

Development and Evaluation of an Algorithm to Automatically Extract Delivery Episodes from Electronic Health Records

Silvia P. Canelón, PhD¹, Heather H. Burris, MD, MPH², Lisa D. Levine, MD, MSCE³, Mary Regina Boland, MA, MPhil, PhD, FAMIA^{1,4-5}

¹Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA

²Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA

³Maternal and Child Health Research Center, Department of Obstetrics and Gynecology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA

⁴Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA

⁵Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia

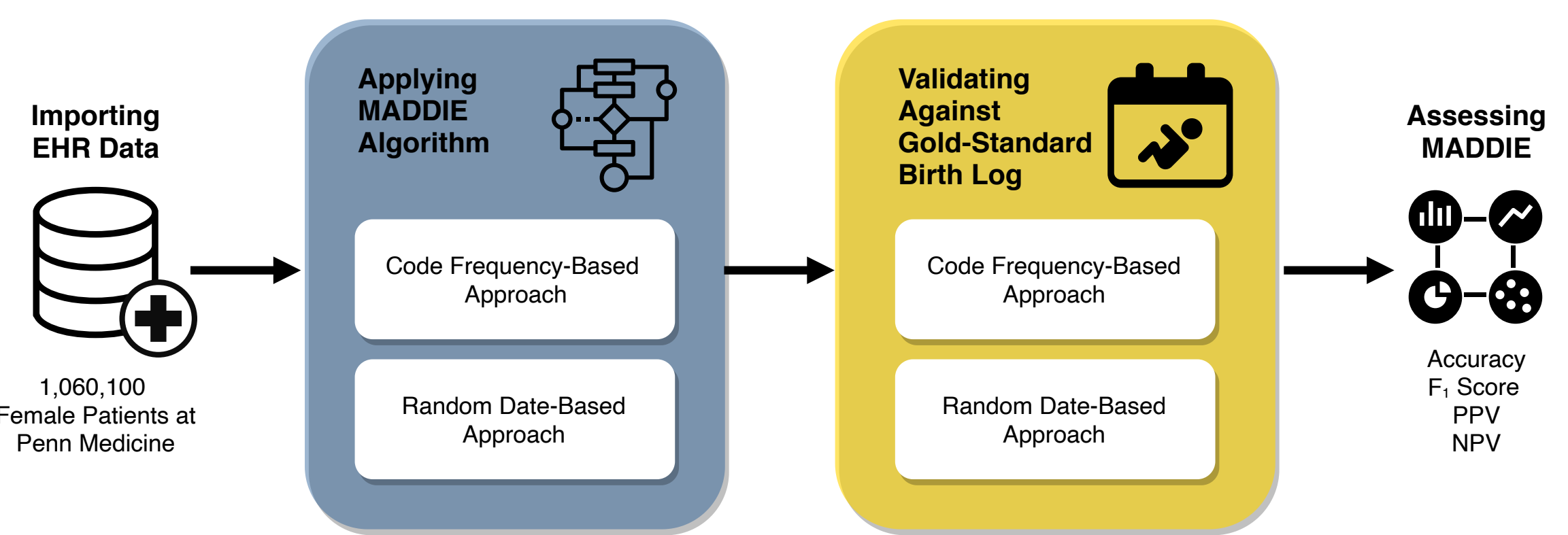
DEPARTMENT of
BIostatistics
EPIDEMIOLOGY &
INFORMATICS



MOTIVATION

- **Electronic health records (EHR)** contain rich information on a patient's medical history that can be used to extract pregnancy-specific details, enabling the study of health outcomes with **greater granularity at the pregnancy-level** rather than solely at the patient-level.
- Existing algorithms extract delivery episodes and dates from these databases using billing codes and have had to rely on **limited manual chart review for validation**.¹⁻⁴
- This study describes an algorithm designed to extract these details and **validated against an independent gold-standard birth log**.⁵

OVERVIEW



We developed an algorithm called **MADDIE: Method to Acquire Delivery Date Information from Electronic Health Records** that infers patient delivery dates and delivery-specific details with high accuracy.

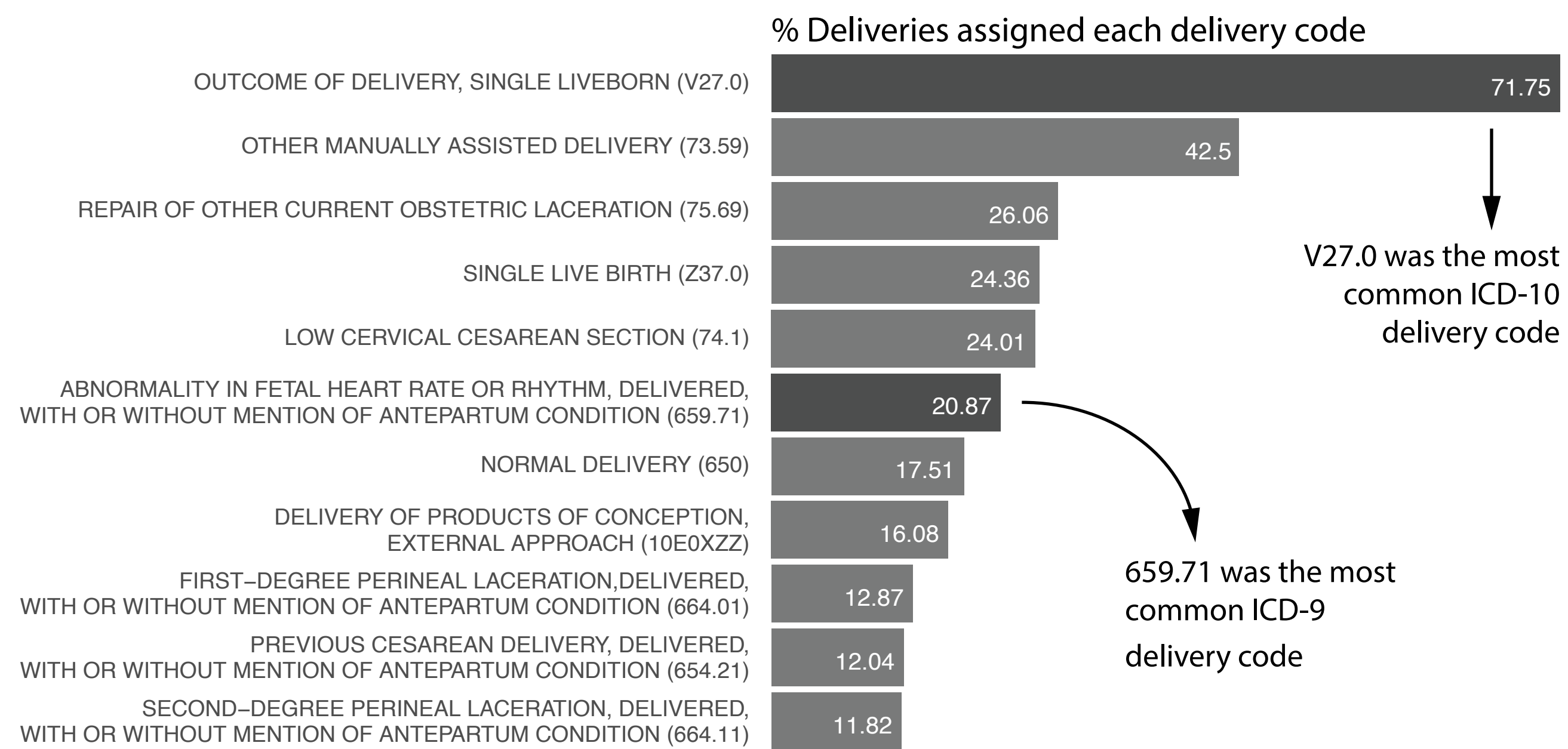
SUMMARY

- **MADDIE used EHR encounter dates** assigned a delivery code, the frequency of code usage, and the time differential between code assignments, to detect distinct deliveries **from 1,060,100 female patients** with visits to the **Penn Medicine** health system (2010-2017).
- **MADDIE identified 50,560 patients with 63,334 distinct deliveries**.
- MADDIE was found to be **98.6% accurate (F₁-score 92.1%)** when compared to the gold-standard birth log. The patient delivery date was on average **0.68 days earlier** than the true delivery date for patients with only **one delivery (±1.43 days)** and **0.52 days earlier** for patients with **more than one delivery episode (± 1.11 days)**.

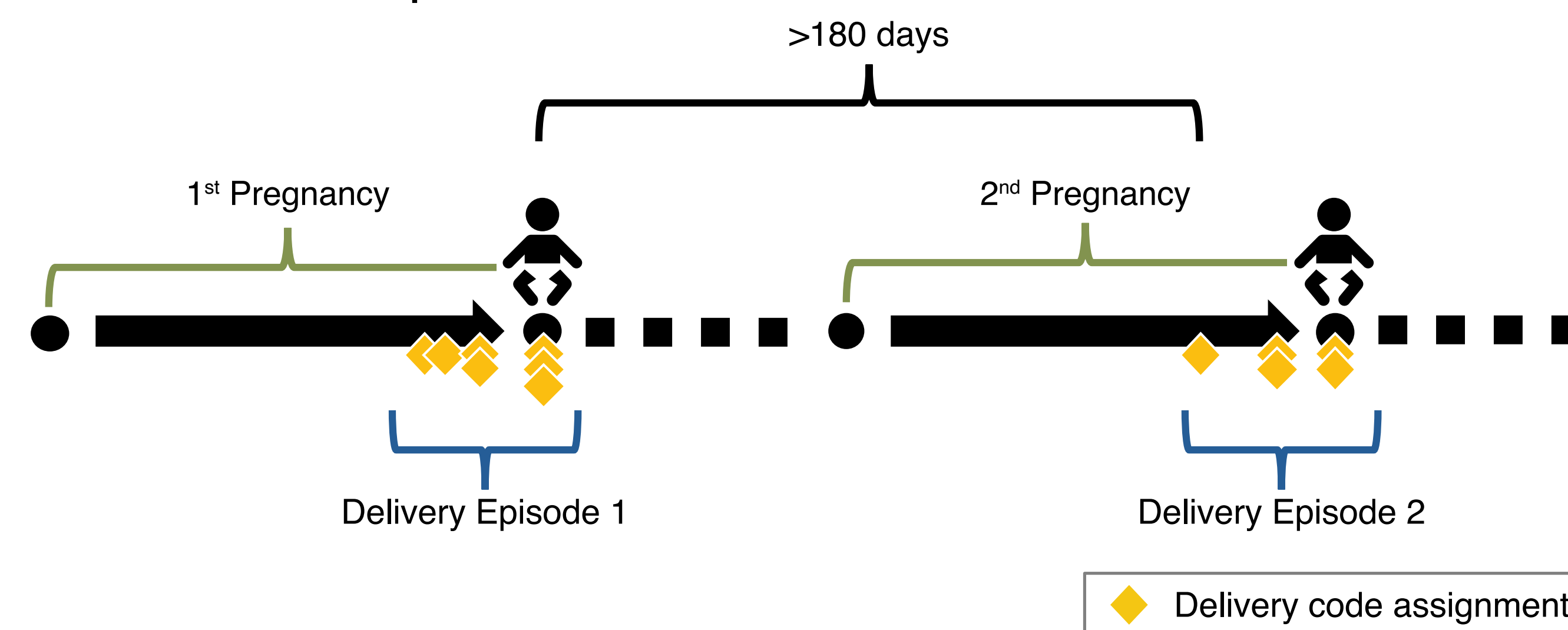
MADDIE: METHOD TO ACQUIRE DELIVERY DATE INFORMATION FROM ELECTRONIC HEALTH RECORDS

STEP 1. ICD version 9 (ICD-9) and version 10 (ICD-10) codes to identify 50,560 patients with delivery diagnoses or delivery procedures during any inpatient or outpatient clinic visit to Penn Medicine 2010-2017.

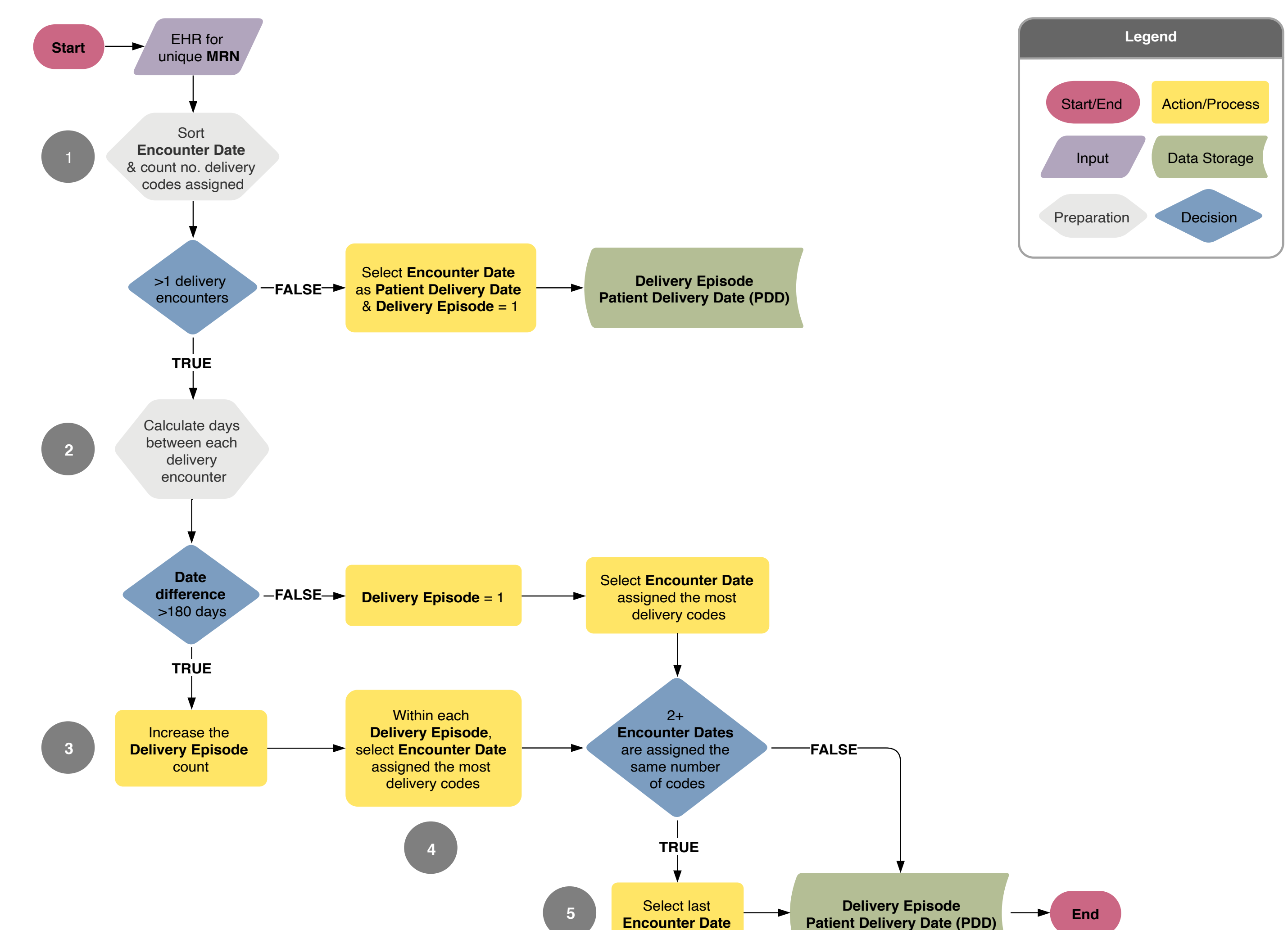
The most common ICD delivery codes assigned across all deliveries



STEP 2. In cases where the patient had a delivery code assigned only during one encounter, the encounter date became the delivery date. In more complex cases, where patients had a delivery code assigned on more than one date, those dates were grouped into a "delivery episode". Delivery code groupings with >180 days between them were considered two separate deliveries.



STEP 3. For each delivery episode, MADDIE computed the frequency of code usage and assumed the date assigned the most delivery codes was likely to be the delivery date. If two or more encounter dates had the same frequency of delivery codes, the last encounter date as selected as the patient delivery date (see validation and assessment).



MADDIE VALIDATION & ASSESSMENT

Validation against the gold-standard. MADDIE was validated against gold-standard birth log records available for only a subset of deliveries (25,676 patients). These records were independently kept by the Hospital of the University of Pennsylvania.

Validation Populations	Birth Log	MADDIE-Derived Cohort
Total population	Patients (%) 25676 (100)	Patients (%) 23271 (100)
Age (years), average:	28.3 ± 6.2	29.5 ± 6.1
Race/Ethnicity^a		
Black or African American	17193 (67)	15363 (66.0)
White	4595 (17.9)	4383 (18.8)
Asian	1614 (6.3)	1539 (6.6)
Other/Not Available ^b	1277 (4.9)	1113 (4.8)
Hispanic	1034 (4.0)	904 (3.9)
Native Hawaiian or other Pacific Islander	30 (0.1)	27 (0.1)
American Indian or Alaskan Native	15 (0.1)	14 (0.1)

^aRace/ethnicity descriptions are "non-Hispanic" unless otherwise indicated.
^bOther includes race description categories Other, Mixed, and Unknown.
Not Available refers to refers to Birth Log records without a linked encounter record.

To select delivery dates within delivery episodes, MADDIE assumed the date assigned the most delivery codes was likely to be the delivery date. This **code frequency-based approach** was compared with a random date-based approach and was found to be more accurate in inferring delivery dates when compared to the gold-standard birth log.

MRN	Delivery Date from MADDIE	Delivery Date from Birth Log	Difference	Category
112358	2014-02-05	2014-02-05	0 days	Match
112358	2014-02-05	2016-12-12	-1041 days	Mismatch
112358	2016-12-01	2014-02-05	1030 days	Mismatch
112358	2016-12-01	2016-12-12	-11 days	Match

Code frequency-based approach. For each patient (MRN), the differences between the algorithm-selected delivery dates and all potential birth log delivery dates were calculated. Dates with a difference ≤ 30 days were considered "matches". Agreement between MADDIE and the birth log was 99.9% for cases with just one delivery and 98.5% for more than one delivery.

MRN	Encounter Date	Code Frequency	Difference	Delivery Episode
112358	2014-01-18	1	0 days	1
112358	2014-01-23	1	5 days	1
112358	2014-01-30	2	7 days	1
112358	2014-02-05	3	8 days	1
112358	2016-11-14	1	1013 days	2
112358	2016-11-26	2	12 days	2
112358	2016-12-01	2	5 days	2

Performance in detecting patients within the birth log. To assess the overall performance of the MADDIE algorithm, we assessed its ability to accurately detect patients with deliveries occurring at the Hospital of the University of Pennsylvania (HUP).

MADDIE	Gold-Standard Birth Log		Positive Predictive Value	Accuracy
	Delivery Positive	Delivery Negative		
Delivery Positive	True Positive 23,001	False Positive 270	98.8% (Total: 23,271)	98.5%
Delivery Negative	False Negative 2,675	True Negative 176,896	Negative Predictive Value 98.5% (Total: 179,571)	F ₁ Score 92.1%
	Sensitivity 89.6% (Total: 25,676)	Specificity 99.8% (Total: 177,166)	Total Population 202,842 female patients at HUP	

CONCLUSIONS

MADDIE is the first algorithm to successfully infer patient delivery date information using only structured delivery codes and identify multiple deliveries per patient. It is also the first to validate the accuracy of the patient delivery date using an external goldstandard of known delivery dates as opposed to manual chart review of a sample.

The MADDIE algorithm facilitates population-based studies of pregnant patients that are commonly underrepresented in clinical research. It allows for greater granularity in studying pregnancy episode-specific as well as delivery-specific associations using EHR data because it has the ability to distinguish patients having several distinct deliveries in the same EHR or clinical records system.

Random date-based approach. Randomly selected an encounter date within each delivery episode rather than make use of a heuristic (i.e. code frequency). Agreement between MADDIE and the birth log using this approach was 99.5% for one delivery and 98.0% for two or more delivery cases.

Differences between randomly generated delivery date and matched birth log delivery date averaged across 100 random iterations

MADDIE identified 50,560 patients with 63,334 distinct deliveries. Along with age at the time of delivery, and the delivery date, MADDIE augments the EHR with delivery-specific details like the delivery number.

Penn Medicine Population	MADDIE-Derived Cohort
Total population	Patients (%) 50560 (100)
Age, years^a	Deliveries (%) 63334 (100)
< 18	1104 (2.2)
18-24	11062 (21.9)
25-34	28947 (57.25)
35-44	13296 (26.3)
45+	307 (0.6)
Average: 29.5 ± 6.1	317 (0.5)
Race/Ethnicity^a	
Black or African American	23777 (47.1)
White	17034 (33.7)
Asian	4031 (8.0)
Other/Not Available ^b	3305 (6.5)
Hispanic	2509 (5.0)
Native Hawaiian or other Pacific Islander	75 (0.1)
American Indian or Alaskan Native	61 (0.1)

REFERENCES

1. Manson, JM et al. Am J Epidemiol. 2001;154:180-7.
2. Hornbrook MC, et al. Health Serv Res. 2007;42:908-27.
3. Naleway AL et al. Vaccine. 2013;31:2898-903.
4. Matcho A et al. PLoS One. 2018;13:e0192033.
5. Canelón, SP et al. Int J Med Inform. 2020; j.ijmedinf.2020.104339

