Physicians Treating Physicians: Information and Incentives in Childbirth

Erin M. Johnson and M. Marit Rehavi

ONLINE APPENDIX

I. California Data

A. Identifying Births to Physicians

Physician-patients were identified by a probabilistic merge of the California confidential Linked Patient Discharge Data-Birth Cohort File (PDD-Birth) from the California Office of Statewide Health Planning and Development (OSHPD) with a dataset of physicians practicing in California. An OSHPD contractor performed the merge for this project. We provided the contractor with a file of physicians practicing in California and worked with her to develop the merge process. The contractor was given access to the full name of the mother (first, last and maiden) in the PDD-Birth data (OSHPD does not release this information to researchers).[1] Only records with female gender or unknown gender were used in the merge.

The primary physician file is the California Medical Board physician licensure database. It includes the full name, zip code, and year of medical school graduation for all physicians with active California medical licenses during the sample period. We augmented this file with data purchased from BrightPath Marketing. The BrightPath Marketing data includes month and year of birth, physician specialty and gender, but was only available for 16% of the physician licenses.

The merge was undertaken in 4 blocks. First name matches were considered first; then maiden name matches; then last name matches; and then matches on year and month of birth.[2] Agreement weights were calculated for 5 variables in the merge process: first name, last name, year of birth, month of birth and commuting zone. First and last name were available in both databases. Because mothers could have multiple last names, matches were considered using any of the mother's last names and her maiden name from the VS data. The full match weight was applied in cases of exact match, and the Jaro and Bigram Comparators

[1]Only the last name of the father was available in the VS data.

[2]It was not computationally feasible to compare all potential pairs. Blocking on commuting zone and birth year range was also intractable. Also for tractability, very common names were excluded in the first three blocks. These were names with frequencies greater than 1,000 in the vital statistics data or greater than 300 in the physician data.

were used to account for "close" matches.[3] The exact year of birth was only available in 16% of physician records. For the remaining 84% of records, an 8-year range of birth years was imputed from the medical school graduation year.[4] For month of birth only exact matches were assigned the full agreement weight. The final matching variable, the commuting zone, was calculated from the zip code in each dataset to account for moves and/or differences in work and home addresses.[5] The full agreement weight was applied if the commuting zone in the physician file matched at least one of the commuting zones in the VS data.[6]

The confidential VS data included 412,376 unique mothers who were at least 25 years old and had either postgraduate education or unknown education at some point during the study period. The physician data included 182,344 physicians of female or unknown gender with a unique combination of matching variables. The probabilistic record linkage identified 8,922 physician moms as matches using a probabilistic match weight cutoff of 0.4 (of these 3,286 are first births to mothers in our age range). 36% of identified physicians were exact matches on month and year of birth. The mean match probability for physicians is 0.63.

The number of physicians we identify is not far from what one would expect based on the birth rate of physicians in Texas. There were approximately 16,700 female physicians between the ages of 16 and 49 living in CA at the time of the 2000 US Census.[7] Only some female physicians will have a child and even fewer will have their first child during our sample. We impute physician birthrates by combing the 2000 US Census data for Texas with our data. The Texas birth data have self-reported occupation and therefore did not require a match on physician name (and presumably have less measurement error in the physician identifier). The observed births to female physicians living in Texas suggest birthrates of 94.47 and 8.44 per 1,000 female physicians age 16-39 and 40-49, respectively (these are the most detailed age categories we could get for the denominator, the number of female physicians living in Texas). This suggests we should expect to observe 987 births to physician-mothers in California in 2000. We identify 793 births to physician-mothers in California in 2000, or 80% of the number we would expect. Repeating the exercise for first births suggests that we observe 68% of the first births to physician-mothers that we would expect to see in California given the rates of first-births per female physician in Texas. Over 80% of first births to physician-mothers in Texas occur to mothers between the ages of 27 and 36. There are more older female physicians living in California (than Texas) and we

[3]A comparison was deemed a match if the maximum of the Jaro and Bigram Comparator was over 0.7, and in this case the comparator value was used to prorate the agreement weight.

[4]For the physicians with year of birth and year of medical school graduation in the physician file, 88% of birth years fell within the imputed 8-year range of birth years.

[5]This was done using the Census zip code to commuting zone crosswalk derived from 1990 commuting patterns.

[6]The U- and M-probabilities are available on request. The highest U-probability was 0.106 for year of birth and the lowest M-probability was 0.991 for year of birth.

[7]This is the finest age range publicly available for this variable at the state level in the Census. Since there are unlikely to be physicians under 24 years of age, the numbers should be comparable with our sample.

suspect the discrepancy between the hit rates for first and all births is at least partly due to our inability to age adjust across the states at a finer level.

The match identifiers were then merged onto the full VS dataset. Births taking place in hospitals were then linked to the hospital discharge records for both mothers and babies for the 9 months prior through 1 year after the birth date. The final file provided to us included the usual VS-PDD data elements as well as the physician identifier and several indicators providing information on the merge matching process for physicians.

Classification of Delivery Method. — The final dataset provides two different methods for determining the delivery method of births. The birth certificate in California contains information on the delivery method. C-sections are also coded on the hospital discharge record associated with each delivery. Specifically, in the VS data C-sections, vaginal deliveries and surgical vaginal deliveries are coded. In 2005, when California switched to a new birth certificate form, an indicator for whether a trial of labor was attempted prior to C-section was also included on the birth certificate. While this would allow classification of C-sections into scheduled and unscheduled solely using the VS data (and not the PDD) in 2005, we chose to classify deliveries as C-sections using the ICD-9-CM procedure codes in the PDD in all years (any delivery with a procedure code of 74 was classified as a C-section). We then classify scheduled and unscheduled C-sections following the methods of Henry et al. (1995) and Gregory et al. (2002). This method uses diagnosis codes indicating a trial of labor to classify C-sections and was superior to classification using the VS trial of labor field in 2005.[8]

Maternal and Infant Health Outcomes. — The indicators for maternal and infant health outcomes were constructed from the diagnosis codes in the PDD. Maternal morbidity measures with corresponding ICD-9-CM codes in parenthesis are: 3rd and 4th degree lacerations, the more serious of the tears associated with vaginal delivery (664.2 or 664.3), post-partum hemorrhage (666) and infection (including pyrexia, generalized infection and major infection: 672, 659.2, 659.3, 670.3). The infant morbidity measures with corresponding ICD-9-CM codes in parenthesis are: meconium (779.84, 770.11, 770.12, 763.84); respiratory assistance (including oxygen therapy and mechanical ventilation: 93.96 and 93.90), intubation (96.04), infection (771), and trauma (all trauma to the infant excluding minor and relatively common scalp lacerations: 767 excluding 767.1). Respiratory assistance and intubation are procedures, not diagnoses. Obstetric wound complications and anesthesia complications (in mothers and infants) were also observable in the PDD using ICD-9-CM codes but occurred in less than 1% of the sample and were therefore excluded from the analysis. See Table A.1 for summary statistics of these measures.

[8]Trial of labor appears to be substantially under-reported in the VS data. Many observations containing diagnosis codes that indicate labor was tried did not report attempting labor in the VS data.

Additional Treatment Margins. — Indicators for the induction of labor along with the use of forceps and vacuum extraction are available in both the Vital Statistics and PDD data. We use the ICD-9-CM procedure codes from the PDD data in order to avoid concerns about the under-reporting of procedures in the VS data. The procedure and corresponding ICD-9-CM codes are: induction (73.1, 73.4), vacuum (72.7), and forceps (72.0-4). See Table A.1 for summary statistics.

Hospital Travel Time. — Travel distance and travel time to the delivery hospital and the closest HMO and non-HMO hospital were calculated as follows. Mothers' residence location was coded as the centroid of the maternal 5-digit zip code. Hospital geocodes were obtained from hospital addresses. Driving distance and time from the maternal zip code centroid to each hospital was then calculated using the Google Distance Matrix API. Mothers were determined to have delivered at the closest hospital if their delivery hospital is the one that minimizes driving time out of those visited by other mothers' in the zip code. See Table A.1 for summary statistics.

B. Texas Data

The Texas data come from the confidential VS file of the Texas State Department of Health Statistics. The data contain the birth certificate record for every birth registered in Texas from 1995-2003 and 2005-2007. The 2004 data was not used because the hospital identifier was not available. The file includes maternal and paternal demographic information, including self-reported industry and occupation for both. Additionally, the data includes maternal pregnancy history, pregnancy risk factors and delivery complications, and the birth outcome, including the method of delivery. The confidential files additionally include a hospital identifier for all hospital births.[9] A new birth certificate form was introduced in Texas in 2005, making additional variables available from 2005-2007. These include the name of the attending (confidential, but made available to us), and a variable indicating whether labor was tried for C-section deliveries.[10] The Texas VS record is not linked to the hospital discharge record. This means we cannot reliably split C-sections into scheduled and unscheduled categories. It also means we only observe risk factors included on the birth certificate form, not those appearing in the discharge record diagnosis codes.

Identifying Births to Physicians. — In Texas we identify physicians using the occupation field from the birth certificate form. We used a 3-step process to

[9]Due to the introduction of a new birth certificate form in Texas in 2005, the numeric hospital identifier is not consistent across years. Hospitals were linked across years using the name, which is available both before and after 2005.

[10]We found the quality of the trial of labor field to be suspect and so this is not used in the analysis.

categorize occupations as physician or non-physician occupations. We first categorized all occupation entries appearing 100 or more times over the sample period. Next, for less common occupation entries, we categorized any entry including the text strings "med" or "phys." Finally, we categorized all entries for individuals with a doctorate after 2005.[11] Because this field is available for both mothers and fathers, we can use the Texas sample to study treatment of physician-fathers as well as physician-mothers. We also identified nurses and the self-employed using the first step of the method. Finally, we merged the Texas Medical Board physician licensure file, which includes medical specialty, to the Texas Vital Statistics data.

Sample. — There were 4,419,892 registered births in Texas in this period. We restrict the analysis to births taking place in a hospital in Texas (dropping 72,792 observations) and to singleton births over 20 weeks gestation (dropping 121,655 observations). Given the time necessary to complete medical school, there are virtually no physicians in their early twenties. The sample is therefore further restricted to the 2,623,090 mothers at least 24 years of age and 50 years of age or younger. We further exclude observations with missing maternal age, missing maternal zip code, missing gestational age, or missing birthweight (90,663 observations). Finally, we restrict our analysis to the 720,487 first births and then to reduce concerns about comparability between physicians and non-physicians to the 372,345 parents with at least one college degree. Of these roughly 2,619 are families with physician mothers, 5,905 are families with physician fathers, and 1,472 are families with physician fathers and mothers.

Summary Statistics. — Table A.1 summarizes the independent variables used in the Texas analysis. The top panel displays means and standard deviations of parental demographics for physician parents, broken out according to the identity of the physician (physician-mothers, physician-fathers and families with two physician parents), and non-physicians. As in California, physician-parents are slightly older, less likely to be hispanic, and live in zip codes with higher income per capita.

The second panel of Table A.1 summarizes information on the infant. As in California, physician-mothers are slightly more likely to deliver before full-term and slightly less likely to deliver post-dates (42 or more weeks gestation). Physician-mothers are also more likely to give birth to babies that are low or very-low birth weight.

The last panel in Table A.1 summarizes clinical risk factors that potentially affect the C-section decision. While the risk factors for the California sample are derived from ICD-9-CM codes, the Texas risk factors come from the birth certificate survey. The risk factors available for comparison are limited, but

[11]Education is not categorized into masters/doctorate before 2005.

physician-parents do appear to have lower rates of diabetes and smoking than non-physicians, and physician mothers may have infants with higher rates of congenital anomalies.

References

Gregory, Kimberly D., Lisa M. Korst, Jeffrey A. Gornbein, and Lawrence D. Platt, "Using Administrative Data to Identify Indications for Elective Primary Cesarean Delivery," *Health Services Research,* 37 (2002), 1387-1401.

Henry, Olivia A., Kimberly D. Gregory, Calvin J. Hobel, and Lawrence D. Platt., "Using ICD-9 Codes to Identify Indications for Primary and Repeat Cesarean Sections: Agreement With Clinical Records," *American Journal of Public Health,* 85 (1995), 1143-1145.

Table A.1—Summary Statistics: Other Measures

	Non-HMO Hospitals				HMO Hospitals			
	Physicians		Non-Physicians		Physicians		Non-Physicians	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Maternal morbidity (%):								
Laceration	7.63	[26.55]	8.44	[27.80]	10.96	[31.27]	11.68	[32.12]
Post-partum hemorrhage	2.86	[16.66]	2.81	[16.52]	3.08	[17.29]	4.92	[21.64]
Infection	3.36	[18.03]	4.37	[20.44]	5.96	[23.70]	5.35	[22.51]
Infant morbidity (%):								
Meconium	3.83	[19.20]	4.17	[19.99]	2.11	[14.40]	3.62	[18.69]
Respiratory assistance	2.24	[14.80]	2.48	[15.57]	3.46	[18.30]	4.26	[20.19]
Intubation	2.38	[15.26]	2.48	[15.55]	1.73	[13.05]	2.80	[16.50]
Infection	1.84	[13.46]	2.13	[14.45]	0.77	[8.75]	1.22	[10.97]
Trauma	0.94	[9.65]	1.28	[11.23]	0.77	[8.75]	1.09	[10.38]
Treatment quantity (%):								
Induction	17.03	[37.59]	15.68	[36.36]	21.35	[41.01]	16.29	[36.93]
Vacuum extraction	15.62	[36.31]	16.80	[37.38]	12.69	[33.32]	11.66	[32.10]
Forceps	2.75	[16.35]	2.17	[14.59]	1.54	[12.32]	1.18	[10.78]
Hospital charges	17,999	[37,429]	19,119	[53,248]	-	-	-	-
Travel Time:								
Deliver at closest hospital (%)	50.93	[263.73]	53.87	[266.46]	20.72	[17.61]	20.02	[35.72]
Driving time to delivery hosp (mins.)	12.25	[96.68]	13.25	[78.62]	9.29	[7.00]	10.04	[32.24]
Driving time to closest HMO hosp	19.73	[19.82]	22.72	[24.50]	17.13	[16.08]	17.82	[33.51]
Driving time to closest non-HMO hosp	12.45	[96.68]	13.50	[78.69]	9.82	[7.23]	9.99	[14.49]
Observations	2,766		494,077		520		85,165	

Table contains means and standard deviations. Maternal and infant morbidity measures are coded from ICD-9-CM codes. HMO-owned hospitals are not required to report hospital charges.

Table A.2—Summary Statistics: Texas

	Physicians						Non-Physicians	
	Moms		Dads		Both			
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Demographics:								
Age	32.32	[3.87]	31.09	[3.81]	32.07	[3.32]	30.28	[3.99]
Mother's education (%)								
Some college	0	[0]	10.94	[31.22]	0	[0]	11.19	[31.53]
College graduate	0	[0]	41.24	[49.23]	0	[0]	51.35	[49.98]
High education	100	[0]	42.98	[49.51]	100	[0]	32.40	[46.80]
Father's education (%)								
Some college	9.36	[29.13]	0	[0]	0	[0]	15.00	[35.70]
College graduate	31.23	[46.35]	0	[0]	0	[0]	41.73	[49.31]
High education	52.50	[49.95]	100	[0]	100	[0]	32.24	[46.74]
Mother's race								
Black	9.47	[29.28]	3.73	[18.94]	3.60	[18.64]	7.44	[26.24]
Hispanic	11.38	[31.76]	13.40	[34.06]	11.96	[32.46]	16.36	[36.99]
Other race	22.18	[41.56]	19.17	[39.37]	28.60	[45.20]	9.86	[29.82]
Zipcode income	29,763	[12,737]	28,901	[13,042]	31,223	[14,272]	25,308	[10,480]
Married	95.91	[19.80]	97.75	[14.84]	99.05	[9.71]	93.23	[25.12]
Mother working	100	[0]	72.73	[44.54]	100	[0]	81.44	[38.88]
Father working	96.22	[19.08]	100	[0]	100	[0]	96.79	[17.62]
Infant information:								
Female	48.03	[49.97]	48.64	[49.99]	50.14	[50.02]	48.67	[49.98]
Very early term (20-36 weeks)	9.13	[28.80]	7.55	[26.43]	7.20	[25.86]	8.18	[27.41]
Early term (37-39 weeks)	31.92	[46.63]	27.54	[44.67]	31.18	[46.34]	26.18	[43.96]
Post-dates (≤42 weeks)	6.19	[24.09]	7.67	[26.62]	6.45	[24.58]	9.15	[28.83]
Very low birth weight	1.15	[10.64]	0.66	[8.10]	0.88	[9.36]	0.98	[9.85]
Low birth weight	6.22	[24.16]	5.15	[22.10]	5.84	[23.46]	4.90	[21.58]
High birth weight	5.69	[23.17]	5.44	[22.67]	4.69	[21.14]	8.26	[27.53]
Prenatal care	98.70	[11.32]	98.27	[13.03]	98.44	[12.41]	98.60	[11.76]
Risk factors:								
Malpositioned fetus	6.99	[25.50]	5.93	[23.62]	6.73	[25.05]	6.60	[24.83]
Diabetes	2.52	[15.68]	2.73	[16.29]	2.45	[15.45]	3.21	[17.63]
Eclampsia	0.31	[5.52]	0.15	[3.90]	0.14	[3.69]	0.26	[5.07]
Smoking	0.27	[5.16]	0.56	[7.45]	0.20	[4.51]	1.46	[12.00]
Hypertension / pre-eclampsia	6.53	[24.71]	4.47	[20.67]	4.82	[21.43]	6.29	[24.29]
Congenital anomaly	0.15	[3.90]	0.017	[1.30]	0.068	[2.61]	0.088	[2.96]
Observations	2,619		5,905		1,472		362,349	

Table contains means and standard deviations of independent variables used in the empirical analysis. The sample is described in the Data Appendix.