MIS Quarterly

RESEARCH ARTICLE

THE IMPACT OF HEALTH INFORMATION SHARING ON DUPLICATE TESTING

Sezgin Ayabakan
Fox School of Business, Temple University, 1801 Liacouras Walk,
Philadelphia, PA 19122-6083 U.S.A. {ayabakan@temple.edu}

Indranil Bardhan and Zhiqiang (Eric) Zheng
Naveen Jindal School of Management, University of Texas at Dallas, 800 West Campbell Road,
Richardson, TX 75080 U.S.A. {bardhan@utdallas.edu} {ericz@utdallas.edu}

Kirk Kirksey
The University of Texas Southwestern Medical Center, 6303 Forest Park Road,
Dallas, TX 75390 U.S.A. {kirk.kirksey@utsouthwestern.edu}

Appendix

Though our quasi-experiment approach (described in the "Results" section), using a treatment-to-control matching algorithm, is a well-accepted methodology to purge out confounding effects, one may still argue that the $Treatment_h$ variables might still be subject to potential endogeneity. For example, providers with higher duplication rates may be more likely to implement health information sharing technologies. To address the concern of endogeneity, we adopted a control function estimation approach as discussed in Wooldridge (2011) and Rivers and Vuong (1988). The control function approach disintegrates the correlation between endogenous explanatory variables and unobservables affecting the outcome using additional regressors that do not appear in the structural equation (Wooldridge 2010). Suppose y_1 refers to the outcome variable, y_2 refers to endogenous explanatory variable, and $\mathbf{z}$ is the $1 \times L$ vector of exogenous variables, where $\mathbf{z_1}$ is a $1 \times L_1$ strict subvector of $\mathbf{z}$. Then model becomes

$$y_1 = \mathbf{z_1}\boldsymbol{\delta_1} + \alpha_1 y_2 + u_1 \tag{4}$$

with the orthogonality condition of $E(\mathbf{z}'u_1) = 0$. To correct for the endogeneity issue in y_2, we apply a linear projection of y_2 on all other exogenous variables. The reduced form of y_2 can be expressed as

$$y_2 = \mathbf{z}\boldsymbol{\pi_2} + v_2 \tag{5}$$

again with the orthogonality condition of $E(\mathbf{z}'v_2) = 0$. Endogeneity becomes an issue if and only if u_1 is correlated with v_2 which can be expressed as in this linear projection:

$$u_1 = \rho_1 v_2 + e_1 \tag{6}$$

where $\rho_1 = E[v_2 u_1]/E[v_2^2]$. Plugging (6) into (4) gives us the control function model for a probit model

$$y_1 = \mathbf{z_1}\boldsymbol{\delta_1} + \alpha_1 y_2 + \rho_1 v_2 + e_1 \tag{7}$$

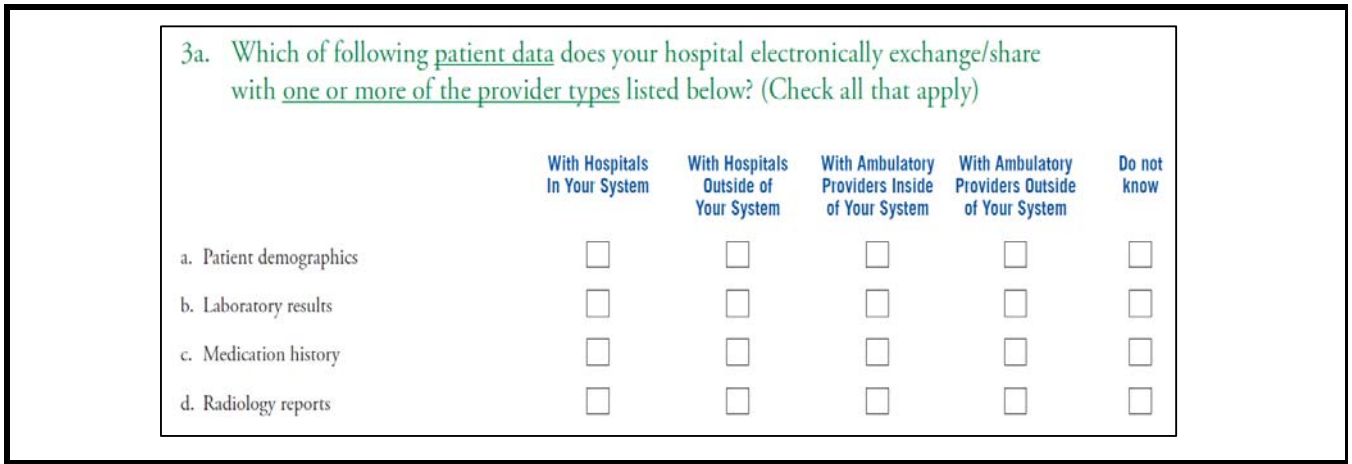
where v_2 is treated as a regressor. Since $E[\mathbf{z_1} e_1] = 0$, $E[v_2 e_1] = 0$ and $E[y_2 e_1] = 0$, model (7) can be estimated in a simple two-step procedure (Wooldridge 2010, 2011). In the first step, we regress y_2 on $\mathbf{z}$, which includes additional regressors that are excluded from $\mathbf{z_1}$. These additional

regressors in **z** will help to breakdown the correlation between u_1 and v_2. Then we estimate the residuals from the first step, $\overline{v_2}$, and plug them into model (7). Finally, we regress y_1 on $\mathbf{z_1}, y_2$ and $\overline{v_2}$ using OLS. The robust standard errors in model (7) can be obtained by bootstrapping to account for first stage estimation (Wooldridge 2011). We estimate model (7) for generic-, intra-, and interorganization information sharing, and include controls for patient payer type, visit type, admission source, age, gender, race, and other provider characteristics.

It remains to be determined as to which variables are to be included in **z** as part of the first stage estimation. Potential candidates should explain the variation in our endogenous variables (i.e., $TreatmentGeneric_h$, $TreatmentIntra_h$, and $TreatmentInter_h$), while they should not be systematically co-determined with *Duplicate_Rate* (Kumar and Telang 2012). One possible variable is the age of a hospital in terms of the number of years that it has been in operation (Age_Clinic_{ht}).[1] Relatively new providers would be more likely to implement health information sharing technologies, while older providers are usually slow adopters of such systems due to the difficulty of replacing legacy systems. At the same time, the age of a provider clinic/facility may not be systematically co-determined with its duplication rate.

We use the two variables, Age_Clinic_{ht}, $Age\_Clinic^2_{ht}$, and the interactions $Age\_Clinic_{ht}*Post_{ht}$, $Age\_Clinic^2_{ht}*Post_{ht}$ as additional variables in the first stage for all three cases, generic-, intra-, and interorganization information sharing. Since the control function approach resembles two-stage least squares (2SLS) estimation, we first check if these additional variables also satisfy the exogeneity and relevance properties of instrument variables in 2SLS (Greene 2011). The exogeneity assumption implies that IVs should be uncorrelated with the error term and the relevance assumption implies that IVs should be correlated with the independent variables (Greene 2011). To test the exogeneity of IVs, we use a test of over-identifying restrictions via Hansen's (1982) commonly employed Hansen-Sargan test. The Sargan statistic is distributed as χ^2 with degrees of freedom equal to the number of exclusion restrictions less the number of endogenous variables. Accordingly, we obtain $\chi^2_{(3)} = 5.474$ with $p = 0.14$, $\chi^2_{(3)} = 0.068$ with $p = .99$, and $\chi^2_{(3)} = 0.371$ with $p = 0.94$ respectively for generic, intra-, and interorganization information sharing variables. These statistics fail to reject the null hypothesis and implies that the instruments that we have selected are valid.

To test the relevance assumption, we employ a weak identification test on the IVs using the Anderson canonical correlations likelihood-ratio test statistic in which the null hypothesis suggests that the model is under identified or instruments are weak. Overall, the statistic is distributed as chi-squared with degrees of freedom equal to number of instruments less the number of regressors plus one. We report the statistics as $\chi^2_{(4)} = 600.9$ with p = 0.00, $\chi^2_{(4)} = 215.9$ with $p = 0.00$, and $\chi^2_{(4)} = 497.0$ with $p = 0.00$ respectively for generic-, intra-, and interorganization information sharing variables. Our results suggest that the IVs are not weak.



3a. Which of following patient data does your hospital electronically exchange/share with one or more of the provider types listed below? (Check all that apply)

	With Hospitals In Your System	With Hospitals Outside of Your System	With Ambulatory Providers Inside of Your System	With Ambulatory Providers Outside of Your System	Do not know
a. Patient demographics	☐	☐	☐	☐	☐
b. Laboratory results	☐	☐	☐	☐	☐
c. Medication history	☐	☐	☐	☐	☐
d. Radiology reports	☐	☐	☐	☐	☐

Figure A1. AHA IT Supplement 2012 Data Questions on Radiology and Laboratory Tests

[1]We manually collected information on the age of the outpatient clinic, measured as the number of years it had been in operation.

Table A1. Correlation Matrix

Variable	V#	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
DupRateRad	***V1***	1													
DupRateLab	***V2***	0.31	1												
InsuranceType†	***V3***	-0.04	-0.05	1											
VisitType†	***V4***	-0.38	-0.19	0.05	1										
AdmissionSource†	***V5***	0.12	0.06	0.01	0.39	1									
Female†	***V6***	-0.08	-0.04	0.07	0.11	-0.03	1								
White†	***V7***	-0.13	0.002	0.12	0.28	-0.06	0.09	1							
Age	***V8***	-0.1	-0.01	0.40	0.13	0.03	0.21	0.43	1						
CMI	***V9***	0.02	-0.07	0.01	0.06	-0.02	-0.19	-0.41	-0.24	1					
Teaching†	***V10***	0.10	-0.05	-0.11	-0.11	0.18	-0.30	-0.59	-0.50	0.77	1				
Urban†	***V11***	0.23	-0.11	0.004	-0.37	0.29	-0.29	-0.62	-0.50	0.66	0.59	1			
Log(Beds)	***V12***	-0.07	-0.11	-0.005	0.18	-0.07	-0.08	-0.28	-0.16	0.56	0.94	0.17	1		
Days_Between_Visits	***V13***	0.07	-0.06	0.009	-0.13	0.13	-0.13	-0.20	-0.15	0.25	0.41	0.53	0.1	1	
ER_Charge†	***V14***	0.53	0.35	-0.10	-0.91	0.22	-0.16	-0.40	-0.27	0.01	0.34	0.55	-0.05	0.21	1

†These variables are categorical variables, as shown below.

InsuranceType: *MedicareB*, *Selfpay*, *Private*, *Other_Insurance*, *MedicareA*, *Medicaid.*

VisitType: *Elective_Visit*, *Emergency_Visit*, and *Other_Visit.*

AdmissionSource: *Referral_source*, *Transfer_source*, and *Other_source.*

Between continuous variables, Pearson correlations are reported.

Between continuous and categorical variables, polyserial correlations are reported.

Between categorical variables, polychoric correlations are reported.

Correlations > 0.4 are highlighted in grey boxes.

Table A2. Results of Control Function Approach (30-Day Time Window)				
	All-Providers		Interorganization	
Treatment	-0.136	(0.190)	0.130	(0.193)
Post	0.118	(0.134)	0.200	(0.163)
Treatment*Post	-0.111	(0.173)	-0.618***	(0.224)
Selfpay	0.00801	(0.212)	0.0695	(0.214)
Private	-0.0337	(0.244)	-0.178	(0.267)
Other_Insurance	-0.161	(0.173)	-0.187	(0.163)
MedicareA	-0.357**	(0.169)	-0.307*	(0.166)
Medicaid	0.161	(0.229)	0.227	(0.197)
Emergency_Visit	0.275**	(0.126)	0.455**	(0.182)
Other_Visit	0.379***	(0.115)	0.334***	(0.119)
Transfer_Source	0.874*	(0.478)	-0.155	(0.494)
Other_Source	-0.444*	(0.257)	-0.688**	(0.288)
Female	-0.120*	(0.0656)	-0.109*	(0.0658)
White	0.0157	(0.0859)	0.0269	(0.0817)
Age	0.00451*	(0.00272)	0.00367	(0.00300)
CMI	0.0497	(0.191)	0.256	(0.220)
Teaching	-0.415**	(0.182)	0.0629	(0.188)
Urban	0.331**	(0.140)	0.114	(0.186)
Log(Beds)	0.0183	(0.0826)	-0.172**	(0.0752)
Days_Between_Visits	0.000209	(0.00364)	0.000673	(0.00449)
ER_Charge	0.305***	(0.0819)	0.204**	(0.0960)
Residuals	0.509***	(0.190)	0.434**	(0.189)
Correlation Coef. (R^2)	0.116		0.119	
AIC	3.13		3.10	
LogLikelihood	-2353.15		-1907.83	
N	1518		1246	
Standard Dev(Y)	0.513		0.502	

Marginal effects at the mean values of the variables are reported. Bootstrap standard errors (200 replications) are reported in parentheses.
*p < 0.10, **p < 0.05, ***p < 0.0

Table A3. Heckman Correction Results with Mills Ratio Estimation

	All-Providers		Inteorganization	
Treatment	-0.178	(0.123)	0.238*	(0.133)
Post	-0.0387	(0.104)	0.253**	(0.125)
Treatment*Post	-0.0545	(0.132)	-0.654***	(0.155)
Mills_Ratio	0.211***	(0.0703)	0.0647	(0.0789)
Selfpay	0.122	(0.155)	0.155	(0.154)
Private	-0.0689	(0.181)	—	—
Other_Insurance	-0.137	(0.123)	-0.108	(0.113)
MedicareA	-0.261**	(0.121)	-0.191*	(0.112)
Medicaid	-0.00852	(0.157)	0.108	(0.156)
Emergency_Visit	0.226**	(0.0963)	0.345***	(0.128)
Other_Visit	0.262***	(0.0843)	0.189*	(0.0970)
Transfer_Source	0.784*	(0.418)	-0.00710	(0.402)
Other_Source	-0.109	(0.199)	-0.392	(0.276)
Female	-0.0667	(0.0485)	-0.0349	(0.0540)
White	0.0464	(0.0594)	0.0222	(0.0641)
Age	0.00104	(0.00200)	0.000257	(0.00214)
CMI	0.186	(0.165)	0.428**	(0.176)
Teaching	-0.327***	(0.121)	-0.0173	(0.153)
Urban	0.264**	(0.108)	0.00754	(0.156)
Log(Beds)	-0.0293	(0.0526)	-0.217***	(0.0601)
Days_Between_Visits	0.00393***	(0.000964)	0.00427***	(0.00105)
ER_Charge	0.251***	(0.0625)	0.165**	(0.0704)
Constant	-0.340	(0.354)	0.183	(0.391)
Correlation Coef. (R^2)	0.087		0.095	
F Statistic	3.12		3.09	
N	2568		2080	
Standard Dev(Y)	0.515		0.509	

First stage involved probit estimation using additional variables Age_Clinic_{ht}, $Age\_Clinic2_{ht}$, Age_Clinic_{ht}*$Post_{ht}$, $Age\_Clinic2_{ht}$*$Post_{ht}$ as exclusion restriction.

Robust standard errors are reported in parentheses.

*p < 0.10, **p < 0.05, ***p < 0.0

Table A4. Results of Control Function Estimation with Information Sharing of Patient Medication History

	All-Providers		Interorganization	
Treatment	-0.138	(0.110)	-0.124	(0.106)
Post	0.127	(0.0827)	0.142*	(0.0820)
Treatment*Post	-0.120	(0.131)	-0.136	(0.121)
Selfpay	0.103	(0.106)	0.102	(0.102)
Private	-0.394***	(0.108)	-0.390***	(0.118)
Other_Insurance	0.101	(0.0807)	0.101	(0.0831)
MedicareA	-0.176**	(0.0694)	-0.174**	(0.0749)
Medicaid	0.138	(0.109)	0.134	(0.108)
Emergency_Visit	0.383***	(0.0695)	0.390***	(0.0647)
Other_Visit	0.249***	(0.0652)	0.247***	(0.0696)
Transfer_Source	0.235	(0.305)	0.232	(0.344)
Other_Source	-0.339***	(0.0978)	-0.349***	(0.0929)
Female	0.0484	(0.0318)	0.0483	(0.0294)
White	0.0731*	(0.0432)	0.0745*	(0.0438)
Age	-0.00492***	(0.00131)	-0.00496***	(0.00151)
CMI	0.178	(0.154)	0.194	(0.160)
Teaching	-0.727***	(0.0733)	-0.734***	(0.0704)
Urban	0.406***	(0.100)	0.396***	(0.0936)
Log(Beds)	0.132***	(0.0349)	0.130***	(0.0350)
Days_Between_Visits	0.00144**	(0.000696)	0.00145**	(0.000704)
ER_Charge	0.361***	(0.0598)	0.366***	(0.0566)
Residuals	0.134	(0.124)	0.119	(0.117)
Correlation Coef. (R^2)	0.12		0.103	
AIC	3.08		3.08	
LogLikelihood	-8959.80		-8929.93	
N	5828		5811	
Standard Dev(Y)	0.497		0.497	

Marginal effects at the mean values of the variables are reported. Bootstrap standard errors (200 replications) are reported in parentheses.
*p < 0.10, **p < 0.05, ***p < 0.0

Table A5. Glossary of Acronyms	
ACA	Affordable Care Act
AHA	American Hospital Association
CBC	Complete blood count
CHF	Congestive heart failure
CMI	Case mix index
CMS	Center for Medicare and Medicaid
COPD	Chronic obstructive pulmonary disease
CPOE	Computerized provider order entry
CPT	Current Procedural Terminology
CT	Computed tomography
DFWHC	Dallas–Fort Worth Hospital Council
DICOM	Digital imaging and communications in medicine
DID	Difference in difference
ECG	Echocardiography
EDI	Electronic Data Interchange
EHR	Electronic health record
EMR	Electronic medical record
ER	Emergency room
FFS	Fee-for-service
HIE	Health information exchange
HITECH	Health Information Technology for Economic and Clinical Health Act
ICD	International Classification of Diseases
LIS	Laboratory information systems
LOINC	Logical observation identifiers names and codes
MRI	Magnetic resonance imaging
OECD	Organization for Economic Cooperation and Development
ONC	Office of the National Coordinator
PACS	Picture archival and communication systems
PHR	Personal health records
REMPI	Regional master patient index
RIS	Radiology information systems
VIF	Variance inflation factor

References

Greene, W. H. 2011. *Econometric Analysis* (7th ed.), New York: Pearson.

Hansen, L. P. 1982. "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica: Journal of the Econometric Society* (50:4), pp. 1029-1054.