Published in final edited form as:

Am J Perinatol. 2015 April ; 32(5): 475–480. doi:10.1055/s-0034-1390351.

The utility of ICD9-CM codes in identifying induction of labor

Lisa D. Levine, MD, MSCE^{1,*}, Meghana Limaye, MD¹, and Sindhu K. Srinivas, MD, MSCE¹

¹ University of Pennsylvania Perelman School of Medicine, Maternal and Child Health Research Program, Department of Obstetrics and Gynecology, Philadelphia, PA, USA

Abstract

Background—Data evaluating the accuracy of ICD9-CM codes in identifying inductions are limited. Our objective was to examine the test characteristics of ICD9-CM coding for induction of labor and to identify differences between those captured by coding and those not.

Methods—We performed a retrospective cohort study of ICD9-CM codes in identifying charts of induced women at our institution from 2005-2009. Review of the medical record was the gold standard. Characteristics of the charts were compared using Mann-Whitney U tests and chi-square tests where appropriate.

Results—3263 women were included, 708 with ICD9-CM coding for induction (screen positive). 422 women were randomly sampled from those not coded as induction (screen negative). The sensitivity of ICD9-CM coding for induction was 51.4%, specificity 98.8%, PPV 96.6%, NPV 74.7%. False negative charts (25%) were more likely to be women induced for premature rupture of membranes (40% vs. 8%, $p<0.001$) or with oxytocin (51% vs. 33%, $p<0.001$) when compared to screen positive charts.

Conclusions—It is reassuring that 97% of charts coded for induction by ICD9-CM codes are, in fact, patients that were induced. With this degree of accuracy, we can be confident that charts coded as induction are unlikely to be miscoded.

Keywords

Accuracy; Coding; ICD9; Induction of labor

Introduction

Elective delivery prior to 39 weeks gestation has been identified as an important quality measure in obstetric care¹. Specifically, the Joint Commission has identified non-medically indicated deliveries between 37-38 6/7 weeks, including those who undergo an induction of labor, as one of the four core perinatal metrics². In order to comply with the Joint Commission, hospitals are now required to abstract chart information to assess their compliance with this metric.

*Corresponding author: Lisa D. Levine MD, MSCE: Hospital of the University of Pennsylvania, Department of Obstetrics and Gynecology, 3400 Spruce Street, 2000 Courtyard, Philadelphia, PA 19104. Telephone: 516-456-6427, Fax: 215-349-5625, lisa.obgyndoc@gmail.com.

Disclosures: The authors report no conflict of interest

Current limitations exist in evaluating quality metrics in obstetrics, such as induction of labor, including the lack of available nationally collected data and the perception of inadequately coded administrative data¹. These limitations highlight the need to evaluate individual procedures to determine if current discharge data can be used to assess quality of care and to offer support to improve discharge data.

Induction of labor is often identified in discharge data and retrospective studies using International Classification of Disease, 9th Edition-*Clinical Modification* (ICD9-CM) codes. The ICD9-CM coding system is the official system in the United States of assigning codes to both diagnoses and procedures that are associated with hospital utilization³. Although initially intended for utilization as statistical tracking measures, ICD9-CM codes are now used for billing, measurement of health services utilization, reimbursement from insurers, and identification of patient charts for research purposes⁴.

The use of ICD9-CM codes to correctly identify diagnoses and procedures has been studied in the obstetric literature for cesarean deliveries and pregnancy complications with variations in the sensitivity and predictive abilities⁵⁻¹⁰. Data specifically evaluating the accuracy in identifying charts of women that have undergone an induction of labor are limited¹¹⁻¹³. Induction of labor is one of the most common obstetric procedures in the United States with more than 20% of pregnant women undergoing an induction¹⁴. With the increasing frequency of this obstetric procedure and the identification of this procedure as a potential quality metric when utilized for non-medical indications, it is essential to continue to evaluate outcomes related to induction and to evaluate the accuracy of the current practices used to identify them.

Therefore, the objective of our study is to examine the test characteristics of ICD9-CM coding in hospital discharge data in identifying the charts of women who underwent an induction of labor and to identify the characteristic differences between those identified by coding and those not. Our hypothesis is that ICD9-CM coding accurately identifies the charts of women that have undergone an induction of labor; however, the charts of women who underwent an induction of labor in the absence of a cervical ripening agent are more likely to be missed by ICD9-CM coding.

Materials and Methods

This was a planned secondary analysis of a large retrospective cohort study evaluating pregnancy outcomes among women with a history of induction of labor compared to those with a history of spontaneous labor from 2005-2009¹⁵. Approval from the Institutional Review Board was obtained prior to the parent study.

This study used hospital discharge data to evaluate the test characteristics of ICD9-CM procedural codes in identifying the charts of women that underwent an induction of labor. The ICD9-CM codes studied included: 73.01 (induction of labor by artificial rupture of membranes), 73.1 (surgical induction of labor), and 73.4 (medical induction of labor). In our hospital, chart abstractors that collect hospital discharge data are trained to identify and code

a chart as an induction of labor if the provider has specifically documented that this procedure was being performed.

All charts from the parent study that were identified as an induction of labor by ICD9-CM codes were included. These women were considered the “screen positive” group. Detailed medical record review and chart abstraction was performed by two trained obstetricians to confirm that the charts of the women identified as an induction by ICD9-CM codes did in fact undergo an induction of labor as per our strict definition. This detailed review of the complete medical record was considered the gold standard. We defined induction of labor as (1) the use of any cervical ripening agent (prostaglandin or cervical foley), (2) artificial rupture of membranes or oxytocin use in the setting of contractions with cervical dilation <4 cm, or (3) artificial rupture of membranes or oxytocin use in the absence of contractions if cervical dilation was ≥ 4 cm. The charts of the women that met our definition were considered “true positives.” The charts of women that were coded as induction by ICD9-CM codes but did not undergo an induction of labor based on chart review were considered “false positives.”

All charts from the parent study that were not identified as an induction by ICD9-CM codes were considered “screen negative.” We randomly sampled approximately 20% of these screen negative charts as predetermined by the parent study¹⁵. Of the screen negative charts reviewed, charts that did not have an induction were considered “true negatives” while charts that did have an induction were “false negatives.”

The two trained obstetricians performing the chart reviews were aware of whether or not the charts of the women were in the screen positive or screen negative group.

Our analysis occurred in two parts. First, we calculated test characteristics of ICD9-CM codes in identifying charts of women who underwent an induction of labor. We estimated the sensitivity (the probability of being coded as an induction if the woman truly underwent an induction of labor), specificity (the probability of not being coded as an induction if the woman did not undergo an induction of labor), positive predictive values (PPV) (the probability that a woman had an induction of labor if she was coded as an induction), and negative predictive value (NPV) (the probability that a woman did not have an induction of labor if she was not coded as an induction). We assumed there were no systematic differences between the randomly sampled screen negative charts ($n=451$) and the entire screen negative cohort ($n=2555$). We therefore applied the NPV from the random sample of screen negative charts to the entire screen negative cohort to determine the sensitivity and specificity.

The second part of the analysis compared demographic characteristics for screen positive charts identified as inductions by ICD9-CM coding and the false negative charts (the screen negative charts that were ultimately found to be inductions on chart review). This analysis was performed to identify systematic differences between those charts captured with ICD9-CM coding for induction and those missed. Mann-Whitney U tests were used to compare non-parametric data and chi-square tests were used to compare categorical variables. Data

analysis was performed using STATA 12.0 for Windows (STATA Corporation, College Station, TX). Statistical significance was set at $p < 0.05$.

Results

Overall, there were 3263 charts identified for the parent study. All charts that had ICD9-CM codes for induction were reviewed ($n=708$, screen positive) and included in this study's analysis. Of those charts that were not coded as induction ($n=2555$, screen negative), 451 were then randomly sampled as noted in the primary study¹⁵ and included in our analysis (Figure 1).

Table 1 displays the test characteristics of ICD9-CM codes in identifying charts of women undergoing an induction of labor. Assuming there were no systematic differences between the 451 screen negative charts that were randomly sampled and the entire cohort of 2555 screen negative charts, the NPV and false negative rates (74.7% and 25.3%) were applied to the entire 2555 cohort of screen negative women to determine the sensitivity and specificity of ICD9-CM coding in identifying inductions (Table 1).

For our next analysis, we compared demographic characteristics of the screen positive charts (those identified as induction by ICD9-CM coding) and the false negative charts (the screen negative charts identified as inductions on chart review). These groups were compared to identify systematic differences between those charts captured with ICD9-CM coding for induction and those missed. The demographic information for these two groups, as well as the information for the entire cohort, is presented in Table 2. Overall, there was no difference in maternal age, body mass index, parity, or mode of delivery between the two groups. However, this table highlights the differences in both indication for induction and method of induction between the screen positive and false negative charts. The false negative, or "missed" charts were more likely to be from women induced for premature rupture of membranes (PROM) (40% vs. 8%, $p < 0.001$) or induced with oxytocin (51% vs. 33%, $p < 0.001$) when compared to the screen positive charts.

Comments

In this study, we evaluated the test characteristics of ICD9-CM coding in identifying women who underwent an induction of labor. We found strong test characteristics, with a PPV of 96.6%, NPV of 74.7%, specificity of 98.8%, and sensitivity of 51.4%. With a sensitivity of 51%, it is reasonable to conclude that ICD9-CM codes alone are not sufficient to screen for inductions. However, the strong PPV signifies that approximately 97% of charts that are coded as induction of labor by ICD9-CM codes are, in fact, patients that were induced. With this degree of accuracy, we can be confident that charts coded as induction are unlikely to be miscoded.

We also compared demographic characteristics of charts identified as induction with ICD9-CM coding and those induction charts that were missed. We found that charts were more likely to be missed by ICD9-CM coding if the induction was performed with oxytocin or for the indication of PROM.

This study not only specifically evaluates the test characteristics of ICD9-CM codes in identifying charts of women who have undergone an induction of labor, but also evaluates the systematic differences between those charts captured by coding versus those who were not. Few studies have specifically targeted ICD9-CM codes for induction of labor in the last decade. Yasmeen et al. found the use of ICD9-CM codes to be 45% sensitive with a PPV of 88% for induction of labor when evaluating California hospital discharge data using similar ICD9-CM codes as our current study¹¹. Lydon-Rochelle found a PPV of 86.4% for induction of labor when using both birth certificates and hospital discharge data together but reported low PPV rates when using birth certificate (52.4%) of hospital discharge data (72.9%) separately¹³. The data analyzed in these two studies; however, were from over a decade ago and substantial changes in medical coding and documentation have occurred over time, which can explain the significant difference between the test characteristics in their studies and those in the present study. Roberts et al. examined the accuracy of hospital discharge data and birth data using ICD10¹⁶ codes for a variety of maternal outcomes including induction of labor¹². They found that the reporting of induction of labor was accurate and reliable with a sensitivity of 92.5% and positive predictive value of 96.1% when using population based birth data; however, using hospital discharge data alone only had a sensitivity of 78% and PPV of 95.4%.

There are possible explanations for why the sensitivity of ICD9-CM coding in our study is low and why the charts of women who are induced with oxytocin and those induced for PROM are more often missed. Hospital discharge coders rely on provider documentation that a procedure was performed, including induction of labor. At our institution, oxytocin is often used for inductions that have a favorable starting cervical exam or for those undergoing a PROM induction. When an induction starts with oxytocin and not with a cervical ripening agent, the physician may be less likely to document it as an induction of labor and perhaps more likely to document it as an augmentation of labor. Using a strict definition of induction of labor, as was done in this study, would limit this systematic misclassification bias.

Our study has several strengths. It is one of the first studies to examine specific characteristics of inductions in order to identify which induction charts are missed when using ICD9-cm coding alone. Additionally, we used physicians rather than using trained medical coders to complete the gold standard chart review with strict definitions of induction of labor. Since we adhered to strict definitions of induction, our gold standard not only verifies that the chart abstractors coded correctly at the time of discharge, but also verifies that the patient truly had an induction. Lastly, the sub-analysis of characteristics of those missed by ICD9-CM coding makes our findings clinically relevant and applicable to those using ICD9-CM codes for induction of labor in future research studies.

Limitations of our study include the following. We analyzed ICD9-CM coding data from one large academic institution. Given differences in provider documentation and coding variation among institutions, our data may not be generalizable to other institutions, although using data from one hospital minimizes practice and documentation variation during labor and delivery. Secondly, ICD10 codes will be introduced in October 2014¹⁶; however, charts previously coded with ICD9-CM will continue to be used as a measurement

of healthcare utilization and identification of charts for research purposes. Therefore, it remains important to ascertain the degree of accuracy of ICD9-CM coding in hospital discharge data despite the upcoming conversion to ICD10.

Our study provides an updated estimate of the test characteristics and usefulness of ICD9-CM codes in identifying charts of women that underwent an induction of labor from hospital discharge data. Unique to our study is the evaluation of characteristics of those inductions missed by ICD9-CM coding, which helps to better understand which charts would not be captured if relying on ICD9-CM coding alone for identifying inductions. While sensitivity is lacking when relying on ICD9-CM codes alone, the 97% PPV is reassuring that we are not improperly coding inductions. This allows researchers to be confident that the majority of charts of women that are coded as an induction did indeed have an induction of labor. Additionally, our study identifies specific areas where documentation can be improved in order to increase the sensitivity and decrease the false negative rate for ICD9-CM codes in identifying inductions. If providers are cognizant of the importance of clear documentation when an induction of labor occurs, this will further improve the test characteristics of ICD9-CM coding and will increase the utility of using these codes for billing, research, and evaluation of healthcare utilization.

In the era where a national database to allow for large scale obstetric studies is lacking, it is reassuring and encouraging that ICD9-CM codes accurately classify the charts of women that undergo an induction of labor and supports their use in discharge data. The lack of a uniform national obstetric database in the United States highlights the need for this type of work to evaluate the reliability and accuracy of ICD9 based data. Additionally, the Joint Commission has identified non-medically indicated deliveries between 37-38 6/7 weeks, including those who undergo an induction of labor, as one of the four core perinatal metrics². Currently, this information is collected through labor intensive chart review. Only with evaluation of this ICD9-CM data and standardization of documentation and chart abstraction, will we be able to evaluate quality metrics on a national level with the goal of improving maternity care and the outcomes of mothers and babies.

Acknowledgements

Adi Hirshberg, MD assisted with some of the data abstraction. Dr. Hirshberg is a resident in Obstetrics and Gynecology in the Department of Obstetrics and Gynecology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA.

Sources of funding: This study was supported in part by a National Institute of Health Reproductive Epidemiology Training Grant: 5T32HD007440-15 and a career development award in Women's Reproductive Health Research: K12-HD001265-14.

References

1. Janakiraman V, Ecker J. Quality in Obstetric Care: Measuring What Matters. *Obstet Gynecol.* 2010; 116:728–32. [PubMed: 20733459]
2. <http://www.jointcommission.org/assets/1/6/Perinatal%20Care.pdf>. Retrieved December 1, 2013
3. International classification of diseases, ninth revision, clinical modification (ICD-9-CM). <http://www.cdc.gov/nchs/icd/icd9cm.htm>. Updated 2013. Retrieved November 24, 2013

4. O'Malley K, Cook K, Price M. Measuring diagnoses: ICD code accuracy. *Health Serv Res.* 2005; 40(5 Pt 2):1620–39. [PubMed: 16178999]
5. Scholes D, Yu O, Raebel MA, Trabert B, Holt VL. Improving automated case finding for ectopic pregnancy using a classification algorithm. *Human Reproduction.* 2011; 26:3163–68. [PubMed: 21911435]
6. Brubaker L, Bradley CS, Handa VL, et al. Anal sphincter laceration at vaginal delivery: is this event coded accurately? *Obstet Gynecol.* 2007; 109:1141–5. [PubMed: 17470596]
7. Henry OA, Gregory KD, Hobel CJ, Plat LD. Using ICD-9 Codes to identify indications for primary and repeat cesarean sections: Agreement with clinical records. *Am J Public Health.* 1995; 85(8 Pt 1):1143–6. [PubMed: 7625515]
8. Geller SE, Ahmed S, Brown ML, Cox SM, Rosenberg D, Kilpatrick SJ. International classification of diseases-9th revision coding for preeclampsia: How accurate is it? *Am J Obstet Gynecol.* 2004; 190:1629–33. [PubMed: 15284758]
9. Romano PS, Yasmeen S, Schembri ME, Keyzer JM, Gilbert WM. Coding of perineal lacerations and other complications of obstetric care in hospital discharge data. *Obstet Gynecol.* 2005; 106:717–25. [PubMed: 16199627]
10. Goff S, Pekow P, Markenson G, Knee A, Chasan-Taber L, Lindenauer P. Validity of using ICD-9-CM codes to identify selected categories of obstetric complications, procedures and co-morbidities. *Paediatr Perinat Epidemiol.* 2012; 26:421–9. [PubMed: 22882786]
11. Yasmeen S, Romano PS, Schembri ME, Keyzer JM, Gilbert WM. Accuracy of obstetric diagnoses and procedures in hospital discharge data. *Am J Obstet Gynecol.* 2006; 194:992–1001. [PubMed: 16580288]
12. Roberts CL, Bell JC, Ford JB, Morris JM. Monitoring the quality of maternity care: How well are labour and delivery events reported in population health data? *Paediatr Perinat Epidemiol.* 2008; 23:144–52. [PubMed: 19159400]
13. Lydon-Rochelle MT, Holt VL, Nelson JC, et al. Accuracy of reporting maternal in-hospital diagnoses and intrapartum procedures in Washington state linked birth records. *Paediatr Perinat Epidemiol.* 19:460–71. [PubMed: 16269074]
14. Martin J, Hamilton B. Births: Final data for 2006. *Natl Vital Stat Rep.* 2009; 57:1–102.
15. Levine LD, Bogner H, Hirshberg A, Elovitz MA, Sammel MD, Srinivas SK. Term induction of labor and subsequent preterm birth. *Am J Obstet Gynecol.* Oct.2013 Epub ahead of print.
16. ICD-10. <http://www.cms.gov/Medicare/Coding/ICD10/index.html?redirect=/icd10/>. Updated 2013. Retrieved August 18, 2013

Figure 1. Flowchart of women included

ICD9-CM: International Classification of Disease, 9th Edition-*Clinical Modification*

Table 1

Test characteristics for ICD9-CM identifying charts of women undergoing an induction of labor

	Induction noted on chart review	No induction after chart review
ICD9-CM codes for induction, screen positive n=708	684	24
No ICD9-CM code for induction, screen negative n=451 (sampled out of 2555)	114	337
PPV: $684/708 = 96.6\%$ NPV ^a : $337/451 = 74.7\%$ False negative ^b : $114/451 = 25.3\%$ Sensitivity ^c : 51.4% Specificity ^c : 98.8%		

PPV: positive predictive value, NPV: negative predictive value

^aUsing a NPV of 74.7% and applying it to entire screen negative cohort, there are 1909 out of 2555 screen negative charts that would be true negatives.

^bUsing a false negative rate of 25.3% and applying it to the entire screen negative cohort, there are 646 out of 2555 screen negative charts that would be false negatives.

^c: NPV and false negative rates (74.7% and 25.3%) were applied to the entire 2555 cohort to determine the sensitivity and specificity

Table 2

Baseline demographic characteristics of the screen positive and false negative groups

	a	b	c	p-value ^a
	Overall cohort of charts evaluated (n=1159)	Identified as inductions by ICD9-CM: "screen positive" (n=708)	Not identified as induction by ICD9-CM: "false negative" (n=114)	
Maternal age (years)^b	24 [19-29]	24 [19-30]	24 [20-29]	0.7
BMI (kg/m²)^b	26.5 [22.4-30.9]	27 [22.9-31.9]	25.9 [21.9-30.3]	0.1
Race	77	77	76	0.9
African American	16	16	17	
Caucasian	7	7	7	
Other				
Nulliparous	63	66	67	0.7
Cesarean delivery rate	20	24	24	--
Indication for induction:	9	8	40	<0.001
Premature ROM	12	17	8	
Post-dates	21	30	22	
Maternal	23	33	21	
Fetal	7	9	9	
Elective	28	3	NA	
Not induced				
Methods of induction	36	33	51	0.006
Oxytocin	53	56	37	
Prostaglandin	7	7	6	
Cervical foley	4	4	6	
ROM				

Data are presented as % unless otherwise specified

BMI: Body mass index, ROM: Rupture of membranes

^a p-value is comparing columns b and c.^b Median [Inter-quartile range]