Breakout Session 2: Track A

Enabling AI / ML Readiness and Modernization of Longitudinal Pregnancy and Cardiovascular Health Data: Lessons Learned

Dr. Rebecca McNeil
Senior Research Statistician, RTI International

nuMoM2b Heart Health Study

Enabling AI/ML Readiness and Modernization of Longitudinal Pregnancy and Cardiovascular Health Data

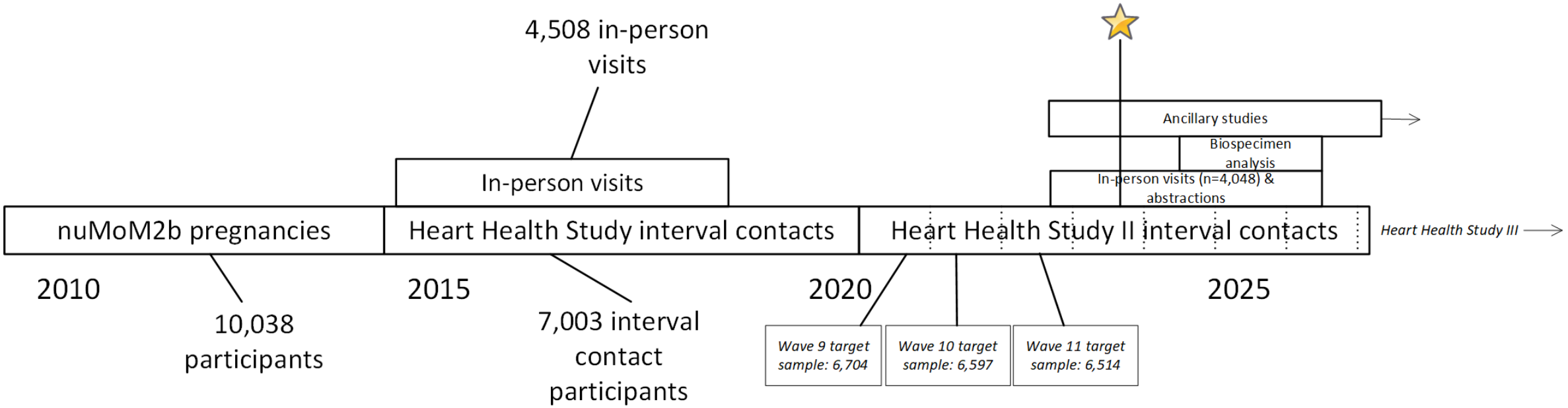
NOT-OD-22-067
Becky McNeil, PhD
RTI International
Contact PI, nuMoM2b-HHS2

NIH National Heart, Lung, and Blood Institute

RTI INTERNATIONAL

Project summary

Overview of study timeline



nuMoM2b

AIMS – Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-be

1. Determine maternal characteristics, including genetics, epigenetics, and physiological response to pregnancy as well as environmental **factors that influence and/or predict adverse pregnancy outcomes (APOs)**
2. Identify specific aspects of placental development and function that lead to APOs
3. Characterize genetic, growth, and developmental parameters of the fetus that are associated with APOs

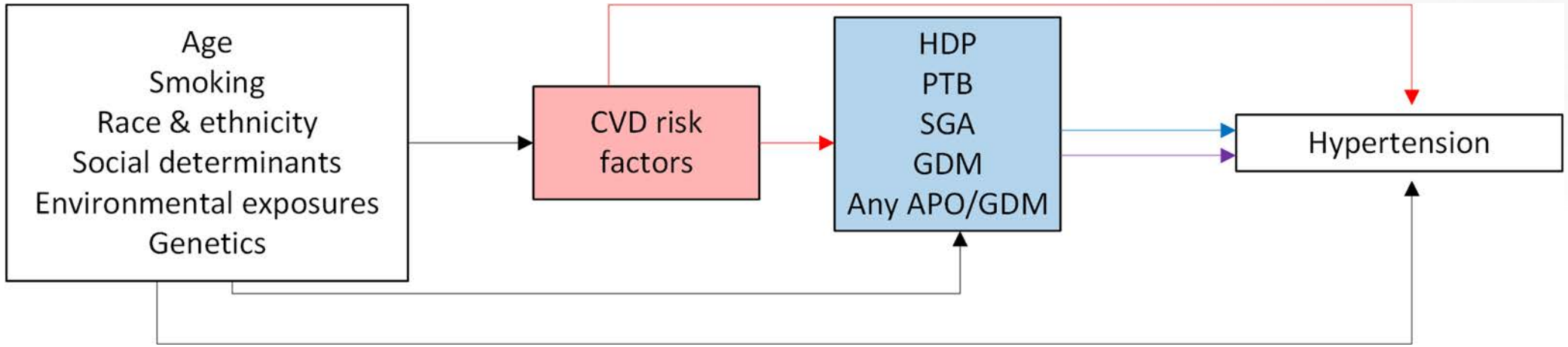
- Sponsor: NICHD
- Multi-site prospective cohort study
- 10,038 nulliparas with singleton pregnancies were enrolled at 6 – $13^{6/7}$ weeks gestation
  - Enrollment 2010–2013; deliveries 2010–2014
- Age at enrollment: 13–45 years
- Participants were followed through 14 days postpartum and newborn infants were followed through hospital discharge
- Study visits at up to 4 timepoints:
  - 6^{0} – 13^{6}
  - 16^{0} – 21^{6}
  - 22^{0} – 29^{6}
  - Delivery

nuMoM2b Heart Health Study

Pregnancy: a window to future cardiovascular health?

Prior studies: association between APOs (PTB, SGA, GDM) and maternal cardiovascular disease later in life

- Some APOs have underlying vascular mechanisms
- Is there a common cause?
- Is pregnancy a “stress test” that might reveal cardiovascular risk?



nuMoM2b Heart Health Study

nuMoM2b Heart Health Study (HHS1)

- Goal: Evaluate the association between APOs and maternal health later in life in an effort to develop future screening and preventive strategies.
- Funded jointly by NHLBI and NICHD
- Continued follow-up (2014-2020) of nuMoM2b participants who agreed to recontact, with:
  - Known pregnancy outcomes
  - Age 18 years or more at time of first recontact
  - (In-person visit) Not currently pregnant and at least 6 months postpartum
- Interval contacts (phone or online survey) every 6 months until in-person visit, then at 12-month intervals
- In-person visit at 2 or more years after nuMoM2b pregnancy
- In-home sleep breathing substudy after the in-person visit (for nuMoM2b sleep breathing study participants)
- Chart abstraction for subsequent pregnancies if APO or multiple gestation pregnancy was reported
- Chart review for reports of CV-related event or death
- Central laboratory analysis of biospecimens: lipids, glucose, HbA1c, insulin, hsCRP, NT-proBNP, urine albumin & creatinine

Domain	Instrument or Clinical Measure	Dataset by Study Visit					
		1st	2nd	3rd	D	IC	IP
Medical and sociodemographic history	Medical and Sociodemographic History	V1A*	V2A*	V3A*	V4A	Txx*	V5A,V5S*
	Lactation History					Txx*	
	Family Medical History		V2A				V5A
	Residential History	V1A*	V2A*	V3A*			V5A
	Medical Record Abstraction: Pregnancy/Neonatal				Cxx*	P5A-F*	P5A-F*
	Medical Record Abstractions: CVD events				CMA,CMB	CRX,CRM	CRX,CRM
Physical activity	Physical Activity Log (BRFSS)	V1A*	V2A*	V3A*			V5A
	Modifiable Activity Questionnaire (MAQ)						V5Q*
Nutrition	Modified Block FFQ	V1D*					V5D*
	Three Factor Eating Questionnaire - Disinhibition						V5P
Health literacy	Rapid Estimate of Adult Literacy (REALM-SF)		V2A				
Psychosocial	Experiences of Discrimination (EOD)		V2A				
	Reaction to Race	V1A					
	Multidimensional Social Support (MSPSS)	V1G					
	Pregnancy Experience Scale, Brief (PES-Brief)			V3J			
	Connor-Davidson Resilience Scale (CD-RISC)		V2I				
Stress & mental health	Perceived Stress Scale (PSS-10)	V1A		V3A			V5R
	Edinburgh Postnatal Depression Scale (EPDS)	V1C		V3C			V5C
	State-trait Anxiety Index (Trait) (STAI-T)	V1H					
Sleep	Sleep Breathing Substudy	V1K		V3K			V5N
	Sleep Patterns and Quality Substudy		V2M				
	WHI Insomnia Rating Scale	V1F/V1L		V3F/V3L			V5N
	Berlin Questionnaire for Sleep Apnea	V1F/V1L		V3F/V3L			V5O
	Epworth Sleepiness Scale, and Restless Legs	V1F/V1L		V3F/V3L			V5N
	Restless Legs Syndrome Diagnostic Criteria	V1F/V1L		V3F/V3L			V5N
	PROMIS Sleep Disturbance & Impairment						V5N
Clinical measures	Blood Pressure	V1B*	V2B*	V3B*			V5B*
	Pulse						V5B*
	Weight	V1B*	V2B*	V3B*			V5B*
	Height	V1B*					V5B*
	Waist, Hip, and Neck Circumference	V1B*					V5B*

* Ancillary datasets contain derived content from these forms and from biospecimen assays and electronically captured data

Overview of data collection across nuMoM2b & HHS1

Enabling AI/ML Readiness … Aims

1. To convert existing epidemiologic datasets to machine-readable structures with complete metadata, with modernized documentation, and to provide access to the resulting data and materials on BioData Catalyst.
2. To complete the preparation and submission of currently disparate multi-omics datasets to dbGaP and BioData Catalyst.
3. To expand awareness and improve usability of the AI/ML ready data by creating a template workflow for the application of ML tools and completing analyses for a clinical use case that will be made available as a public markdown project.

Challenges

Enabling AI/ML Readiness … Aim 1

Aim 1: To convert existing epidemiologic datasets to machine-readable structures with complete metadata, with modernized documentation, and to provide access to the resulting data and materials on BioData Catalyst.

- Machine-readability: Convert existing files (130+) from SAS datasets to JSON format
  - In existing data, missing data codings were applied inconsistently and many variables were not labeled. Extensive work to update prior to JSON conversion, still underway
- Harmonization: assign clinical ontologies to clinical characteristics and endpoints (obstetric and cardiovascular)
  - Deferred until addition progress made on machine-readability
- Documentation modernization: Convert current documentation (300+ page word document + supporting metadata) to file-specific datasheets
  - Proceeding well, but time-consuming

nuMOM2b Heart Health Study

Enabling AI/ML Readiness … Aim 2

Aim 2: To complete the preparation and submission of currently disparate multi-omics datasets to dbGaP and BioData Catalyst.

- Data sharing: submission of GWAS, exosome and plasma proteomics, WGS, and RNASeq data and/or datasheets for public access on BioData Catalyst
  - TOPMed data (WGS, RNAseq) delayed
  - Technical issues with dbGaP accessions delayed submission of GWAS data
- Data interoperability: support multi-omic analyses by performing common group comparisons and storing results under common metadata structures
  - Deferred start on this
- Data harmonization: create phenotype datasets using new metadata standards used by other major cardiovascular health cohort studies
  - Deferred start on this

Enabling AI/ML Readiness … Aim 3

Aim 3. To expand awareness and improve usability of the AI/ML ready data by creating a template workflow for the application of ML tools and completing analyses for a clinical use case that will be made available as a public markdown project.

All Aim 3 activities pending additional progress on JSON dataset creation and omics data submissions.

- Template workflow: create template for basic ML workflow and analysis using commonly used cohort datasets
- Data paper: submit a data paper describing the shared datasets and project information
- Use case: complete analysis for a clinically relevant application of the AI/ML ready datasets and make it available as a public markdown project

Future work

Future work

- Complete existing planned work
- Apply processes to new datasets to prepare them for public use
- Update / maintain new resources as appropriate

Thank you!