Published in final edited form as:

Med Care. 2012 August ; 50(8): 714–721. doi:10.1097/MLR.0b013e31824e32cf.

Very Low Birth Weight Hospital Volume and Mortality – An Instrumental Variables Approach

George L. Wehby, Ph.D., MPH [Assistant Professor],

Dept. of Health Management and Policy, University of Iowa, 105 River Street, N248 CPHB, Iowa City, IA 52242, Phone: 319-384-5133, Fax : 319-384-5125

Fred Ullrich [Research Specialist], and

Dept. of Health Management and Policy, University of Iowa, Iowa City, IA 52242

Yang Xie, Ph.D., MPH [Associate Director]

Global Outcomes Research, Merck & Co., Inc., One Merck Drive, WS-2E55, Whitehouse Station, NJ 08889, Phone: 908-423-4091, Fax: 908-775-1688

George L. Wehby: george-wehby@uiowa.edu; Yang Xie: yang.xie@merck.com

Abstract

Background—Previous studies of very low birth weight (VLBW) hospital volume effects on inhospital mortality have used standard risk-adjusted models that only account for observable confounders but not for self-selection bias due to unobservable confounders.

Objective—To assess the effects of hospital volume of VLBW infants on in-hospital mortality while explicitly accounting for unobservable confounders and self-selection bias using an instrumental-variables model.

Methods—The sample includes 4,553 VLBW infants born in 63 hospitals in 2000–2004 in New Jersey. We employ instrumental-variables analysis using as instruments the differences between the patient's distances to the nearest low (<50 VLBW infants annually), moderate (51–100 infants annually) and high (>100 VLBW infants annually) volume hospitals. We evaluate several volume measures and adjust for observable infant and hospital characteristics.

Results—We find beneficial volume effects on survival that are significantly underestimated in classical risk-adjusted models, under which low and moderate volumes compared to high volumes increase mortality odds by 1.8 and 1.88 times, respectively (risk ratios of 1.4 and 1.5, respectively). However, using an instrumental variables approach, we find that low and moderate volumes increase mortality odds by 5.42 and 3.51 times, respectively (risk ratios of 2.76 and 2.21, respectively). These findings suggest unobservable confounders that increase the selection of infants at a greater mortality risk into higher volume hospitals.

Conclusions—Accounting for unobserved self-selection bias reveals large survival benefits from delivering and treating VLBW infants at high-volume hospitals. This supports policies regionalizing the delivery and care for pregnancies at-risk for VLBW at high-volume hospitals.

Keywords

Very low birth weight; hospital volume; neonatal mortality; infant mortality; neonatal intensive care; instrumental variables

Introduction

Infants born at very low birth weight (VLBW) – less than 1500 grams – have high neonatal/infant mortality rates. In 2001 in the US, where 1.4% of live births were VLBW, neonatal mortality was 21.38%.¹ While infant mortality rates have decreased significantly over the past two decades, the decline in mortality for VLBW infants has been approximately 50% of that for higher birth weight infants,¹ indicating increasing survival disparity between VLBW infants and those at higher birth weight.

Regionalizing delivery and care for VLBW infants has received considerable research and consumer-group attention as an approach for enhancing survival.^{2, 3} The underlying theory is that higher-volume hospitals acquire knowledge and experience from treating more cases and increase their staff human capital and technological resources, which in turn improve care quality and patient outcomes.

Few studies have evaluated the effects of VLBW hospital volume on in-hospital mortality in the US and other developed countries.^{2, 4–7} Most studies have generally found small to moderate reductions in mortality with increasing volume.^{2, 4–6} However, one study has reported small reductions in mortality with volume up to 50 VLBW infants annually, and small mortality increases with volume above this threshold.⁷

A main analytical challenge in volume-outcome studies that only rely on adjusting for observable confounders is the potential bias in volume-effect estimates due to self-selection into hospitals based on volume *and* “unobserved” individual-level confounders, such as maternal/fetal health risks. The direction of such bias is theoretically ambiguous. Volume benefits may be underestimated if more severe/complicated VLBW cases are more likely to be delivered at higher-volume hospitals (adverse self-selection). In contrast, volume benefits may be overestimated if less severe/complicated cases select more into higher-volume hospitals (favorable self-selection). All previous studies of VLBW-volume effects on mortality have only adjusted for observable confounders and none has explicitly accounted for self-selection bias from unobservable confounders.^{2, 4–7}

In this study, we assess the effects of hospital volume of VLBW infants on in-hospital mortality using an instrumental-variable model in order to account for self-selection into hospital volume based on unobservable confounders.

Methods

Study sample and Measures

The study sample includes 4,553 VLBW infants born in 63 hospitals in New Jersey (NJ) between 2000 and 2004 and identified from the State Inpatient Dataset (NJ-SID).⁸ The NJ-SID includes the entire population of in-hospital births in NJ. The data are from inpatient hospital-discharge records systematically reported from hospitals to the NJ Department of Health and Senior Services and are released under the auspices of the Agency for Healthcare Research and Quality (AHRQ). The 2000–2004 American Hospital Association Annual Surveys of Hospitals are used to obtain data on hospital beds, staffing level, and teaching status. Data from the 2009 American Academy of Pediatrics list of neonatal intensive care units (NICUs) are used to assign hospital-NICU levels. NICU data for years 2000–2004 are unavailable. However, it is unlikely that hospital-NICU levels frequently change over a few years, and we only use NICU level as a covariate. Furthermore, we find that the results are overall insensitive to excluding NICU level altogether. We include zip-code level average school years completed among adult females and median-household income from the 2000

U.S. Census and county-level unemployment rates from the U.S. Bureau of Labor Statistics for years 2000–2004.

We limit the main sample to in-hospital delivered VLBW infants who were either discharged home after delivery or died in the hospital. Infants transferred to another hospital are excluded from the main analyses for two reasons. First, we focus on identifying the effects of “*being born and cared for*” at lower-versus higher-volume hospitals. This is the main question of interest for researchers and policymakers when evaluating the effectiveness of care for VLBW infants at lower-volume hospitals. Excluding transferred infants should not bias the volume effect for non-transferred infants who remain hospitalized at lower-volume hospitals. In addition to this conceptual justification, we face a data limitation; even though we observe discharge status for each infant, we cannot link with certainty pre- and post-transfer data for transferred infants since the dataset has no unique-infant identifiers. Previous studies with identifiers have generally assigned transferred infants into the volumes of the hospitals to where they were transferred. This may bias the volume effect downward by ignoring the pre-transfer care at lower-volume hospitals. An opposite bias may occur if transferred infants are assigned to the volumes of the transferring hospitals. We employ sensitivity analyses described below that explicitly account for infant transfers and find that our inference is robust to excluding transferred infants.

We exclude infants of mothers residing outside of NJ for whom we cannot calculate the distance instruments defined below and infants with additional health conditions (including congenital anomalies, birth weight less than 500 grams, gestational age less than 24 weeks, and multiple birth) that may modify the volume effects but are a small group to study alone. Finally, we exclude infants with incomplete data on study variables.

The outcome is in-hospital mortality of VLBW infants, defined by death after delivery and before discharge to home based on hospital discharge status. The main volume measures are two binary indicators for ranges of the number of VLBW infants born in the hospital in a given year. Using indicators for volume ranges accounts for potential non-linearity in volume effects. We designate hospitals as low, moderate, and high volumes where the number of VLBW births per year was ≤ 50 , 51–100, and >100 , respectively, and estimate the effects of low and moderate volumes relative to high volume (reference category). These ranges are consistent with cutoffs in previous studies,^{4, 6} allowing comparison to their results, and have well-balanced frequencies given the sample’s distribution of the hospitals’ numbers of VLBW infants (about 19%, 34%, and 47% of infants are in low, moderate, and high volume categories, respectively).

We also evaluate a continuous volume measure representing the hospital annual number of VLBW infants. Furthermore, we evaluate the effects of this continuous measure within two volume thresholds: up to 50 VLBW infants annually and above 50 VLBW infants in order to compare to previous research.⁷ We do this by adding an interaction term between continuous volume and a binary threshold indicator (≤ 50 versus >50 VLBW infants).

Instrumental Variables/Statistical Analysis

Instrumental variables (IV) are used to identify treatment effects accounting for self-selection into treatments (hospital volume here) based on unobserved confounders.^{9–11} The method employs variables, called instruments that should be: 1) strongly related to treatment selection; and 2) unrelated to the outcome except through their effect on treatment selection (i.e. the instruments should not affect the outcome directly or correlate with unobserved confounders). Under these conditions, the instruments exploit exogenous variation in the treatment that is independent of self-selection bias.

There are numerous IV applications.^{12–15} Among those studying hospital treatments, the distance between the patient's residence and the nearest hospital that provides the treatment, or the difference between this distance and the distance to the nearest hospital that does not provide the treatment – differential-distance instrument – is a commonly used instrument.^{13–15} The strength of such instruments is that distance is strongly predictive of treatment selection and can be theorized, particularly in the case of differential-distance instruments, to be unrelated to unobserved confounders affecting treatment selection and outcomes.

We use as instruments the differences between each of the distances from the mother's residence to the nearest low- and moderate-volume hospitals and the distance to the nearest high-volume hospital. Specifically, we subtract the distance to the nearest high-volume hospital from the distances to the nearest low- and moderate-volume hospitals. Since data on mothers' addresses are only available at the zip-code level, distances (measured in estimated drive time minutes) are calculated from the mother's zip-code centroid to the nearest hospital in each volume category (full addresses are available for hospitals). We use these two instruments for all the binary and continuous volume measures.

The differential-distance instrument has a theoretical advantage over distances to the nearest hospital in each volume category as instruments.^{16, 17} If any unobserved confounders influence residential location in general, it is unlikely that they determine residential choice based on the difference in distances to hospitals of various volumes. Unobserved maternal/fetal health risks are likely the main unobserved confounders. Pregnancy complications due to poor maternal health (hypertension, diabetes, infections, or other chronic/acute conditions) or fetal health problems (e.g., very poor fetal growth) that may predispose to delivering at higher-volume hospitals but also increase VLBW mortality are unlikely to relate to the differential-distance instrument. However, it is *impossible* to *fully* test this assumption due to the unobservable factors, which highlights the importance of selecting instruments based on theory.

We estimate the instrumental-variable model using the two-stage residual-substitution method.¹⁸ The first stage regresses the volume measures, including the interaction term between continuous volume and volume-threshold indicator when using that measure,¹⁹ on the instruments and all covariates using ordinary-least-squares (OLS) and obtains the residual terms. The second stage adds these residuals as regressors into a logistic regression for mortality. This controls for unobserved confounders correlated with both volume and mortality. The usual standard errors in the second-stage IV-logistic regression are biased due to including the first-stage residuals. Therefore, we bootstrap the IV model (both stages) with 2,000 replications in order to consistently estimate 95% confidence intervals for volume effects. Bootstrap is an appropriate method for estimating variance in this model.¹⁸

This IV method requires a consistent estimator of the first stage residuals. Since OLS consistently estimates these residuals, it is an appropriate choice for estimating the first stage.¹⁸ While estimating the first stage for the three-category volume measure by multinomial logit may seem appealing, it is unclear how to define and consistently estimate residuals from that model that would have appropriate properties for the IV residual substitution method. Since both residual terms from the two first stage OLS regressions for the low and moderate volumes are included in the mortality function, this accounts for any correlations between these error terms due to unobservable confounders related to both volumes.

Both stages are adjusted for relevant observed characteristics including demographics (race, sex), insurance status, infant's birth weight, and an indicator for antenatal and delivery

complications (hematologic, intrauterine growth restriction, oligohydramnios, RH isoimmunization, breech presentation, chorioamnionitis, cord prolapse, and fetal distress problems). There is no continuous measure of gestational age in the dataset but only an indicator for gestational age less than 24 weeks (an exclusion factor in our study). We also adjust for zip-code level average schooling years among females older than 18 years and mean household income, county-level unemployment rate, and hospital-level characteristics including numbers of beds and full-time staff, teaching status, and NICU level. In order to compare to the IV model, we estimate a classical risk-adjusted logistic regression for mortality with standard errors clustered at the hospital-level.

We report odds ratios (ORs) for volume effects in order to compare to previous studies, which only reported ORs. Since ORs may not accurately estimate relative risk for relatively common outcomes such as mortality in this sample, we also report risk ratios (RRs) for the main models.

We evaluate the significance of the instruments in the first-stage regressions using a partial F-statistic. Since the instruments are calculated at the mothers' zip-code level, we cluster the first-stage standard errors at the zip-code level. As a partial assessment of whether the instruments may satisfy the second IV assumption, we evaluate their relationships with observable individual-level confounders by regressing them on these variables; we find no significant relationships, supporting the exogeneity of the instruments. All analyses were done using *Stata 11*.

Sensitivity analyses

In order to gauge the sensitivity of volume effects to the exclusion of transferred infants, we define a three-category outcome that includes transferred infants in a separate category in addition to the two categories of dead or discharged alive. Next, we simultaneously estimate the volume effects on these three categories using multinomial logistic regression. This analysis formally accounts for differences in transfers by volume when estimating the volume effects on mortality and allows for volume effects to vary between transfers and mortality. As shown below, we find that our general inference is insensitive to accounting for transfers. In another analysis, we evaluate interactions between volume and NICU level but find no significant interactions. Finally, in order to evaluate the effect of potential measurement error in NICU level, we estimate a model that excludes NICU level.

Results

Sample Description

The total sample that meets the study inclusion criteria includes 6,176 infants. Of these, 1,623 infants meet the exclusion criteria, resulting in a final sample of 4,533 infants for the main analysis. Figure 1 shows a flow chart of sample exclusions.

Table 1 reports the distribution of the study outcome and variables for the total sample and subgroups stratified by VLBW volume. About 13.7% of the sample infants died before discharge. The unadjusted mortality rates were higher in low- and moderate-volume hospitals than high-volume hospitals (15.8%, 16.3%, and 11%, respectively). Hospital volume was significantly correlated with all observed characteristics except female birth. Birth weight and rates of antenatal and delivery complications were higher at low- and moderate-volume hospitals compared to high-volume hospitals. Publicly insured, uninsured, and black infants were more likely to be delivered at low- and moderate-volume hospitals compared to high-volume ones.

Instrument Quality

Table 2 reports the first-stage OLS regressions. The instruments have significant effects on hospital-volume choice, with a partial F-statistic of 22 for low volume and 275 for moderate volume, which exceed the minimum-threshold of 10 for non-weak instruments.²⁰ Mothers living closer to high-volume than low-volume hospitals (greater positive differential distance between the nearest low- and high-volume hospitals) are less likely to deliver at a low-volume hospital and vice-versa. A 10-minute increase in differential driving time to a low- versus high-volume hospital increases the probability of delivering at a high-volume hospital by 0.056. In contrast, an increase in distance to the nearest moderate-volume hospital relative to distance to the nearest high-volume hospital increases the probability of selecting a low-volume hospital. Similar instrument effects are observed for moderate-volume choice. A 10-minute increase in differential driving time to a moderate-versus high-volume hospital increases the probability of delivering at a high-volume hospital by 0.152. The F-statistics for instrument effects on continuous volume and its interaction with the volume-threshold indicator are 79.4 and 13.7, respectively.

Table 3 reports the F-statistics from the regressions of the instruments on individual-level covariates, one at a time. The instruments are only marginally significantly related to insurance status but not to other variables, providing support that they are exogenous.

Volume Effects on Mortality

Table 4 reports the low- and moderate-volume effects relative to high volume on in-hospital mortality as estimated from the IV model and classical logistic regression. Supplementary Table A1 reports full regression results.

We first describe the ORs as reported in previous studies. Using risk-adjusted logistic regression, low and moderate volumes compared to high volume are significantly associated with 1.8-fold and 1.88-fold increases in mortality odds, respectively. Slightly smaller effects are observed in the unadjusted regression. The volume effects are significant and noticeably larger under the IV model than the classical logistic regression. Using the IV model, low and moderate volumes compared to high volume increase mortality odds by 5.4 and 3.51 times, respectively. The hypothesis that volume measures are exogenous and that estimates from classical risk-adjusted logistic regression are unbiased is rejected for moderate volume based on a Hausman-type endogeneity test for the significance of the first-stage residual coefficients in the mortality function ($p < 0.05$), indicating unobservable confounders related to both volume selection and mortality.^{21, 22}

As expected, ORs overestimate relative risk and exceed RRs. However, overall inference is unaffected by using RRs. Under the classical risk-adjusted model, low and moderate volumes have RRs of 1.45 and 1.49, respectively, compared to 2.76 and 2.21 under the IV model.

Similar differences between the classical and IV estimates are observed for continuous volume (Table 4). In the unadjusted logistic regression, an increase in VLBW volume by 10 infants is associated with a 4% decrease in mortality odds ($OR = 0.96$); the adjusted effect is insignificant ($OR = 0.95$). In the IV model, an increase in VLBW volume by 10 infants reduces mortality odds by 16% ($OR = 0.84$). The classical risk-adjusted estimates are rejected based on significance of the first-stage residual in the mortality function (significant at $p < 0.01$). The RRs for an increase in volume by 10 infants are 0.97 in the classical model and 0.9 in the IV model.

The classical model suggests insignificant volume-change effects within either range of 50 infants or less or more than 50 infants. In contrast, the IV model suggests a larger and

significant effect of volume increases at volumes above 50 infants with a volume increase by 10 infants decreasing mortality odds by 17% (OR=0.83); RR is 0.89. However, volume increases below the 50-infant threshold have smaller and insignificant effects.

Sensitivity Analyses

Supplementary Table A2 reports the volume effects on mortality from the multinomial logistic regression including transferred infants in a separate category in addition to discharged alive and dead. This analysis adds 1,777 transferred infants that have complete data on study variables and pass other exclusion criteria for the main sample. A similar pattern of results is observed to the main analysis excluding transferred infants. The volume effects on mortality using the classical risk-adjusted model change little. The IV estimates of volume effects on mortality from this multinomial model are even larger than the ones excluding transfers. Therefore, the main finding of underestimated volume benefits in classical risk-adjusted models is not sensitive to including or excluding transferred infants. Furthermore, our classical estimates are within range of those from previous studies with post-transfer information,⁶ which provides further assurance that transfer status is not biasing our results.

We find an overall similar pattern of results for volume effects when excluding NICU level from the model (supplementary Table A3). Therefore, it is unlikely that potential measurement error in NICU level is significantly biasing our volume effects.

Discussion

We find beneficial volume effects for in-hospital survival using the IV-model that are significantly underestimated in classical risk-adjusted models. The analysis rejects the estimates from classical risk-adjusted models that only account for observable confounders. The results suggest adverse self-selection into higher-volume hospitals. Specifically, infants with “unobserved” characteristics that increase their mortality risks are more likely to be delivered at higher-volume hospitals. These may include poor fetal growth and other maternal/fetal health risks that may be identified during prenatal care, all of which are often inadequately captured in secondary data sources. Ignoring self-selection on unobservable confounders may seriously underestimate the improvement in survival with delivery and care for VLBW infants at higher-volume hospitals. The study, which is the first to explicitly account for unobserved self-selection bias in this context, highlights the importance of accounting for unobservable confounders for obtaining accurate estimates of hospital-volume effects on infant health. Since administrative datasets cannot measure all relevant confounders, direct risk-adjustment alone may not remove self-selection bias. Therefore, it is important to employ models that explicitly account for such bias.

The results support regionalizing the delivery and care for VLBW infants at high-volume hospitals (>100 VLBW infants annually). Based on our IV estimates, 87 infants (64%) of the 136 infants who died in low-volume hospitals would have survived if they were delivered at high-volume hospitals. Similarly, 138 infants (55%) of the 250 infants who died at moderate-volume hospitals would have survived if they were delivered at high-volume hospitals. Therefore, policies that increase the access of at-risk pregnancies to high-volume hospitals will have large survival returns.

The study results should be considered within some qualifications. While the study sample is fairly large and population-based, it is based on one state and is smaller than those of previous VLBW volume-mortality studies using classical models. Therefore, it is important to replicate this study in larger samples from other states. Also, the model does not significantly reject the classical estimate for the low- versus high-volume effect ($p=0.13$).

However, the nearly significant p-value and rejecting the classical estimates for the moderate and continuous volume measures suggest that the classical estimate for low volume may have not been rejected because of limited power, which is a common limitation of such tests. Also, we cannot capture pregnancies from NJ delivered in nearby states. However, insurance coverage restrictions on seeking care from non-contracted hospitals reduce such occurrences. This does not bias our estimates, but limits the generalizability of the results to births in NJ hospitals to resident mothers. Finally, the dataset does not allow for linking maternal and infant data. Observing maternal characteristics is useful for risk adjustment and evaluating the sources of self-selection bias. However, adjusting for general maternal characteristics available in birth registry data such as prenatal care use or general health conditions is unlikely to fully account for self-selection bias, as these characteristics were included in previous studies reporting comparable classical estimates. Future studies that evaluate the effects of hospital volume on health outcomes of surviving infants are needed. Also, evaluating if hospital volume effects vary by socioeconomic and clinical characteristics is important for identifying any disparities in benefiting from improved care effectiveness at larger volumes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Data analysis was in-part supported by NIH grant 1R03 DE018394. The NIH had no role in the study design, data collection and analysis, or manuscript writing and submission.

References

1. Arias E, MacDorman MF, Strobino DM, Guyer B. Annual summary of vital statistics--2002. *Pediatrics*. 2003; 112(6 Pt 1):1215–1230. [PubMed: 14654589]
2. Phibbs CS, Bronstein JM, Buxton E, Phibbs RH. The effects of patient volume and level of care at the hospital of birth on neonatal mortality. *JAMA*. 1996; 276:1054–1059. [PubMed: 8847767]
3. LeapFrog Group. [accessed on December 13, 2011] Fact Sheet: Evidence-Based Hospital Referral. http://www.leapfroggroup.org/media/file/FactSheet_EBHR.pdf
4. Phibbs CS, Baker LC, Caughey AB, Danielsen B, Schmitt SK, Phibbs RH. Level and volume of neonatal intensive care and mortality in very-low-birth-weight infants. *N Engl J Med*. 2007; 356:2165–2175. [PubMed: 17522400]
5. Bartels DB, Wypij D, Wenzlaff P, Dammann O, Poets CF. Hospital volume and neonatal mortality among very low birth weight infants. *Pediatrics*. 2006; 117:2206–2214. [PubMed: 16740866]
6. Chung JH, Phibbs CS, Boscardin WJ, Kominski GF, Ortega AN, Needleman J. The effect of neonatal intensive care level and hospital volume on mortality of very low birth weight infants. *Med Care*. 2010; 48:635–644. [PubMed: 20548252]
7. Rogowski JA, Horbar JD, Staiger DO, Kenny M, Carpenter J, Geppert J. Indirect vs direct hospital quality indicators for very low-birth-weight infants. *JAMA*. 2004; 291(2):202–209. [PubMed: 14722146]
8. New Jersey State Inpatient Database (SID), Healthcare Cost and Utilization Project (HCUP), Agency for Healthcare Research and Quality, 2000–2004.
9. Angrist JD, Imbens GW, Rubin DB. Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*. 1996; 91:444–455.
10. Newhouse JP, McClellan M. ECONOMETRICS IN OUTCOMES RESEARCH: The Use of Instrumental Variables. *Annual Review of Public Health*. 1998; 19:17–34.
11. Angrist JD, Krueger AB. Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives*. 2001; 15:69–85.

12. Wehby GL, Murray JC, Castilla EE, Lopez-Camelo JS, Ohsfeldt RL. Quantile effects of prenatal care utilization on birth weight in Argentina. *Health Economics*. 2009; 18:1307–1321. [PubMed: 19142894]
13. McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *JAMA*. 1994; 272:859–866. [PubMed: 8078163]
14. Brooks JM, Chrischilles EA, Scott SD, Chen-Hardee SS. Was breast conserving surgery underutilized for early stage breast cancer? Instrumental variables evidence for stage II patients from Iowa. *Health Serv Res*. 2003; 38(6 Pt 1):1385–1402. [PubMed: 14727779]
15. McConnell KJ, Newgard CD, Mullins RJ, Arthur M, Hedges JR. Mortality benefit of transfer to level I versus level II trauma centers for head-injured patients. *Health Serv Res*. 2005; 40:435–457. [PubMed: 15762901]
16. Kahn JM, Ten Have TR, Iwashyna TJ. The relationship between hospital volume and mortality in mechanical ventilation: an instrumental variable analysis. *Health Serv Res*. 2009; 44:862–879. [PubMed: 19674428]
17. Tsai AC, Votruba M, Bridges JF, Cebul RD. Overcoming bias in estimating the volume-outcome relationship. *Health Serv Res*. 2006; 41:252–264. [PubMed: 16430610]
18. Terza JV, Basu A, Rathouz PJ. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of Health Economics*. 2008; 27:531–543. [PubMed: 18192044]
19. Kelejian HH. Two-Stage Least Squares and Econometric Systems Linear in Parameters but Nonlinear in the Endogenous Variables. *Journal of the American Statistical Association*. 1971; 66:373–374.
20. Staiger D, Stock JH. Instrumental Variables Regression with Weak Instruments. *Econometrica*. 1997; 65:557–586.
21. Wooldridge, JM. *Econometric analysis of cross section and panel data*. Cambridge and London: MIT Press; 2002.
22. Hausman JA. Specification Tests in Econometrics. *Econometrica*. 1978; 46:1251–1271.

Figure 1.
Sample exclusions

Table 1
Description of Study Variables

Variable	Total sample (N=4,553)	Low volume 50 (N=863)	Moderate volume 51–100 (N=1,535)
VLBW in-hospital mortality (%)	13.66	15.76 ^{***}	16.29 ^{***}
Continuous annual VLBW volume (in increments of 10)	9.25 (4.81) [0.1–18.9]	2.73 ^{***} (1.61) [0.1–5]	6.79 ^{***} (1.09) [5.1–9.2]
Low volume (%) ^a	18.95		
Moderate volume (%) ^a	33.71		
Female (%) ^b	49.84	48.78	49.84
Black (%) ^c	29.80	32.91 ^{***}	31.07 ^{***}
Other non-White (%) ^c	27.08	31.29 ^{***}	22.21 ^{***}
Medicaid/CHIP (%) ^d	17.20	23.99 ^{***}	19.22 ^{***}
Other non-private insurance/out of pocket payment (%) ^d	7.45	11.70 ^{***}	7.49 ^{***}
Antenatal/delivery complications (%) ^e	4.57	5.10 ^{**}	5.47 ^{**}
Birth weight (in grams)	1059.71 (282.65) [500–1,498]	1088.81 ^{***} (290.89) [500–1,498]	1070.31 ^{***} (285.26) [500–1,498]
Education rate (average number of school years for females > 18 years)	12.73 (1.12) [9.78–15.72]	12.48 ^{***} (1.04) [9.78–15.72]	12.62 ^{***} (1.00) [9.78–15.57]
Median household income	53725.69 (22141.54) [13,421–185,466]	47656.48 ^{***} (19365.88) [13,421–185,466]	50394.70 ^{***} (19149.61) [13,421–132,373]
Unemployment rate	5.01 (1.26) [2.3–8.1]	5.42 ^{***} (1.25) [2.6–7.9]	4.79 ^{***} (1.21) [2.3–8.1]
Staffed hospital beds	562.98 (367.40) [10–2,121]	373.83 ^{***} (124.17) [10–650]	564.77 ^{***} (441.26) [10–2,121]
Total FTE hospital personnel	2917.15 (1540.34) [353–7,796]	2090.24 ^{***} (1125.85) [353–5,955]	3129.41 ^{***} (1802.57) [472–7,796]
Teaching hospital (%) ^f	66.40	28.04 ^{***}	64.43 ^{***}
NICU I or II level (%) ^g	10.17	37.20 ^{***}	9.25 ^{***}
NICU IIIA level (%) ^g	4.28	14.02 ^{***}	4.82 ^{***}
NICU IIIB level (%) ^g	74.19	44.03 ^{***}	68.21 ^{***}
Differential distance to low volume (drive time minutes)	–20.37 (24.26)	–31.6 ^{***} (30.5)	–31.2 ^{***} (24.67)
Differential distance to moderate volume (drive time minutes)	–6.74 (24.7)	–12.54 ^{***} (25.29)	–24.75 ^{***} (23.2)

Note: The table reports the frequency (%) for binary 0/1 variables and means, standard deviations in parentheses, and ranges in brackets for continuous variables.

^a 0/1 indicator with high volume of more than 100 VLBW infants annually as the reference category;

^b 0/1 indicator with male as the reference category;

^c 0/1 indicator with White as the reference category;

^d 0/1 indicator with private insurance as the reference category;

^e 0/1 indicator with no complications as the reference category;

^f 0/1 indicator with non-teaching hospital as the reference category;

$\frac{g}{0/1}$ indicator with NICU IHC level as the reference category.

and indicate significant association between the variable and hospital volume at $p < 0.05$ and $p < 0.01$, respectively.

The associations are evaluated using a chi-square test of independence between a variable combining all categories for categorical variables (such as insurance or race) and the three-category volume measure. The associations for continuous variables are evaluated by regressing these variables on the two binary hospital volume indicators.

Table 2

First Stage Regression Coefficients of the IV Model

	Low volume	Moderate volume	Continuous volume	Continuous volume X Low volume
Differential distance to low volume	−0.0056 *** (0.0011)	0.0049 *** (0.0011)	−0.2880 (0.1772)	−0.0205 *** (0.0039)
Differential distance to moderate volume	0.0063 *** (0.0010)	−0.0152 *** (0.0009)	1.7949 *** (0.2223)	0.0212 *** (0.0043)
Female ^a	−0.0047 (0.0084)	0.0108 (0.0120)	−1.5272 (1.9996)	−0.0251 (0.0315)
Black ^b	0.0568 *** (0.0163)	−0.0848 *** (0.0231)	1.4289 (4.1213)	0.2588 *** (0.0650)
Other non-White ^b	0.0056 (0.0131)	−0.0874 *** (0.0198)	−0.9140 (3.9597)	0.0589 (0.0508)
Medicaid/CHIP ^c	0.0412 *** (0.0155)	0.0240 (0.0165)	−10.5447 *** (3.7405)	0.1279 ** (0.0585)
Other non-private ^c insurance/out of pocket payment	0.0639 *** (0.0227)	0.0209 (0.0233)	−21.9545 *** (4.6190)	0.2225 ** (0.0914)
Antenatal/delivery complications ^d	0.0174 (0.0238)	0.0302 (0.0297)	−17.6220 *** (4.6702)	0.0855 (0.0845)
Birth weight	0.00003 (0.00002)	0.0001 *** (0.00002)	−0.0134 *** (0.0039)	0.0001 (0.0001)
Education rate	0.0151 (0.0190)	0.0035 (0.0168)	−0.6202 (3.9542)	0.1182 (0.0756)
Median household income	−0.000001 (0.000001)	−0.000001 (0.000001)	0.0003 * (0.0002)	−0.0000 ** (0.0000)
Unemployment rate	0.0453 *** (0.0073)	−0.0614 *** (0.0077)	−3.5326 *** (1.2648)	0.1046 *** (0.0238)
Hospital beds	−0.0005 *** (0.0001)	−0.0001 (0.0001)	0.0756 *** (0.0126)	−0.0019 *** (0.0003)
Total FTE personnel	0.0001 *** (0.00002)	0.0001 *** (0.00001)	−0.0246 *** (0.0035)	0.0004 *** (0.0001)
Teaching hospital ^e	−0.2152 *** (0.0272)	0.0620 ** (0.0308)	47.5559 *** (5.7523)	−0.5901 *** (0.1135)
NICU I or II level ^f	0.5956 *** (0.0582)	−0.3557 *** (0.0504)	−29.3692 *** (10.0992)	0.7061 *** (0.2136)
NICU IIIA level ^f	0.3721 *** (0.0587)	−0.0349 (0.0653)	−21.2670 *** (8.1168)	1.0638 *** (0.2457)
NICU IIIB level ^f	0.0613 * (0.0362)	−0.1976 *** (0.0286)	45.8802 *** (5.6633)	0.2745 * (0.1471)
Constant	−0.2366 (0.2123)	0.6632 *** (0.1864)	108.4737 ** (42.6161)	−1.4763 * (0.8216)
F (2, 456) for instrument effects	22.4 ***	274.9 ***	79.42 ***	13.68 ***
R-squared	0.43	0.41	0.49	0.22
Sample size	4,553	4,553	4,553	4,553

Note: The Table reports the coefficients and standard errors (in parentheses) from the regressions of volume measures on the instruments and other control variables. Standard errors are adjusted for clustering at the patient's zip-code level.

* $p < 0.1$,

** $p < 0.05$,

*** $p < 0.01$.

The reference categories are:

^a male;

^b White;

^c private insurance;

^d no complications;

^e non-teaching hospital; and

^f NICU IIIC level.

Table 3

F-Statistics for Effects of Individual-Level Characteristics on the Instruments

Individual-level characteristic	Differential distance between nearest low- and high-volume hospitals	Differential distance between nearest moderate- and high- volume hospitals
Infant's sex	1.40	2.07
Insurance status	2.65 *	2.84 *
Race	0.18	1.79
Antenatal/delivery complications	0.20	1.25
Birth weight	0.56	0.08

Note: Table 3 reports the F-statistics of the models that regress separately the two instruments on the individual-level characteristics, one at a time.

* indicates $p < 0.1$.

Table 4
Volume Effects on In-Hospital Mortality Under Different Logistic Regression Models and Volume Measures

Annual volume	Classical – Unadjusted		Classical – adjusted		IV model	
	OR	RR	OR	RR	OR	RR
<i>Categorical volume</i>						
Low volume (< 50 infants)	1.52 ** [1.06,2.18]	1.44 * [0.99,1.89]	1.80 ** [1.09,2.99]	1.45 * [0.98,1.92]	5.43 ** [1.21,27.47]	2.76 ** [1.12,5.25]
Moderate volume (51–100 infants)	1.58 ** [1.01,2.48]	1.49 * [0.92,2.05]	1.88 ** [1.05,3.37]	1.49 * [0.95,2.03]	3.51 *** [2.05,6.16]	2.21 *** [1.57,3.0]
High volume (>100 infants; reference category)	1.00	1.00	1.00	1.00	1.00	1.00
<i>Continuous volume ^a</i>						
Increase by 10 infants	0.96 ** [0.94,0.99]	0.97 ** [0.94,0.99]	0.95 [0.89,1.01]	0.97 * [0.93, 1.01]	0.84 *** [0.78,0.91]	0.9 *** [0.87,0.94]
Increase by 10 infants within volume of < 50 infants	1.03 [0.89,1.19]	1.03 [0.91,1.15]	0.99 [0.82,1.2]	0.996 [0.89,1.1]	0.95 [0.69,1.33]	0.98 [0.83,1.1]
Increase by 10 infants at volume > 50 infants	0.97 [0.94,1.01]	0.98 [0.95,1.01]	0.95 [0.89,1.03]	0.97 [0.93,1.02]	0.83 ** [0.77,0.90]	0.89 *** [0.85,0.93]

Note: The table reports the odds ratios (ORs) and risk ratios (RRs) for the effects of different volume measures on infant mortality under unadjusted and adjusted classical logistic regression and under the IV model (the IV model is also adjusted for all covariates included in the classical adjusted logistic regression). 95% CI are in brackets.

*, **, and *** indicate $p < 0.1$, $p < 0.05$, and $p < 0.01$, respectively.

RRs are defined as mortality risk at low (or moderate) volume divided by mortality risk at high volume. For continuous volume, RRs are obtained from dividing mortality risk at a volume higher than mean volume by 10 infants by mortality risk at mean volume. All analyses are based on the main sample of 4,553 non-transferred infants.

^aThe effects of the continuous volume measure are shown for an increase in the number of VLBW infants by 10 from the mean of continuous volume.