Routledge
Taylor & Francis Group

Applied Developmental Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hads20

Conducting virtual assessments in developmental research: COVID-19 restrictions as a case example

G. Anne Bogat, Kristyn Wong, Maria Muzik, Joseph S. Lonstein, Amy K. Nuttall, Alytia A. Levendosky, Cara F. Colao, Alanah Hall, Kara Cochran, Kathryn R. Forche, Allison Koneczny, Amanda Gareffa, Olivia Oates, Stephanie Robinson, Alexandra Ballinger & Sara F. Stein

To cite this article: G. Anne Bogat, Kristyn Wong, Maria Muzik, Joseph S. Lonstein, Amy K. Nuttall, Alytia A. Levendosky, Cara F. Colao, Alanah Hall, Kara Cochran, Kathryn R. Forche, Allison Koneczny, Amanda Gareffa, Olivia Oates, Stephanie Robinson, Alexandra Ballinger & Sara F. Stein (2023) Conducting virtual assessments in developmental research: COVID-19 restrictions as a case example, Applied Developmental Science, 27:1, 1-17, DOI: 10.1080/10888691.2021.1989305

To link to this article: https://doi.org/10.1080/10888691.2021.1989305

Published online: 05 Nov 2021.

APPLIED DEVELOPMENTAL SCIENCE
2023, VOL. 27, NO. 1, 1–17
https://doi.org/10.1080/10888691.2021.1989305

Routledge
Taylor & Francis Group

Conducting virtual assessments in developmental research: COVID-19 restrictions as a case example

G. Anne Bogat[a], Kristyn Wong[a], Maria Muzik[b], Joseph S. Lonstein[a], Amy K. Nuttall[a], Alytia A. Levendosky[a], Cara F. Colao[a], Alanah Hall[b], Kara Cochran[a], Kathryn R. Forche[c], Allison Koneczny[b], Amanda Gareffa[a], Olivia Oates[b], Stephanie Robinson[b], Alexandra Ballinger[a], and Sara F. Stein[b]

[a]Michigan State University; [b]University of Michigan; [c]University of Detroit Mercy

ABSTRACT

Developmental researchers face considerable challenges regarding maximizing data collection and reducing participant attrition. In this article, we use our experiences implementing our study on the effects of timing of prenatal stress on maternal and infant outcomes during the COVID-19 pandemic as a framework to discuss the difficulties and solutions for these challenges, including the development of two types of virtual assessments. Specific information regarding use of virtual platforms, confidentiality, engaging children during video conferencing, and modifying the major assessments of our research are discussed. Feasibility data are presented, and data analytic challenges regarding statistical inference are outlined. Finally, we conclude with some of the unintended positive consequences for our research that resulted from making these modifications to our original methods.

Overview

The current article details how our federally-funded longitudinal study, the Prenatal Stress Study, based at two large Midwestern universities, adapted our assessments and procedures for virtual administration in order to allow us to continue data collection during the COVID-19 pandemic. University restrictions on collection of participant data necessitated these changes; however, we believe that our adapted methods provide guidance for developmental researchers who confront numerous logistical issues involving participant data collection.

Due to concerns about the transmissiblity of the COVID-19 virus, our universities shut-down in-person research in March, 2020. At the time, there was no indication when it would be allowed to resume. Between March, 2020 and July, 2021, our universities permitted us to conduct virtual and/or in-person assessments for 41 weeks; for the other 26 weeks, we could only administer questionnaire data via an online platform. Because our research included numerous assessments that involved staff interacting with the women and women interacting with their infants/children, we were concerned about the amount of missing data that would result from the in-person research prohibitions. In order to maximize the amount of data we collected from our participants when in-person data collection was not possible, we developed two different types of virtual assessments: 1) virtual visits without drop-off of materials and 2) virtual visits with (contactless) drop-off of materials. The virtual visits without drop-off were conducted and filmed using a videoconferencing software. The virtual visits with drop-off used the same videoconferencing software, but we also provided the participants materials for some of the key assessments (e.g., toys for the mother child teaching tasks, heart rate monitors to measure heart rate variability). We offered these virtual assessment options to participants when the universities did not allow in-person research to take place as well as when participants were not comfortable or able to come to our offices, even when it was permitted by the universities. For example, we found that virtual assessments were accessible for participants who had moved considerable distances from our universities and would otherwise have attrited.

In this article, we first provide a brief overview of our study design to contextualize our assessments and necessary virtual adaptations. We then describe broad

CONTACT G. Anne Bogat bogat@msu.edu Department of Psychology, Michigan State University, 316 Physics Road, Room 262, East Lansing, MI 48824-1312, USA.

considerations for designing virtual assessments before outlining our specific virtual adaptations. Finally, we discuss the feasibility of virtual assessments, including the participation and retention rates prior to and during COVID-19, as well as both the strengths and problems inherent in our adaptations.

Research methods for the Prenatal Stress Study

Our current research, funded by two R01s (large grants geared to a specific research project) from National Institute for Child Health and Development (NICHD, National Institute of Health), is a multi-site longitudinal study investigating the effects of the timing of prenatal stress on infant and child development. Data collection occurs at three sites and is coordinated by two large, midwestern, Research 1 universities. The goals of the research are to study how various stressors (e.g., food insecurity, intimate partner violence, general life stress, relationship stress, and exposure to community violence) experienced at different times during pregnancy differentially affect mother's stress physiology, mental health, and caregiving activities, and how these factors then relate to individual differences in infants'/children's own stress responsiveness and socioemotional development. A major strength of our project is the comprehensiveness of the stress assessments during pregnancy. There are three pregnancy assessments that take place in university offices that examine physiological markers of stress and weekly surveys addressing specific stressors the women experience. After birth, there are in-person assessments at 1, 6, and 30 months. Building on the two grants, a third grant [an R03 (2 year small grant) also from NICHD] focuses on the effects of intimate partner violence on prenatal and early postnatal bonding. [See Levendosky et al. (2021) for a more thorough explanation of the methods of our research.]

Similar to other studies of high risk parents and children, women are recruited into our study through many avenues, including Facebook advertisements, flyers posted at community sites, OB/GYN clinics, various community agencies serving pregnant and parenting women (e.g., WIC, a federal program providing supplemental nutritional food and information for women, infants, and children), and a hospital registry. Women interested in participating are screened on various criteria. Assessments occur, in-person, at three time points during pregnancy—15-17 weeks, 23-25 weeks, and 32-34 weeks. These time points were chosen based on our hypotheses that they were sensitive periods in fetal development during which perturbations would affect specific brain functions related to later infant/child physiological and behavioral stress sensitivity (e.g., Charil et al., 2010; De Bellis et al., 1999; Ulfig et al., 2003; Weinstock, 2008). Women must participate in at least 2 of the 3 in-person pregnancy assessments in order to continue in the research study. This decision rested, in large part, on the foundational hypotheses of the research regarding timing of prenatal stress that rely on the physiological data we collect repeatedly during these assessments.

In addition, during pregnancy, women complete weekly questions via online surveys that assess experiences of stress. After the birth of the child, in-person assessments of mother and child occur when children are 1, 6, and 30 months. There is also a planned assessment at 4 years of age, but this has not yet occurred and is not a focus of this article. There are other online assessments that occur throughout this time. After the 6 month assessment, women complete online surveys every three months to assess key study variables. When children are 1-, 2-, and 3-years-of-age, women complete more extensive online surveys that focus especially on their child's development.

Considerations for virtual assessments

At the beginning of the COVID-19 pandemic shutdown in our state (late-March, 2020), there was a period of time during which our universities would not allow us to conduct in-person research. That is, we could not recruit new participants nor continue in-person assessments for participants already enrolled in the study. We continued to send enrolled participants online questionnaires to complete as they reached particular waves of data collection. As we waited to find out when our universities would allow in-person research to resume, albeit with revised protocols focused on safety for participants and research staff, we developed two types of virtual assessments: virtual visits *without* drop-off of materials and virtual visits *with* drop-off of materials.

Participants were called to schedule the assessment and a determination was made as to which type of virtual visit would take place. Various factors (e.g., fears of COVID-19 transmission, participant moved domicile) sometimes precluded the scheduling of a virtual visit with drop-off. After scheduling, participants were sent a confirmation email with a reminder of the date of the assessment, a copy of the consent

form, and directions for using the video conferencing platform. Detailed directions were provided for downloading the appropriate video conferencing application as well as how to use the application for their visit. This was complicated because of the multitude of different platforms that participants might use (e.g., iPhone, tablet, PC computer, etc.). See Appendix A for a copy of the instructions for one video conferencing platform.

Both types of visits varied somewhat depending on which wave of data we were collecting. The specifics for each of the major assessments (i.e., mother-child free play, affect knowledge test, and heart rate variability) are described later in the article, but here we provide a general overview of each virtual visit.

Virtual visit without drop-off

All waves of assessments involved mother-child interactions. For the virtual visit without drop-off, the mother used her own laptop, cellphone, or tablet to record the mother-child interaction. Research assistants (RAs) were on a videoconference call with the participant to aid in the placement of the camera. The mother was instructed to find toys in the home that were comparable to those we used for in-person assessments. For the gift delay task, we sent the mother a wrapped present from an online source. Other tasks (e.g., teaching tasks, heart rate monitoring, collecting biomarkers of stress) could not be done without dropping off specific materials at the participant's home.

Virtual visit with drop-off

COVID-19 restrictions necessitated specific procedures for drop-off of materials that might not be necessary for researchers without these concerns. In our study, two RAs were involved with each virtual visit with drop-off. The first RA left the materials at the participant's doorstep and then called the participant to let them know that the materials were delivered. This RA watched from a distance to determine when the mother had retrieved the materials. Once this occurred, a second RA was notified, and this RA conducted the virtual visit from a private home or university office. When the visit was completed, the first RA returned to the home to pick up the materials.

When participants allowed us to drop off materials, we were able to collect somewhat more data than we could with the virtual visit without drop-off. In order to aid the recording of the mother-child interaction, we sent video cameras and tripods to the participants' homes. Again, the RAs were on a videoconference call with the participant to aid in set up of the equipment. We also provided materials for the teaching task. These were assembled in plastic containers, each of which was clearly labeled and organized by task to indicate how and when the participant was expected to use them. Heart rate monitors were provided and the participant was instructed on the placement of the monitor on herself and her child. Again, however, we could not collect biomarkers of stress.

Because of COVID-19 contamination concerns, we were careful to purchase materials that could be cleaned and disinfected after each visit. However, regardless of whether assessments are taking place during a pandemic, this is an important consideration for research with children. Commercially available disinfectant wipes were used for hard plastic or wood materials. For materials such as cloth, disinfectant ultraviolet light wands were used. We also informed participants of our hygiene procedures to instill confidence in the safety of the drop-off materials.

Development of the virtual assessments

We developed virtual assessments because of several concerns. The first was that currently enrolled women would drop out of our research because of reduced opportunities to earn compensation for participation. As with all longitudinal research, in order to increase retention, we establish a trusting relationship with our participants and maintain regular contact with them. Participants are informed in advance of the incentive structure for participating in our research, and the shut-down of in-person research meant we would not be able to meet those expectations. Second, because we could not recruit new, pregnant participants nor conduct in-person assessments with those currently enrolled, we were concerned that a substantial pause in our research would seriously undermine data collection that we deemed essential to the major hypotheses of the research related to timing of prenatal stress. These concerns led us to explore and develop the two types of virtual assessments.

When we began deliberating how we could collect data virtually, there was little guidance for our procedures in the research literature. However, there was some advice written by clinicians. For example, Pearson Assessments, the copyright owner of the Wechsler Intelligence Scales, provides psychologists with information for administering virtual

assessments, acknowledging that not all Wechsler scales are possible to administer virtually (e.g., block design). Citing Eichstadt et al. (2013), Pearson recommends that psychologists give careful attention to the tele-assessment equipment and environment (https://www.pearsonclinical.co.uk/Sitedownloads/WAIS-IV/PDFs/admin-wais%E2%80%93iv-via-tp.pdf). For example, professionals should consider the stability and reliability of the wireless connectivity that will be used, the importance of a reasonably large digital image for both parties to view, the ability to screen share, and, if possible, the use of a "stand-alone peripheral camera" that can more easily capture the movements of the participant. Eichstadt et al. (2013) also emphasize the importance of good lighting as well as minimizing disruptions and distractions. These were all factors that we considered in our virtual adaptations of the in-person assessments.

Both types of virtual visits raised issues regarding the quality of the data as well as confidentiality. Farmer et al. (2020) note, when administering virtual clinical assessments, the distractions of the home environment might affect performance and thus the quality of the data obtained. Many of our participants live in small homes and, in addition, have multiple family members or friends living with them. These distractions might also compromise confidentiality. We gave specific instructions to our participants regarding privacy when we scheduled the appointment and then, again, when the assessment began. We emphasized that the interview should be done privately and with no or few distractions. We also acknowledged the difficulty of this given the current circumstances where families were spending more time than usual with each other given the COVID-19 restrictions in our state. However, even without COVID-19 restrictions, researchers should be sensitive to participants who live in multi-family households and homes where space is at a premium.

Because many women in our study were in relationships where intimate partner violence occurred, we did a specific screening for safety before scheduling either type of virtual visit. Figure 1 provides the questions and decision tree we used to determine whether it was safe for the woman (and/or child) to participate in a virtual visit. If the participant and the RA felt that the woman would be safe, the visit was scheduled. On the day of the visit, we emphasized that if, for any reason, participants could not complete the assessment or felt it was unsafe to continue, they could stop the videoconference at any time. Although there were some distractions that occurred during these virtual assessments (e.g., other children or adults coming into the room), none of our participants indicated that their confidentiality or safety was compromised.

Prenatal Stress Study assessments

Our research employs numerous self-report and maternal-report questionnaires that did not need modification as a result of COVID-19 restrictions. However, some of the developmental assessments and methods for determining observed parenting, child emotion knowledge, child self-regulation, and physiological indicators of stress that involve mothers interacting with their children or children interacting with RAs were modified. Each of these assessments/methods was chosen because of its widely documented use in the field and within the existing literature.

Parent-child interaction paradigms in our study include a free play task and teaching task, both of which are commonly employed to observe parent-child interactions (Kerig & Lindahl, 2000). Free play interactions are widely used to assess various aspects of parent-child relational quality, affect, and parenting (e.g., Crowell & Fleischmann, 1993; Deater-Deckard, 2000; Landry et al., 2008; Owen, 1992). In addition, because free play paradigms are widely conducted in early childhood research, they provide the potential to be coded with multiple parent-child coding schemes (Kerig & Lindahl, 2000). Many types of structured parent child interaction tasks are also utilized within the literature, with teaching tasks being commonly used with dyads beginning at 12 months and into preschool (Egeland et al., 1983; Loop et al., 2017; Zeanah et al., 2000). Teaching tasks provide a unique opportunity to observe the dyad interacting within a standardized, increasingly stressful context.

In addition, we assess several biomarkers of stress (i.e., salivary cortisol and alpha amylase) and other physiological measures including maternal and child heart rate. Both salivary cortisol and alpha amylase are well-known metabolites that respond to social stressors (Barbazanges et al., 1996; Davis & Granger, 2009; De Weerth & Buitelaar, 2005; Granger et al., 2007; Hill-Soderlund et al., 2015; Koss & Gunnar, 2018). Maternal and child heart rate are well-established measures for capturing autonomic nervous system activity as a proxy for physiological stress reactivity (e.g., Entringer et al., 2010; Lunkenheimer et al., 2018; Michels et al., 2013).

Child assessments used in our study and discussed in more detail below included two affect knowledge

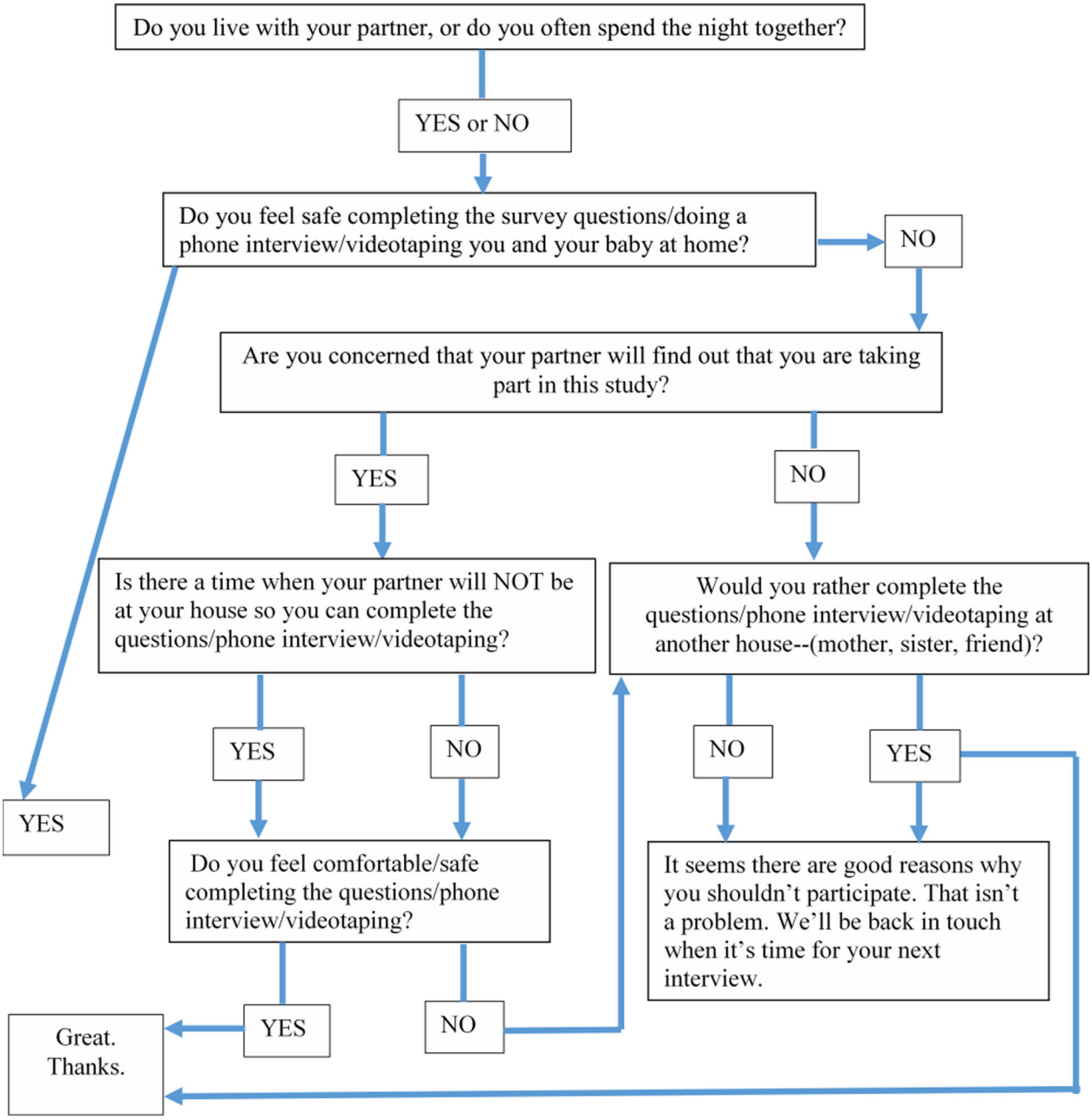

Figure 1. Decision tree to determine participant's safety for virtual visits

tasks and a gift delay task. The affect knowledge tasks were adapted from Denham's (1986) battery for assessing preschooler social cognition and emotion because of its established reliability and validity (Denham et al., 2003). Existing studies have used these assessments to link preschooler emotion knowledge to child self-regulatory outcomes and later academic success (Denham et al., 2012; 2014). Children in our study also participate in a gift delay task, which was a modification of the "wrapped gift" task (Kochanska et al., 2000). The wrapped gift and/or gift delay task has been widely used to examine various aspects of young children's emotion regulation and effortful control (e.g., Carlson, 2005; Joyce et al., 2016; Kochanska et al., 2000).

There were some assessments that needed little to no modifications for the virtual visits. These included a maternal receptive vocabulary test (Dunn, 2019), used to control for maternal verbal intelligence, and a Baby Stroop Test (Hughes & Ensor, 2007), a developmentally appropriate version of the frequently used test to assess effortful control (Kochanska et al., 2000). For virtual visits with drop-off of materials, we were further able to collect cheek cells from buccal swabs,

which are proposed to be used in the future to examine DNA methylation, as well as deliver saliva collection tubes so the mothers could provide samples, at their homes, for salivary diurnal cortisol and alpha amylase analysis. These samples were stored in participants' freezers and retrieved several days later. The maternal diurnal cortisol and alpha amylase levels are proposed as mechanisms through which prenatal stress may affect infant and early childhood regulatory functioning (see Glover et al., 2010 for a review).

There were some assessments that could not be adapted for the virtual visits. For example, all tasks of The Laboratory Temperament Assessment Battery (Lab-TAB; Goldsmith & Rothbart, 1999), to assess children's behavioral reactivity and regulation, must be done in-person, using standardized materials and procedures, and thus could not be conducted virtually. Although the Trier Social Stress Test (Kirschbaum et al., 1993), used for collection of challenged cortisol and alpha amylase for the mothers, has been reliably administered virtually (e.g., Gunnar et al., 2021), we did not do so. Our concern was that the participant's environment would have too many distractions and possible stressors (e.g., other children running in and out of the room) to ensure that the Trier was the only stressor and that the baseline and challenged saliva samples could be collected at the specific times necessary to assay for cortisol and salivary alpha amylase. In addition, many of the women in our sample are high risk, including those who experience intimate partner violence. As far as we are aware, no virtual Triers have been administered and validated with at-risk samples. We could also not measure child height and weight for the virtual visits because of the need to use calibrated equipment that was not easily portable.

However, we were able to adapt many of our assessments for the virtual visits. In the following sections we describe the primary assessments that required significant adaptations for virtual assessments: (1) parent-child interactions, (2) child assessments, and (3) heart rate data collection. We first describe the standard procedures for that task followed by modifications for the two types of virtual visits. Modifications were also made for in-person data collection during the COVID-19 pandemic, but these are not discussed in detail. Generally, those modifications involved participants and RAs wearing masks, minimizing the amount of time that RAs and participants were in the same room together (e.g., providing instructions via videoconference call from a separate room), and RAs not touching or holding the children.

Parent-child interactions

All mothers were given standard instructions prior to recording the parent-child interactions including remaining seated so that their faces were in view of the camera as well as refraining from significant movement and cell phone use. During the virtual visits, we provided checklists to the mothers to ensure the best possible recordings. There were two parent-child interactions: a mother-child free play at 1-, 6-, and 30-months and a mother-child teaching task at 30 months.

Mother-child free play

1-month-old free play. In our original design, 1-month assessments took place in the participant's home. RAs brought a mirror, bouncy seat, heart rate monitor, and recording equipment to the visit. A 2-minute baseline heart rate of the mother was recorded. Then RAs left the room and dyads completed an 8-minute free play interaction that was video recorded while mothers were wearing the heart rate monitor. Mothers were instructed to keep the baby seated in the bouncy seat and to do their best to keep the baby awake for the entire interaction. If the baby became upset, mothers could hold the baby. Also, if a mother was uncomfortable sitting on the floor, the RA helped find an alternative place for the interaction to take place.

Virtual visit without drop-off for 1-month-old free play. Participants accessed our video platform and were instructed how to best set up their recording devices (i.e., cell phone, laptop, or tablet) for the mother-baby interaction. Mothers with bouncy seats were instructed to sit on the floor facing the baby. If the mother did not have a bouncy seat, she was instructed to use supports such as pillows or blankets to prop the baby facing her while the two of them sat on either a couch or a bed. The RA screen-shared still shots of sample camera angles to assist participants in setting up the video and offered suggestions for propping up the recording device using materials available in the home. RAs focused on ensuring that backlighting or distracting sounds did not hinder videotaping. Mothers were then provided with the same directions as for pre-COVID-19-in-person home visits.

Virtual visit with drop-off for 1-month-old free play. Participants were provided with a tripod, video camera, heart rate monitor, and a bouncy seat. Using our video equipment reduced a number of difficulties we had encountered when the mother used her own equipment, including suboptimal videos due to participant recording devices and unstable internet

connections. Participants were instructed by the RA how to set up the tripod and camera, and guidance was provided to ensure that both the mother's and infant's face and body could be seen in the camera frame.

6- and 30-month old free play. The 6-month old mother-child free play interaction took place in university offices in our original design. Baseline heart rate for both mother and infant were recorded for 2 minutes. Mother-infant dyads were seated on a play mat and mothers were asked to play with their child as they normally would for 8 minutes, during which time heart rate data were also recorded. If the mother was uncomfortable sitting on the floor, alternative arrangements were made. At this visit, dyads were provided with a set of age appropriate toys that included soft books, stacking rings, toy cars, and farm animals on a string. We had not begun the 30-month assessment prior to COVID-19; thus, all in-person procedures for the 30- month-olds were administered after the COVID-19 pandemic began and thus adhered to COVID-19 safety procedures.

Virtual visit without drop-off for 6- and 30-month-old free play. Mothers were instructed prior to the scheduled visit to find specific types of toys in the home that were similar to the standardized toys we used for the in-person visits (e.g., stacks of rings, dollhouse, etc.), avoiding toys that were electronic or noisy. On the day of the virtual visit, participants used their own devices (i.e., cell phone, laptop, or tablet) to access the video platform, and the RA assisted them in setting up their recording devices and placing themselves and their children in the correct positions. All other directions for the free play administration remained the same as the in-person visit.

Virtual visit with drop-off for 6- and 30-month-old free play. Participants were provided with a tripod, video camera, and a set of the same toys used at the in-person assessments. Similar to the 1-month virtual visit with drop-off, these assessments were administered via videoconferencing, and the interaction was recorded with the camera provided. Instructions to ensure that both the mother's and infant's face and body were visible were the same as for the 1-month free play.

Mother-child teaching task

The teaching task was based on the Caregiver Child Structured Interaction Procedure (Crowell & Fleischmann, 1993).

In-person assessment. In our original design, dyads engaged in four increasingly difficult teaching tasks for approximately 3-5 minutes each. The first two tasks were at or below the child's developmental level and the third and fourth tasks were above the child's developmental level. Mothers were provided a large plastic bin with four different bags of toys; each bag was labeled for mothers to retrieve when the research assistant instructed them to do so. The RAs provided verbal instructions, but written instructions on index cards were also provided to the mother.

Virtual visit without drop-off. Because the teaching task required specific sets of toys of increasing difficulty, this task was not administered during the virtual visit without drop-off.

Virtual visit with drop-off. In addition to the other materials dropped off for the 30-month-old assessment, we provided the identical bin and labeled bags of toys as for the in-person assessment. RAs provided instructions via videoconferencing regarding how mothers should interact. Video recordings followed identical procedures described above for the free play interactions.

Child assessments

One of the challenges of virtual assessments was relying on mothers to help keep their children in front of the camera without also coaching the child or helping them with any of the assessments. We developed detailed instructions for the mothers and the RAs to mitigate this problem.

Engaging the children and keeping their attention also presented challenges. As researchers who assess child participants know, one uses many techniques to keep children interested and focused on the tasks. However, many of these (e.g., liberal use of stickers as an incentive for completing tasks) are not possible with virtual visits. Building rapport during virtual visits presented difficulties. Families often used cell phones for their virtual visits. The small size of the screen made it difficult for the child to see the RA's face at all times. In addition, RAs also paid close attention to the backdrops and lighting in the settings in which they conducted the virtual assessments. It was much easier to get children's attention when they could clearly see the RA's face and the objects she was holding. Although there were many challenges, we developed a number of other strategies for making child virtual assessments run as smoothly as possible, when the children were 30 months old. We believe these strategies would be useful for any researcher conducting a virtual assessment with a child.

First, we always asked the mother whether the child had any experience with video calls (e.g., does the child talk to grandparents, or does the child attend virtual daycare/preschool). If the child did have these experiences, we asked the mother to use these as a framework to explain how our staff would be interacting with the child. If the child had no video call experience, then the staff or the mother explained more about how it worked and what was important (e.g., stay in front of the camera). In discussing clinical assessment, Farmer et al. (2020) noted that children with less exposure to video chatting may engage as they would in-person, but their behavior with the assessor may be atypical because of their unfamiliarity with the medium. Children may be less likely to attend to or complete the tasks that are administered virtually.

Second, for tasks that we needed to discuss with the mother and that we did not want the child to hear the instructions, we screen shared some typewritten text explaining what was about to happen and what the mother should do during the task. The screen sharing sometimes included pictures of the task (e.g., a drawing or screen shot of the ideal camera set up for the mother-child free play and teaching task).

Third, in the same way that tone and demeanor of the interviewer are important when interacting in-person with a child, these are also important considerations in the virtual assessment. We encouraged our staff to be energetic and happy—perhaps even more so than when they were assessing a child in-person.

Fourth, we found there was often a lag in audio on our video conferencing software. When training staff, we emphasized the need to be patient and pause and wait somewhat longer than normal to give the child time to respond.

Fifth, similar to in-person assessments, we used language that made the child feel they were assisting the staff person or that the staff person and the child were playing a fun game together. This became extremely important during virtual assessments as there were sometimes distractions in the home environment. We were constantly talking to and reinforcing the child's behavior (e.g., "I really appreciate you helping me with these questions," "This was much easier/more fun with your help," or "Now we're going to play a fun game with some spoons").

Affect Knowledge Test

(only administered at the 30-month assessment, which began after our COVID-19 campus shutdowns)

In-person assessment. Two paradigms from Denham's (1986) Affect Knowledge Task were modified for the assessment. For the first task, children were administered the emotion faces task in which they were presented with a board laminated with four different emotional expressions–happy, sad, mad, and scared. For the second task, children were administered a puppet situations paradigm.

The RA presented the child with three different puppets, the four faces from the previous task (happy, sad, mad, scared), and eight different story stems, in a predetermined order. Children were asked "How does the puppet feel" and to "Give the puppet a face" after each story stem. The original task instructions required one adaptation due to the COVID-19 pandemic context. The original instructions would have had the RA acting out each of the four emotions while showing the child the matching face. During COVID-19 assessments, RAs wore face masks, therefore, we made a video of an individual acting out each emotion while holding up the corresponding cartoon face. This modification had the advantage of standardizing the instructions and visual prompts across in-person and virtual assessments.

Virtual visit with and without drop-off. A PDF version of the emotion faces board was created. For the first portion of the task, the RA shared the screen of the emotion faces board with the child and directed the mother to verify the child's choice. We numbered the faces so the mother would not state the correct (or incorrect) answer by verbalizing what the child had pointed to (e.g., the "happy face"). For the second portion of the task, the child watched the prerecorded video of an RA acting out the emotions and then completed the puppet situations task.

Gift delay

(only administered at the 30-month assessment, which began after COVID-19 campus shutdowns)

In-person assessment. A standard gift delay task was administered (Kochanska et al., 1996; 2000). A wrapped gift was placed on a table in front of the child. The RA indicated she had to step out of the room to retrieve a bow for the gift. Children were instructed to wait and not touch the box until the RA returned, and the mother was told not to talk or react to the child during this time. The RA left for approximately 3 minutes before returning.

Virtual visit without drop-off. Prior to the visit, a gift was sent to each participant's home from an online vendor. The gift arrived as a wrapped present, and

mothers were told to not show the gift to the child prior to the assessment. During the assessment, mothers were instructed by the RA to have the child sit on the floor in front of the camera and place the gift in front of the child. The mothers then left the room for 3 minutes and told the child not to touch the gift until they returned. Upon her return, mothers told the child that they could open the gift.

Virtual visit with drop-off. The only modification was that the wrapped gift was provided in the set of materials left at the door rather than being mailed to the participant's home.

Heart rate variability

Original in-person assessment

Electrocardiography (ECG) data were recorded via the Bittium Faros wearable 3-lead ambulatory device (Bittium Corporation, Oulu, Finland) at multiple waves of the study. Mother's heart rate was recorded during various tasks during the third pregnancy visit and the 1-month post-birth visit. At the 6- and 30-month assessments, heart rate data were collected for both mother and infant/child. Prior to adhering the electrodes to the skin, the RA cleaned the skin with an alcohol pad, and then attached the three leads to the collarbone and chest. For the first two minutes, a baseline reading was obtained, and then heart rate was recorded during the tasks.

Virtual visit without drop-off

If participants chose the fully virtual visit, we could not provide the heart rate monitor. Therefore, heart rate data were not collected.

Virtual visit with drop-off

For virtual visits where participants agreed to a contactless drop-off of materials, ECG collection instruments and accessories were provided. These included the heart rate monitor (or monitors, if a child was also going to wear one) with leads and electrode pads already attached, alcohol swabs for cleaning the skin prior to placing the device, and a laminated photograph of correct placement of the device on the body. If two ECG devices were being dropped off, they were in separate bags labeled "Mom" and "Child," and the devices themselves were also labeled to ensure there would be no confusion as to which device held which person's data. During the visit, the RA verbally instructed the participant how to attach and turn on the heart rate monitors via videoconference. A short video depicting how to attach the monitors and turn them on was also available as needed. The materials were then retrieved from the participant's doorstep via contactless pick-up.

Feasibility of implementation of adaptations

As outlined above, we were able to adapt many of our assessment methods to virtual visits without drop-off or virtual visits with drop-off of materials. The success of such adaptations to research hinges, in part, on the feasibility of implementing these adaptations and whether we were able to retain participants in our study using these techniques. We have some preliminary data that support both feasibility and retention.

The following information describes assessments conducted between March, 2020 and July, 2021 (67 noncontiguous weeks). For 26 weeks, between these two dates, we were not allowed to conduct any research involving human participants (virtual or not). During the remaining 41 weeks, assessments occurred virtually, virtually with drop-off of materials, and in-person. However, because we were given permission much later during this time period to conduct in-person assessments, few of these were given through July, 2021. Our experience is that many participants prefer in-person assessments when the option is given (and in this case, when it was allowed by our universities). However, those women living in geographic areas with high rates of COVID-19 transmission were more reluctant to agree to in-person assessments. When in-person assessments were not possible, participants readily engaged in the virtual assessments.

Between March, 2020 and July, 2021, we conducted 240 assessments, on 147 participants, across all waves of our study, an 83% completion rate. The data we present focuses on participants already in the study when COVID-19 began, hence newly recruited participants, engaging in the first pregnancy visit, are not included. For the second and third pregnancy assessments we had a 91% and 95% completion rate, respectively. Eighty-six percent of these were fully virtual assessments (there was no virtual with drop-off possible, as explained earlier), and 14% were in-person.

Overall, for our post-pregnancy assessments at 1-, 6-, and 30-months, we were able to complete 80% of those scheduled. The percent varied across waves: 81% for 1-month; 93% for 6-months, and 66% for 30 months. Of the post pregnancy assessments, 19% were in person, 69% were fully virtual, and 12% were virtual with drop-off of materials. Our participants

opted for both types of virtual visits because of COVID-19 concerns regarding transmission of the virus. However, some participants did not have a choice; they were assessed virtually because, for some period of time, one of our universities did not allow children younger than 2 years of age in university buildings. Other participants had moved out of the area in which our universities were located. The 30 month assessment had a lower completion rate (66%), possibly because of the length of time between the 6 month and 30 month assessments. Although we keep in touch with our participants via email contact and surveys between those times, there are greater challenges scheduling a visit with a 30 month old compared to a 1- or 6-month old (e.g., the mother may have another child or have resumed employment).

We also compared assessment completion rates (assessments completed ÷ assessments due) for the year prior to COVID-19 and during COVID-19. All assessments had higher completion rates during COVID-19. We were able to complete *more* second and third pregnancy assessments (91% vs. 78%; 95% vs. 78%), more 1-month assessments, (81% vs. 74%), and more 6-month assessments (93% vs. 89%). (Because the 30-month assessment began during COVID-19, we do not have comparisons with the prior year.) We do not know whether these numbers reflect the reduced burden on the participants (that is, is it always easier for participants to schedule a virtual assessment) or whether the restrictions of movement during COVID-19 (e.g., individuals working from home, unemployment, businesses not open for in-person transactions) made our participants more generally available.

In examining the comparisons between the two time periods (pre- and during COVID-19), the 1-month assessment had the lowest completion rate for both time periods; however, the differential favors the virtual assessment. Prior to COVID-19, all 1-month assessments took place in the participant's home. During COVID-19, the assessments could only be virtual or virtual with drop-off of materials. However, the visit had to be scheduled at a convenient time for both the RA and the participant. The 1-month assessment is our shortest. Perhaps women were able to squeeze in an appointment for a virtual assessment more easily than scheduling with our project staff.

In summary, our assessment completion rates indicate that fully virtual and virtual visits with drop-off are viable, and, in fact, we had higher completion rates for the assessments when compared to the year prior to COVID-19. During COVID-19, the use of virtual assessments has allowed (and continues to allow) us to retain participants and collect as much data as possible from them. However, the adaptations also create data analytic challenges that we discuss next.

Implications of adapting research protocols for data analysis

Once data collection in our longitudinal study is complete we will need to consider how the assessment adaptations influence our data analysis plans. Therefore, next we discuss problems that we will be confronting as we think about the statistical analyses using data from our study that were collected in multiple ways (e.g, in-person and virtual) as well as during different times (e.g., before and during the COVID-19 pandemic). We discuss three of these challenges here.

Fidelity

Fidelity criteria allow us to maximize internal validity. In order to assess the fidelity of our pregnancy assessments (which all have and still occur in-person), we have checklists that allow us to determine whether our RAs are implementing the assessment in a standardized fashion. Because the assessments are standardized, we can be relatively confident that we have minimized the introduction of confounding variables into the assessment.

On the other hand, there are now four types of assessments for the post-pregnancy waves: in-person before COVID-19, in-person with COVID-19 adaptations, virtual visits without drop-off, and virtual visits with drop-off of materials. We train our RAs to reliability on each of these as well as watch videos of their assessments, post training, to determine whether the administration is standardized. However, we do not yet know whether these differences in data collection methods have an effect on the participants' responses. Ideally, researchers wishing to use virtual assessments would only use virtual assessments to prevent this issue. However, in the case of the COVID-19 pandemic this was not possible for our study. This is likely also to be true in other contexts, such as when in-person studies wish to use virtual assessments for retention of participants who have moved away.

Measurement invariance must be determined when administration of an assessment is standardized, but the assessment might not measure the same construct

across different participants (e.g., age, culture). However, here the concern is that participants might respond differently to assessments administered using different methods. For example, because of COVID-19, as well as the need to reach underserved populations, virtual assessments for Autism Spectrum Disorder have been developed, but their psychometric properties are not yet known (Berger et al., 2021). Various testing companies have been quite interested in the issue of in-person vs. virtual equivalence of assessment instruments. For example, Pearson Assessments summarizes technical reports showing the equivalence of digital vs. traditional formats for administration of the Wechsler scales for both clinical and non-clinical populations; however, these are not peer-reviewed studies (https://www.pearsonclinical.co.uk/Sitedownloads/WAIS-IV/PDFs/admin-wais%E2%80%93iv-via-tp.pdf). Psychological Assessment Resources cites some additional research with their assessment instruments; these are white papers not published in refereed journals. One of these is from Wright (2018) who reports results for a case-control match design format of the Reynolds Intellectual Assessment Scales—Second Edition. The sample was quite small, with a wide range of ages tested (age 3 to age 19). Findings indicated no significant difference for the various subtests. However, the Speeded Processing Index did show a difference—scores were significantly higher when administered in-person compared with virtually. Farmer et al. (2020) note that taken as a whole, studies assessing equivalence between in-person and virtual formats "are limited to statements about condition (i.e., *group score equivalency*), and cannot address whether raw scores obtained by an individual are equivalent across formats (i.e., *individual score equivalency*)" (page 479, italics in the original text).

As researchers begin to integrate virtual assessments into their studies, it is important to determine the fidelity of the virtual administrations. Some research on this has already been reported. For example, Manning et al. (2020) compared child language samples obtained in-person or online during mother-child play. They found that both methods yielded the same number of usable samples and the same speech and language characteristics. Importantly, group analyses and within-child comparisons were equivalent. Other researchers have demonstrated psychometric equivalence between in-person and online administration of specific assessments (e.g., Brock et al., 2015 for psychological, physical, and sexual aggression questionnaires). Another method to test equivalence is to replicate published in-person research by conducting online versions of that research and making comparisons (e.g., Nussenbaum et al., 2020). When conducting assessments with young children, one concern is that parents might intervene when the assessments are conducted online and in the home. However, at least one study found that parental interference occurred in less than 1% of the sample (Rhodes et al., 2020).

To make use of these fidelity methods requires advanced planning. Unfortunately, for our longitudinal study that was in process when COVID-19 began, this was not possible. We will have to conduct fidelity checks *post hoc*. One option, similar to tests for measurement invariance, would be to run a factor model for each situation in which the variable that is observed under different conditions is related to a selection of other variables (always using the same other variables). If fidelity exists, the factor models should be the same across the different administrations.

Missingness

Missing data in research studies is inevitable. In any cross-sectional study, some data can be expected to be missing due to non-response, which is data missing for a portion of the protocol. In longitudinal studies, data may additionally be missing due to attrition, which is data missing for an entire case within one or more measurement occasions (Little, 2013). Unfortunately, missing data are not always properly considered when performing statistical analyses. For example, a review of developmental studies found that 82% of researchers used inappropriate methods to account for missingness, with the most common error being deletion methods resulting in high levels of bias (Jelicić et al., 2009).

Modern missing data techniques include model-based approaches, such as full information maximum likelihood (FIML), and data-based approaches, such as multiple imputation, that attempt to correct for the potential bias introduced by missing data and regain power (Enders, 2010). Use of these approaches requires considering the reason the data are missing. There are three mechanisms characterizing data that are missing: Missing completely at random (MCAR), a truly random process; missing at random (MAR), a measured and predictable process; and missing not at random (MNAR), an unmeasured and unpredictable process (Little & Rubin, 2002; Rubin, 1976, 1987). In the case of MAR, measured variables that explain the

missing data mechanism may be incorporated into the missing data technique in order to recover the missingness; these are called auxiliary variables (e.g., Enders, 2008; Graham, 2003). When data are missing due to the MCAR or the MAR process, modern missing data techniques appropriately address these processes (e.g., Enders, 2010). However, in any study it is likely that missing data are attributable to all three of these missing data processes (e.g., Little, 2013).

In our study, some missing data are directly attributable to the COVID-19 pandemic. For example, as stated earlier, all women had to come for at least two of the three in-person assessments during pregnancy. During the COVID-19 pandemic, some women may have missed an appointment because they were uncomfortable coming to our offices for in-person data collection. Although we were able to proceed with data collection virtually under the adaptations described above, and this prevented the introduction of missing data for some tasks, which is the goal of the present paper, there are tasks we were unable to adapt (e.g., the Trier Social Stress Task and the collection of cortisol and salivary alpha amylase reactivity). Because such missing data are attributable to an individual's experiences or appraisals of the pandemic, the missing data process is MAR. We expect that beyond the COVID-19 pandemic there will be tasks that other research teams will not be able to adapt for virtual assessments resulting in a MAR missing data mechanism.

During the COVID-19 pandemic we also introduced new questions resulting in missing data under the original (pre-COVID-19) study design. For example, we added COVID-19-specific stress questions that are parallel in form to our original stressor-specific questions. This was critical for our study given our interest in stress and assessment of specific stressors. Adding questions or measures to a study mid-data collection is not uncommon and can usually be thought of as MCAR because the order in which participants are enrolled into the study is, in theory, random. In this context, however, data are missing due to the timing of pregnancy and the COVID-19 pandemic and added questions are not relevant prior to data collection; therefore, these data are intentionally missing and do not require a missing data technique (cf. Sackett & Yang, 2000).

Threats to internal validity related to the COVID-19 pandemic

In our research, participants are enrolled on a continuous basis until the expected final sample size is reached. Thus, some participants will be pregnant during COVID-19 and others will not. However, they will all be parenting during COVID—but at different points in the child's life. A history effect threatens internal validity by impacting the independent and dependent variables and doing so differentially across participants (Shadish et al., 2001). In the case of our study, COVID-19 is a stressor and given our focus on stress in pregnancy (our independent variable) and, for example, child emotion regulation (a dependent variable), COVID-19 presents a clear confound of the effect of stress on our dependent variable. Given our rolling enrollment into the study, the timing of the COVID-19 pandemic differs across participant's individual timing of pregnancy gestation and child age. Thus, the COVID-19 pandemic began before pregnancy for some participants, during pregnancy for others, and after pregnancy for others. Therefore, a history effect is a likely concern in our study. Another way to conceptualize this internal validity concern is to view our longitudinal study design as a cohort-sequential design (Nesselroade & Baltes, 1979) where each participant's pregnancy timing is a cohort effect. Such a design is well-suited to understanding age effects while controlling for cohort effects (which in our study is our central question of the gestational/child age timing of stress) yet susceptible to history effects, because it is difficult to disentangle the history effect from the age by cohort interaction (Little, 2013). We have attempted to offset this threat to validity by adding COVID-19 specific stressor items that parallel our other independent variables in order to explicitly assess the impact of COVID.

Positive consequences of virtual adaptations

Although earlier in this article we focused on some of the research method problems exacerbated by COVID-19, we want to acknowledge that there were positive outcomes that resulted from making the adaptations.

The first is that we learned to conduct virtual assessments. This is a data collection method we probably would have not attempted if we had not been forced to adapt. Because of the ubiquity of cell phones owned by participants as well as the ability to provide participants relatively inexpensive smartphones with data plans, conducting virtual assessments involving videoconferencing is now relatively easy for researchers. Dropping off video cameras for our participants' use increased the quality of our recordings considerably, and is highly recommended where feasible. Going forward, we advise researchers

to plan their longitudinal studies with virtual assessments in mind, especially when considering the number of participants who are likely to move far from where data collection occurs. Researchers could determine which assessments could be done virtually if it is not possible to administer those in-person.

There is very limited guidance within the literature regarding the methodological issues that arise when conducting assessments outside of a standardized lab environment (i.e., in the home). It is likely that developmental researchers have established protocols which they have adapted for their study populations, but this information has not been widely shared or deposited in a publicly available resource such as the IMPACT Measures Repository (https://ctn.uoregon.edu/projects/impact-measures-repository). Sharing our adapted procedures and methods may help to increase collaboration and open the dialogue among researchers who have adapted to challenges, such as those posed by COVID-19, which prohibit or restrict in-person research.

Our innovative virtual assessment adaptations were introduced to reduce missing data challenges, which they did. We hoped to reduce the overall amount of missingness within a given wave (non-response) by allowing participants to complete a portion of an assessment virtually or across waves (attrition) and by allowing participation when women were unable or unwilling to return to our offices for assessments. The virtual assessments also enabled us to be more flexible, generally. For example, if a child was asleep during the initially scheduled time of the virtual assessment, we could postpone the assessment for a later time or day, with minimal inconvenience to our RAs or the participant. Because we obtained informed consent via telephone, participants could complete study questionnaires online before their visits, which helped reduce the length of the virtual assessment and thus reduced the burden on mother and child. Given that our study recruits participants from three distinct geographical locations, virtual assessments meant that RAs from either university could conduct them. In addition, RAs from different universities could work together on a virtual assessment, thus increasing collaboration and fidelity of administration across sites. Finally, various forms of technology allowed us to maximize staff time and enhance availability to our participants. For example, the use of a video-conferencing telephone app allowed RAs to conduct virtual assessments from their homes (always in a private and confidential space), which meant that time was not spent traveling to and from project offices. In addition, by connecting our office telephone lines to an app, we were able to have multiple RAs communicate with participants throughout the day from their cell phones.

Developmental researchers might consider implementing planned virtual assessments for families confronting challenges that create difficulties for them coming to university offices. For example, low income families likely have various transportation barriers, and these might be heightened further in those families with multiple small children. There may also be families who are not college-educated and might be less familiar with and perhaps intimidated by a university setting. Finally, families from historically marginalized groups may be distrustful of research, in general. All of these families might be more amenable to research that takes place in their homes.

A second positive consequence was the change in communication between the teams at our two universities. Prior to COVID-19, the two teams communicated via videoconference, each team sitting in their separate conference rooms. There were a number of problems with this. First, it was frequently difficult to hear when some people spoke, notably those farther from the microphone. This was exacerbated by side conversations going on in each room. In addition, if staff happened not to be in office during the staff meeting, they were not on the videoconference. These absences sometimes led to lack of knowledge about decision-making that occurred relevant to project procedures. Finally, and probably most importantly, having the project staff at each university separately on one screen in the other's conference room reinforced psychological barriers between the two universities/teams. While unintentional, this lack of unity led to reduced levels of communication between staff at the two universities—although we only realized this when our teleconference meetings changed as a result of COVID-19.

Due to the social isolation enforced during the COVID-19 pandemic, our staff meeting teleconference procedures have changed such that each member of the staff and the PIs sit in their own offices/homes and log on separately. There are significant advantages to this approach, which we would not have realized if we had not been forced to try this approach due to the pandemic. Notably, it is much easier to hear from each person who is speaking and there are no side conversations. We also have much better attendance, as staff do not have to be at the office to attend. Most importantly, we feel like one team across universities, which has led to more positive morale. This facilitates

more frequent and better communication between the staff at each university. We all prefer this approach and we will be maintaining this specific approach to video conferencing going forward.

A final positive outcome involves the main focus of our research–to understand the impact of specific prenatal stressors on mother and child outcomes. We are now particularly well-positioned to also provide new information about the impact of the COVID-19 stressor on the lives of mothers and children compared to the other stressors we measure in our study.

Conclusion

Longitudinal research involves some level of attrition and missingness of data. However, because of the restrictions imposed by our universities, as well as the reluctance of participants to engage in in-person assessments, the COVID-19 pandemic would have resulted in unacceptably large amounts of attrition and missing data if we had not developed alternative methods of data collection. We developed two types of virtual visits (with and without drop-off of materials) that enabled us to reduce attrition and missing data. When COVID-19 concerns are past, but when participants move far from the locations where data collection occurs, we can still administer these protocols.

We are aware that many investigators paused their research, waiting until the COVID-19 pandemic passed. But, as we write this manuscript, the COVID-19 Delta variant, with its high transmissibility rate, is sweeping the country calling into question when the pandemic will end. Thus, investigators who paused their research may well have to begin data collection during the pandemic. This will mean that developmental researchers must make determinations as to whether virtual assessments are appropriate and/or feasible and, if such assessments occur, they will inevitably confront many of the data analytic challenges we discussed earlier.

Importantly, for our multisite study, COVID-19 restrictions, which necessitated meetings held via videoconferencing, helped to strengthen collaboration amongst team members and enabled greater communication across sites. Moving forward, sharing the lessons learned by researchers in the developmental sciences regarding virtual assessments with one another, whether due to COVID-19 or not, can help to facilitate efficiency in data collection and research staff administration as we navigate the research landscape.

Data sharing is not applicable to this article as no new data were created or analyzed.

Acknowledgements

The authors would like to acknowledge Georgia Bayer, Jade Kobayashi, Quantanise Williams, Rachael Mccally, Patty Krause, Will Eckerle, and Arsh Singh for their efforts in the implementation of this research study and Alexander von Eye for comments on an earlier version of this manuscript. We would also like to thank the women and children who graciously gave of their time to participate in our research.

Funding

This work was supported by three grants from the National Institute of Child Health and Development: R01HD085990 to AAL, GAB, and JSL (with MM as site PI and AKN as Co-I); R01HD100469 to AAL, GAB and MM (with JL and AKN as Co-Is); and R03HD096141 to GAB, JSL, and AAL (with MM as site PI).

References

Barbazanges, A., Piazza, P. V., Le Moal, M., & Maccari, S. (1996). Maternal glucocorticoid secretion mediates long-term effects of prenatal stress. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *16*(12), 3943–3949.

Berger, N. I., Wainer, A. L., Kuhn, J., Bearss, K., Attar, S., Carter, A. S., Ibanez, L. V., Ingersoll, B. R., Neiderman, H., Scott, S., & Stone, W. L. (2021). Characterizing available tools for synchronous virtual assessment of toddlers with suspected Autism Spectrum Disorder: A brief report. *Journal of Autism and Developmental Disorders*, https://doi.org/10.1007/s10803-021-04911-2

Brock, R. L., Barry, R. A., Lawrence, E., Rolffs, J., Cerretani, J., & Zarling, A. (2015). Online administration of questionnaires assessing psychological, physical, and sexual aggression: Establishing psychometric equivalence. *Psychology of Violence*, *5*(3), 294–304. https://doi.org/10.1037/a0037835

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, *28*(2), 595–616. https://doi.org/10.1207/s15326942dn2802_3

Charil, A., Laplante, D. P., Vaillancourt, C., & King, S. (2010). Prenatal stress and brain development. *Brain Research Reviews*, *65*(1), 56–79.

Crowell, J. A., & Fleischmann, M. A. (1993). Use of structured research procedures in clinical assessments of infants. In C. H. Zeanah, Jr. (Ed.), *Handbook of infant mental health* (pp. 210–221). The Guilford Press.

Davis, E. P., & Granger, D. A. (2009). Developmental differences in infant salivary alpha-amylase and cortisol responses to stress. *Psychoneuroendocrinology*, *34*(6), 795–804.

De Bellis, M. D., Baum, A. S., Birmaher, B., Keshavan, M. S., Eccard, C. H., Boring, A. M., Jenkins, F. J., &

Ryan, N. D. (1999). A.E. Bennett Research Award. Developmental traumatology. Part I: Biological stress systems. *Biological Psychiatry*, *45*(10), 1259–1270. https://doi.org/10.1016/s0006-3223(99)00044-x

De Weerth, C., & Buitelaar, J. (2005). Physiological stress reactivity in human pregnancy–a review. *Neuroscience and Biobehavioral Reviews*, *29*(2), 295–312.

Deater-Deckard, K. (2000). Parenting and child behavioral adjustment in early childhood: A quantitative genetic approach to studying family processes. *Child Development*, *71*(2), 468–484.

Denham, S. A. (1986). Social cognition, prosocial behavior, and emotion in preschoolers: Contextual validation. *Child Development*, *57*(1), 194–201. https://doi.org/10.2307/1130651

Denham, S. A., Bassett, H. H., Way, E., Mincic, M., Zinsser, K., & Graling, K. (2012). Preschoolers' emotion knowledge: self-regulatory foundations, and predictions of early school success. *Cognition & Emotion*, *26*(4), 667–679. https://doi.org/10.1080/02699931.2011.602049

Denham, S. A., Bassett, H. H., Zinsser, K., & Wyatt, T. M. (2014). How preschoolers' social–emotional learning predicts their early school success: Developing theory-promoting, competency-based assessments. *Infant and Child Development*, *23*(4), 426–454. https://doi.org/10.1002/icd.1840

Denham, S. A., Blair, K. A., DeMulder, E., Levitas, J., Sawyer, K., Auerbach-Major, S., & Queenan, P. (2003). Preschool emotional competence: Pathway to social competence. *Child Development*, *74*(1), 238–256. https://doi.org/10.1111/1467-8624.00533

Dunn, D. M. (2019). *Peabody picture vocabulary test* (5th ed.) [Measurement instrument]. NCS Pearson.

Egeland, B., Sroufe, A., & Erickson, M. (1983). The developmental consequence of different patterns of maltreatment. *Child Abuse & Neglect*, *7*(4), 459–469.

Eichstadt, T. J., Castilleja, N., Jakubowitz, M., & Wallace, A. (2013, November). *Standardized assessment via telepractice: Qualitative review and survey data* [Paper presentation]. Annual meeting of the American-Speech-Language-Hearing Association.

Enders, C. K. (2008). A note on the use of missing auxiliary variables in Full Information Maximum Likelihood-Based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*(3), 434–448. https://doi.org/10.1080/10705510802154307

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.

Entringer, S., Buss, C., Shirtcliff, E. A., Cammack, A. L., Yim, I. S., Chicz-DeMet, A., Sandman, C. A., & Wadhwa, P. D. (2010). Attenuation of maternal psychophysiological stress responses and the maternal cortisol awakening response over the course of human pregnancy. *Stress*, *13*(3), 258–268. https://doi.org/10.3109/10253890903349501

Farmer, R. L., McGill, R. J., Dombrowski, S. C., McClain, M. B., Harris, B., Lockwood, A. B., Powell, S. L., Pynn, C., Smith-Kellen, S., Loethen, E., Benson, N. F., & Stinnett, T. A. (2020). Tele-assessment with children and adolescents during the Coronavirus (COVID-19) pandemic and beyond: Practice and policy implications. *Professional Psychology: Research and Practice*, *51*(5), 477–487. https://doi.org/10.1037/pro0000349

Glover, V., O'Connor, T. G., & O'Donnell, K. (2010). Prenatal stress and the programming of the HPA axis. *Neuroscience & Biobehavioral Reviews*, *35*(1), 17–22. https://doi.org/10.1016/j.neubiorev.2009.11.008

Goldsmith, H., & Rothbart, M. K. (1999). *The Laboratory Temperament Assessment Battery (Prelocomotor Version.*, Edition 3.1*)*University of Wisconsin-Madison.

Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*(1), 80–100. https://doi.org/10.1207/S15328007SEM1001_4

Granger, D. A., Kivlighan, K. T., el-Sheikh, M., Gordis, E. B., & Stroud, L. R. (2007). Salivary alpha-amylase in biobehavioral research–Recent developments and applications. *Annals of the New York Academy of Sciences*, *1098*(1), 122–144. https://doi.org/10.1196/annals.1384.008

Gunnar, M. R., Reid, B. M., Donzella, B., Miller, Z. R., Gardow, S., Tsakonas, N. C., Thomas, K. M., DeJoseph, M., & Bendezú, J. J. (2021). Validation of an online version of the Trier Social Stress Test in a study of adolescents. *Psychoneuroendocrinology*, *125*, 105111. https://doi.org/10.1016/j.psyneuen.2020.105111

Hill-Soderlund, A. L., Holochwost, S. J., Willoughby, M. T., Granger, D. A., Gariépy, J.-L., Mills-Koonce, W. R., & Cox, M. J. (2015). The developmental course of salivary alpha-amylase and cortisol from 12 to 36 months: Relations with early poverty and later behavior problems. *Psychoneuroendocrinology*, *52*, 311–323. https://doi.org/10.1016/j.psyneuen.2014.08.011

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology*, *43*(6), 1447–1459. https://doi.org/10.1037/0012-1649.43.6.1447

Jelicić, H., Phelps, E., & Lerner, R. M. (2009). Use of missing data methods in longitudinal studies: the persistence of bad practices in developmental psychology. *Developmental Psychology*, *45*(4), 1195–1199. https://doi.org/10.1037/a0015665

Joyce, A. W., Kraybill, J. H., Chen, N., Cuevas, K., Deater-Deckard, K., & Bell, M. A. (2016). A longitudinal investigation of conflict and delay inhibitory control in toddlers and preschoolers. *Early Education and Development*, *27*(6), 788–804. https://doi.org/10.1080/10409289.2016.1148481

Kerig, P. K., & Lindahl, K. M. (Eds.) (2000). *Family observational coding systems: Resources for systemic research*. Psychology Press.

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The "Trier Social Stress Test": A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, *28*(1-2), 76–81. https://doi.org/10.1159/000119004

Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology*, *36*(2), 220–232. https://doi.org/10.1037/0012-1649.36.2.220

Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young

children and its role in emerging internalization. *Child Development*, *67*(2), 490–507.
Koss, K. J., & Gunnar, M. R. (2018). Annual research review: Early adversity, the hypothalamic-pituitary-adrenocortical axis, and child psychopathology. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *59*(4), 327–346.
Landry, S. H., Smith, K. E., Swank, P. R., & Guttentag, C. (2008). A responsive parenting intervention: the optimal timing across early childhood for impacting maternal behaviors and child outcomes. *Developmental Psychology*, *44*(5), 1335–1353.
Levendosky, A. A., Bogat, G. A., Lonstein, J. S., Muzik, M. M., & Nuttall, A. (2021). A longitudinal prospective study examining the effects of the timing of prenatal stress on infant and child regulatory functioning: The Michigan Prenatal Stress Study protocol. *BMJ Open*, *11*(9), e054964. https://doi.org/10.1136/bmjopen-2021-054964
Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford Press.
Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data.* (2nd ed.) Wiley. https://doi.org/10.1002/9781119013563
Loop, L., Mouton, B., Brassart, E., & Roskam, I. (2017). The observation of child behavior during parent-child interaction: The psychometric properties of the Crowell Procedure. *Journal of Child and Family Studies*, *26*(4), 1040–1050. https://doi.org/10.1007/s10826-016-0625-0
Lunkenheimer, E., Busuito, A., Brown, K. M., & Skowron, E. A. (2018). Mother–child coregulation of parasympathetic processes differs by child maltreatment severity and subtype. *Child Maltreatment*, *23*(3), 211–220.
Manning, B., Harpole, A., Harriott, E., Postolowicz, K., & Norton, E. (2020). Taking language samples home: Feasibility, reliability, and validity of child language samples conducted remotely with video chat versus in-person. *Journal of Speech, Language, and Hearing Research : Jslhr*, *63*(12), 3982–3990. https://doi.org/10.1044/2020_JSLHR-20-00202
Michels, N., Sioen, I., Clays, E., De Buyzere, M., Ahrens, W., Huybrechts, I., Vanaelst, B., & De Henauw, S. (2013). Children's heart rate variability as stress indicator: Association with reported stress and cortisol. *Biological Psychology*, *94*(2), 433–440.
Nesselroade, J. R., & Baltes, P. B. (1979). *Longitudinal research in the study of behavior and development*. Academic Press. http://hdl.handle.net/21.11116/0000-0003-61EB-7
Nussenbaum, K., Scheuplein, M., Phaneuf, C., Evans, M., & Hartley, C. (2020). Moving developmental research online: comparing in-lab and web-based studies of model-based reinforcement learning. *Collabra: Psychology*, *6*(1), 17213. https://doi.org/10.1525/collabra.17213
Owen, M. (1992). *The NICHD study of early child care mother–infant interaction scales. Unpublished manuscript*. Timberlawn Psychiatric Research Foundation.
Rhodes, M., Rizzo, M., Foster-Hanson, E., Moty, K., Leshin, R., Wang, M. M., Benitez, J., & Ocampo, J. D. (2020). Advancing developmental science via unmoderated remote research with children. *Journal of Cognition and Development : official Journal of the Cognitive Development Society*, *21*(4), 477–493. https://doi.org/10.1080/15248372.2020.1797751
Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. https://doi.org/10.1093/biomet/63.3.581
Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley. https://doi.org/10.1002/9780470316696
Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *The Journal of Applied Psychology*, *85*(1), 112–118. https://doi.org/10.1037/0021-9010.85.1.112
Shadish, T. D., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
Ulfig, N., Setzer, M., & Bohl, J. (2003). Ontogeny of the human amygdala. *Annals of the New York Academy of Sciences*, *985*, 22–33. https://doi.org/10.1111/j.1749-6632.2003.tb07068.x
Weinstock, M. (2008). The long-term behavioural consequences of prenatal stress. *Neuroscience and Biobehavioral Reviews*, *32*(6), 1073–1086. https://doi.org/10.1016/j.neubiorev.2008.03.002
Wright, A. J. (2018). *Equivalence of remote, online administration and traditional, face-to-face administration of the Reynolds Intellectual Assessment Scales*. (2nd ed.) [White paper]. Retrieved from https://pages.presencelearning.com/rs/845-NEW-442/images/Content-PresenceLearning-Equivalence-of-Remote-Online-Administration-of-RIAS-2-White-Paper.pdf
Zeanah, C. H., Jr., Larrieu, J. A., Heller, S. S., & Valliere, J. (2000). Infant–parent relationship assessment. In C. H. Zeanah, Jr. (Ed.), *Handbook of infant mental health*. (pp. 222–235) The Guilford Press.

Appendix A

Instructions for Video Conferencing Using the Zoom Platform

'Could you tell me, do you plan to use a cell phone, tablet, or a laptop with a camera to take the recording of you and your baby?'

If phone, *'What kind of cell phone do you have?'*

If tablet, *'What kind of tablet is it?'*

If laptop with webcam, *'What kind of a browser do you use?'*

Offer correct directions below depending on device.

CELL PHONES:

- Iphone: Download the 'Zoom Cloud Meetings' app from the Apple App Store (the icon is a white camera on a blue background).

Link to send: https://apps.apple.com/us/app/zoom-cloud-meetings/id546505307

- Android Phone: Download the 'Zoom Cloud Meetings' app from the Google Play Store (the icon is a white camera on a blue background).

Link to send: https://play.google.com/store/apps/details?id=us.zoom.videomeetings

TABLETS:

- Kindle: Download 'Zoom Cloud Meetings' from the Amazon App Store

Link: https://www.amazon.com/Zoom-Video-Communications-Inc-Meetings/dp/B00B5L5JRM

- Android Tablet: Download the 'Zoom Cloud Meetings' app from the Google Play Store (the icon is a white camera on a blue background).

Link to send: https://play.google.com/store/apps/details?id=us.zoom.videomeetings

- I pad: Download the 'Zoom Cloud Meetings' app from the Apple App Store (the icon is a white camera on a blue background).

Link to send: https://apps.apple.com/us/app/zoom-cloud-meetings/id546505307

If participant has a LAPTOP/COMPUTER with webcam:

RA needs to schedule a Zoom meeting on the agreed upon date and time and enable 'join before host.' RA then needs to email the participant the zoom link and have them click the link and it should open in a browser.

- **GOOGLE CHROME (this is the ideal browser):** Once the link has been clicked it sends you to a window that says 'A download should start automatically in a few seconds' then a file should download to your computer. Click and run this file. Once ran, Zoom opened and asked to join with video, then click with audio. They can then close out of the zoom meeting and just click the link on the day the recording will take place.

Note: All other browsers are more complicated. See details below

- **FIREFOX:** Once the link has been clicked it should open a Firefox browser which takes them to an MSU zoom page. Browser should say if you see no download happening then click 'download and run zoom'. They will need to save file 'zoom.exe'. The file should save but may not automatically launch, so participant may have to navigate to their downloads and open the file to run it. Once that happens Zoom opened, and will ask to join with video, then have them click join with audio. They can then close out of the zoom meeting and just click the link on the day the recording will take place.
- **INTERNET EXPLORER:** Once the link has been clicked it should open an Internet Explorer browser which takes them to an MSU zoom page. If nothing prompts from browser, have them click 'download' and run Zoom". Browser should ask if they want to 'run or save or cancel'. Have them hit run, and it should download Zoom and open it. Have participant click join with video, then click to join with audio. They can then close out of the zoom meeting and just click the link on the day the recording will take place.
- **SAFARI:** Once the link has been clicked it should open a browser in Safari which takes them to general MSU zoom page. Have them click 'open zoom.us' if you see the system dialog. If nothing prompts from browser it will say "If you cannot download or run the application, join from your browser". Click "join from your browser" text. Then enter your name, then "Join", then it opens meeting.

To wrap up say *'Okay, now that we have that all set up, I am going to email you some information about the phone interview, if you have the chance, please look over this email and read the consent form so we can answer any questions you may have that day. Thanks so much and talk to you then!'*