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

Silvia P. Canelón, PhD[1] , Heather H. Burris, MD, MPH[2], Lisa D. Levine, MD, MSCE[3], Mary Regina Boland, MA, MPhil, PhD, FAMIA[1,4-5]

[1]Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA
[2]Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA
[3]Maternal and Child Health Research Center, Department of Obstetrics and Gynecology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA
[4]Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA
[5]Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia

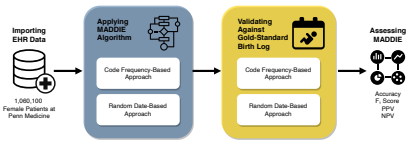
Perelman School of Medicine University of Pennsylvania

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



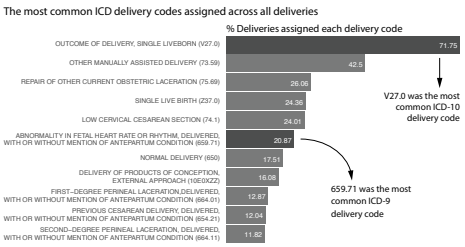
We developed an algorithm called **MADDIE: M**ethod to **A**cquire **D**elivery **D**ate **I**nformation from **E**lectronic Health Records that infers patient delivery dates and delivery-specific details with high accuracy.

SUMMARY

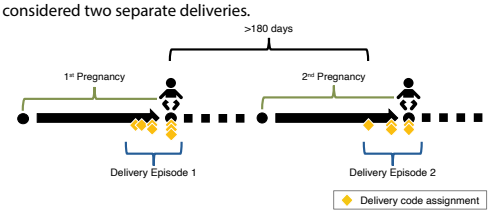
- **MADDIE used EHR encounter dates** assigned a delivery code, the frequency of code usage, and the time differential between code assignments, to detect distinct deliveries **from 1,060,100 female patients** with visits to the **Penn Medicine** health system (**2010-2017**).
- **MADDIE identified 50,560 patients with 63,334 distinct deliveries.**
- MADDIE was found to be **98.6% accurate (F_1-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

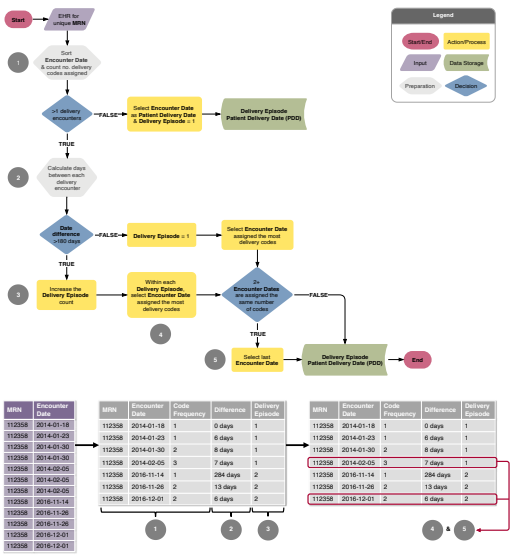
STEP 1. ICD version 9 (ICD-9) and version 10 (ICD-10) codes to identify 50,560 patients with delivery diagnoses or delivery procedures during any inpatient or outpatient clinic visit to Penn Medicine 2010-2017.



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.

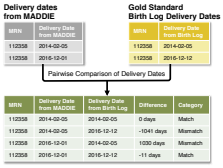
Validation Populations	Birth Log	MADDIE-Derived Cohort	
	Patients (%)	Patients (%)	Deliveries (%)
Total population	25676 (100)	23271 (100)	28427 (100)
Age (years), average:	28.3 ±6.2	29.5 ±6.1	
Race/Ethnicity[a]			
Black or African American	17193 (67)	15363 (66.0)	19004 (66.85)
White	4595 (17.9)	4383 (18.8)	5301 (18.65)
Asian	1614 (6.3)	1539 (6.6)	1812 (6.6)
Other/Not Available[b]	1277 (4.9)	1113 (4.8)	1312 (4.6)
Hispanic	1034 (4.0)	904 (3.9)	1044 (3.7)
Native Hawaiian or other Pacific Islander	30 (0.1)	27 (0.1)	34 (0.1)
American Indian or Alaskan Native	15 (0.1)	14 (0.1)	17 (0.1)

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

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).

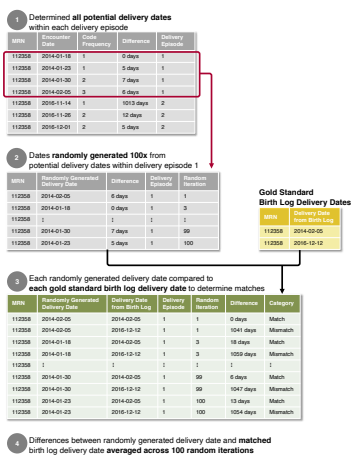
		Gold-Standard Birth Log			
		Delivery Positive	Delivery Negative		
MADDIE	Delivery Positive	True Positive	False Positive	Positive Predictive Value	Accuracy
		23,001	270	98.8% (Total: 23,271)	98.5%
	Delivery Negative	False Negative	Negative Predictive Value	F_1 Score	
		2,675	176,896	98.5% (Total: 179,571)	92.1%
		Sensitivity	Specificity	Total Population	
		89.6% (Total: 25,676)	99.8% (Total: 177,166)	202,842 female patients at HUP	

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.



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.

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.



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	
	Patients (%)	Deliveries (%)
Total population	50560 (100)	63334 (100)
Age, years[a]		
< 18	1104 (2.2)	1129 (1.8)
18-24	11062 (21.9)	12670 (20.0)
25-34	28947 (57.25)	34139 (53.9)
35-44	13296 (26.3)	15079 (23.8)
45+	307 (0.6)	317 (0.5)
Average: 29.5 ±6.1		
Race/Ethnicity[a]		
Black or African American	23777 (47.1)	29965 (47.3)
White	17034 (33.7)	21443 (33.9)
Asian	4031 (8.0)	4985 (7.9)
Other/Not Available[b]	3305 (6.5)	4073 (6.4)
Hispanic	2509 (5.0)	2993 (4.7)
Native Hawaiian or other Pacific Islander	75 (0.1)	94 (0.1)
American Indian or Alaskan Native	61 (0.1)	81 (0.1)

[a]Patient counts reflect the number of unique patients in each age bracket at the time of their delivery encounter. Patients with multiple deliveries are counted multiple times.
[a]Race/ethnicity descriptions are "non-Hispanic" unless otherwise indicated.
[b]Other includes race description categories Other, Mixed, and Unknown.

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.

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**

Penn University of Pennsylvania