harmonization, and Analysis groups	Mabel Carabali
	Brooke Levis
Exposures Working Group	Ana Gorini de Veiga
Harmonization Working Group	Luz Leegstra
	Helen Cerigo
Outcomes Working Group	Guilherme Calvet
Social Science Working Group	Edna Acosta Perez
	Marcela Daza
	Pamela Ximena Diaz Bermudez
	Rosamund Greiner
	Victor Herrera
	Olivia Manders
	Gustavo Matta
	Ester Paiva
Statistical Analysis Working Group	Andrea Benedetti
	Isaac Caicedo-Castro
	Harlan Campbell
	Anneke Damen
	Thomas Debray
	Paul Gustafson
	Ulisses Montarroyos
	Wayner Souza
	Yinghui Wei
Pasteur Institute	Arnaud Fontanet
University of Sao Paulo	Joao Paulo Souza
Institute of Social and Preventive Medicine, University of Bern	Nicola Low
Instituto Nacional de Salud - Colombia	Maritza Gonzales Duarte
Centre Hospitalier Universitaire de la Reunion	Patrick Gerardin
Universidad del Valle de Guatemala	Celia Cordon-Rosales
London School of Hygiene and Tropical Medicine	Phillipe Mayaud
UMC Utrecht	Carl Moons
University of Colorado	Margo Harrison
	Jamie Westcott