

Using Clinical Data Repositories to Assess the Clinical and Financial Burden of Disease: The Example of Mitral Regurgitation

Abhishek Khemka^{1*}, Irmira Gradus-Pizlo⁴, Richard Kovacs⁴, Wanzhu Tu⁴, Ross Hayden², Abdullah A Masud⁴, George J Eckert⁴ and William M Tierney³

¹Irmira Gradus-Pizlo Richard Kovacs, Wanzhu Tu, Abdullah Masud, George Eckert – Indiana University School of Medicine, Indianapolis, IN, USA

²Regenstrief Institute, Indianapolis, IN, USA

³Dell Medical School, University of Texas, Austin, TX, USA

⁴Indiana University School of Medicine, Indianapolis, USA

Abstract

Although there have been tremendous advances in understanding various disease outcomes, there are significant gaps and the associated costs to investigate disease burden can be exorbitant. Clinical data repositories can be a valuable aid for analysing patient and disease characteristics in a faster and most cost-effective manner. We offer our own example, using mitral regurgitation as the illustration of a disease process that was identified through the use of a clinical data repository in a subset of patients, matched with a control population, and then analysed for clinical and financial factors. Increasing adoption of digital systems to store and analyse large volumes of data paired with incentives by the government and various health systems makes the current environment ripe for an explosion of big data to help guide clinical decision making.

Keywords: Mitral regurgitation; Clinical data repositories; Clinical burden of disease; Electronic Health Records (EHR)

Introduction

Since 1960, the spell of using a digital computer system to record, store, and analyse data has enthralled clinicians [1]. There have been many hurdles and to date, the use of clinical data repositories for exploring diseases and their outcomes has been limited due to some of the following reasons [2,3]:

- Non-standardized coding schemes.
- Missing, misclassified or incomplete information.
- Lack of consensus to define specific conditions and outcomes.
- Lack of clinical knowledge among analysts and lack of analytic prowess among clinicians.
- Lack of intercommunication among hospitals, practices and Electronic Health Records (EHR).

However, the era of big data may have finally arrived with the diffusion of EHRs and the mandate from the HITECH Act [4]. Implementation and use of EHRs is now >75% among hospitals and >50% among physicians due to mandates for having and demonstrating use. Physicians and practices that refrain from adoption of EHRs face penalties [5,6]. There has been increasing interoperability of both clinical and non-clinical data. Using this information provides practice based affirmation for evidence based care [7]. The National Institute of Health has dedicated over \$100 million into the Big Data to Knowledge Initiative to extract knowledge from these myriad data [8]. There have been significant advances leading to refined processes to extract data and guide clinical decision making [9-11]. It is in this environment that we endeavoured to better understand repercussions of mitral regurgitation and offer our experience of using a clinical database as a possible guide for future studies.

Over five million Americans are diagnosed with heart valve disease with an estimated three million Americans affected by Mitral Regurgitation (MR) [12,13]. MR is the most common valvular abnormality with an estimated 70% of healthy adults having trivial MR

on transthoracic echocardiography. Surveys such as the Framingham Heart Study and the Strong Heart Study found at least mild MR in 19% of study patients with 1.9% having moderate MR and 0.2% with severe MR [14,15]. MR is an ideal target for demonstrating the value of clinical data repositories because disease progression is pervasive, expensive, disabling, and relatively easily identifiable [16]. There are substantial medical, device based and surgical options that can serve as cure or palliation for patients. To our knowledge, the clinical and financial burden on patients has not been adequately demonstrated. We hope our experiences as demonstrated will highlight the rapidity and efficiency with which researchers can retrieve, analyse and present data to answer clinical questions. The partnership also emphasizes the mutual benefit that can occur when academia and industry collaborate.

Study design and oversight

The Clinical and Health Economic Burden of Mitral Regurgitation study is a retrospective study that utilized clinical data repositories to assess the clinical and financial burden of functional mitral regurgitation in a general population in central Indiana. This study was approved by Institutional Review Board of Indiana University (IU). Data were extracted from two independent health systems with different primary care and referral networks:

IU health: A network of 18 hospitals and affiliated outpatient practices with 1,500 IU Health physicians who deliver both primary and specialty cardiac care to more than three million patients annually throughout the state of Indiana. The hospitals within downtown

*Corresponding author: Abhishek Khemka, MD, MBA, Indiana University School of Medicine, Indianapolis, USA, Tel: 888-484-3258; E-mail: akhemka@iu.edu

Received May 24, 2017; Accepted June 19, 2017; Published June 26, 2017

Citation: Khemka A, Gradus-Pizlo I, Kovacs R, Tu W, Hayden R, et al. (2017) Using Clinical Data Repositories to Assess the Clinical and Financial Burden of Disease: The Example of Mitral Regurgitation. J Health Med Informat 8: 266. doi: 10.4172/2157-7420.1000266

Copyright: © 2017 Khemka A, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Indianapolis function as the quaternary referral centre for paediatric and adult patients with advanced pathologies.

Eskenazi health: The fourth largest safety net health system in the nation with a 360 bed hospital, nine federally qualified community health centres and mental health clinics in Indianapolis, IN. Eskenazi cares for 1 million patients annually within Marion County with a disproportionate share being poor and underinsured.

Both systems allow for training of medical students and residents from the IU School of Medicine. The combination of the two systems allows for an inclusion of a varied patient population representing a multitude of ethnicities, socioeconomic backgrounds, education, and insurance coverage. Each health system made their facility available for clinical research and provided the Regenstrief Institute investigators with full access to the data repositories and resources. The Regenstrief Institute (RI) is a non-profit research organization dedicated to studying and improving health systems and is affiliated with Indiana University and Purdue University. These data resources are stored in databases managed by RI administrators on servers managed by Indiana University’s Information and Technology Service. Synapse Cardiovascular from Fujifilm was used as the database for identifying patients with MR. Synapse is a web-based, multi-modality software system that allows for the viewing, interpretation and reporting of echocardiograms and also allows for storage.

This study was funded by Abbott Vascular which markets a device for repairing mitral valves. In order to minimize the likelihood that the funding source would introduce bias into the methods, analysis, or interpretation of results, we closely adhered to the “Principles and Benchmarks for Ethically Credible Academic-Industry Partnerships” [17].

Study population

Patients were eligible for this study if they had an echocardiogram performed between January 1, 2010 and December 15, 2015 with findings of moderate, moderate-to-severe, or severe MR concomitant with left atrial dilation. The severity was defined based on two-dimensional and Doppler echocardiographic parameters as defined by the American Society of Echocardiography [18]. Patients were excluded if they met any of the prespecified criteria (Table 1).

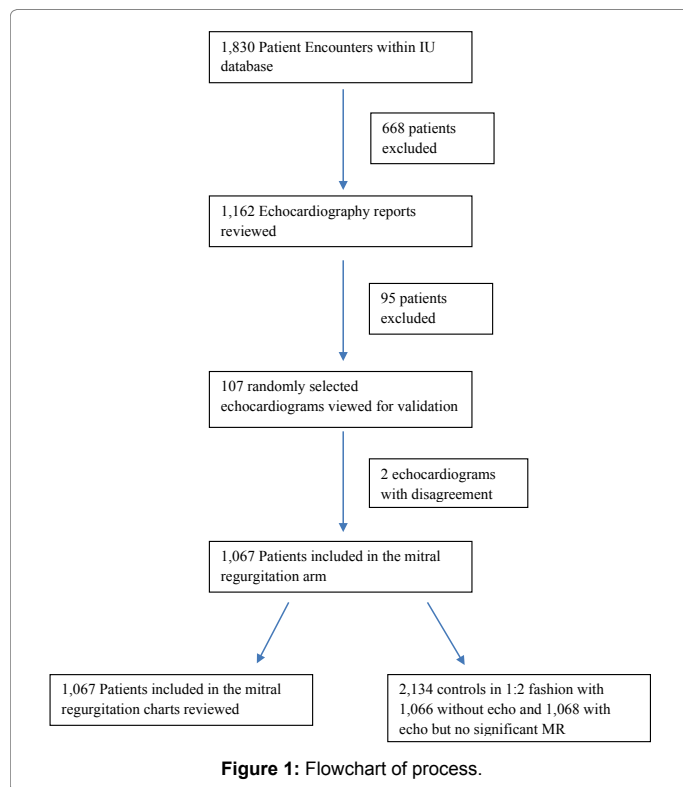
Study procedures

To adequately power the study, researchers estimated needing about 1,000 patients with moderate or worse mitral regurgitation along with 2,000 control patients in a 1:2 fashion:

- 1,000 patients with echocardiograms showing mild or less mitral regurgitation by echocardiogram matched by age and sex.
- 1,000 patients matched by age and sex but with no recorded echocardiogram within IUH’s echocardiography database.

Researchers within the echocardiography lab at IU Health did a preliminary search with the above criteria and identified a test cohort of 541 patients with moderate or worse MR with abbreviated reports indicating the degree of MR and size of left atrium. 171 patients were excluded based on the prespecified criteria after examining the abbreviated reports. The study investigators then reviewed 50 randomly selected echocardiograms from the remaining 370 echocardiograms to confirm the veracity of the report findings, degree of MR, and size of the left atrium. After confirmation of the abbreviated reports and the overall process, additional searches were run identifying a total of 1,830 patients (inclusive of the original search) with moderate or worse

MR of whom 668 were excluded based on the specified criteria (Figure 1). Following this, cardiologists reviewed the complete reports of the remaining echocardiograms (1,162) and rejected another 95 patients with findings that would confound from the concentration of MR (Table 1).



Initial Exclusion Criteria	Patients excluded
Age <18 or Age >85	348
Lack of left atrial dilation	250
Severity of mitral regurgitation less than moderate	57
Location of study (if not from prespecified hospitals including IU Methodist, IU University, IU West, IU Saxony, IU North)	13
Total	668
Exclusions after further review	Patients excluded
Moderate or worse Aortic Stenosis	24
Prosthetic Heart Valve	22
Ventricular Assist Device	10
Moderate to Severe or Worse Aortic Regurgitation	9
Valvular Vegetation or Mass	9
Mitral Valve Prolapse	7
Hypertrophic Cardiomyopathy	3
Mitral Stenosis	2
Rheumatic Mitral Valve	2
Tetralogy of Fallot	2
Mitral Valve Ring	1
Flail Leaflet	1
Cardiac Transplant	1
Large Pericardial Effusion	1
Papillary Rupture	1
Total	95

Table 1: Exclusion criteria.

Next, through a random computer generator, a 10% sample of patients was selected to validate the echocardiograph reports by reviewing the images, including size of left atrium and severity of their MR as well as assessing for confounding factors [19]. If there was disagreement, the study was adjudicated by two senior cardiologists. Of the 1,067 reports included, 107 sample echocardiograms were reviewed. One patient was found to have a bioprosthetic aortic valve not mentioned in the report. A second patient was reported to have moderate MR but the images did not have Doppler profiles saved. However, comments in the report for this study noted it was done at the bedside with cardiologist present reviewing the images and we suspect the Doppler images showed at least moderate MR but those images were not saved. The remaining 105 were judged to meet all the inclusion criteria. All 1,067 patients were kept in the MR subset.

Researchers at RI then searched the database of the Indiana Network for Patient Care (INPC) database, inclusive of both health systems, to find a matched set of patients in 1:2 design with two control patients for every one with MR [20]. The INPC is a local health information infrastructure with 38 major hospital systems including over 100 hospitals as well the county and state public health departments and Indiana Medicaid representing over nine billion data elements [21,22]. The network allows for rapid access to laboratory values, radiology, cardiology, pathology and other specialty reports as well as dictated notes and documents from providers. The data is available for analysis by any investigator and the researchers at RI will also provide an initial feasibility study to help with grant proposals [23]. The INPC was used to extract demographic information, vital signs, cardiovascular and non-cardiovascular comorbid conditions, drug treatments, hospitalizations, and death (Table 2). The participating hospitals are hard wired to the INPC. Internal analysis has shown the time lag between a patient discharge from a hospital or ED and storing of those data in the INPC is less than 3 seconds. Data from lab tests and imaging studies including the echocardiogram results are also instantaneous, in that as the report is generated, it is added onto the INPC database. This is in contrast to most health networks that are not hard wired and capture data in structured increments such as nightly or weekly but not immediately.

A major limitation for all institutions including our own is a paucity of structured data or missing data. Researchers at Regenstrief used a proprietary NLP Data Extraction Providing Targeted Healthcare (nDepth) tool which allows for mining the actual textual reports for data to identify patients as well as to confirm certain conditions and write the information back into the data repository. For example, applications by nDepth include finding patients with metastatic melanoma, identifying pre-diabetic patients for clinical trials, finding reasons for refusal of osteoporosis medications, mapping patient trajectory following cancer treatment and detecting treatment failure in insomnia [24].

Death was noted and confirmed by querying the Indiana State Department of Health's death certificate files maintained by the RI, the family search database maintained by Church of Jesus Christ of Latter Day Saints, and the National Death Index by the National Centre for Health Statistics.

To assess the financial burden of mitral regurgitation claims data from the INPC and charge data provided by the local institutions was analysed for cost. The sources of claims data within the INPC include Indiana Medicaid (almost 4 million patients) and the largest health insurance company in Indiana, Anthem Inc., which has been an INPC member since 2002 and has data on more than more than 7 million

Demographics	Encounter setting
Gender	Emergency Department
Race	Outpatient
Age	Inpatient
Height	
Weight	Medications
BMI	Drug class
	Drug Name
	Dose
Insurance Type	Adherence (based on pharmacy refills)
Commercial	
Medicare	
Medicaid	Procedures Performed
Self-pay	Right Heart Catheterization
	Left Heart Catheterization
	Corrective heart valve procedure
Substance Use	
Tobacco Use	
Alcohol Abuse	Outcomes
	Hospitalization for heart failure
Lab Values and Clinical Outcomes (closest to index date)	Hospitalization for cardiovascular problem (heart failure, myocardial infarction, acute coronary syndrome)
B-type Natriuretic Peptide	Hospitalization for any cause
Creatinine	Death
Potassium	
Sodium	Comorbidities
Ejection Fraction	Diabetes
Systolic Blood Pressure	Heart Failure
Diastolic Blood Pressure	Kidney Disease
Heart Rate	Hypertension
New York Heart Association Class	Stroke
	Atrial Fibrillation
Charge Data	Ischemic Heart Disease
Inpatient charge data	Pulmonary Hypertension
Outpatient charge data	

Table 2: Descriptive statistics for MR and control populations.

patients in the INPC. Although charge data is not a true reflection of cost, estimation of true hospital costs can be made from cost-to-charge ratios that all hospitals must provide to CMS annually for each hospital cost centre [25,26]. Furthermore, true hospital costs have been estimated using cost: charge ratios through the INPC database [27,28]. One major drawback with this accelerated timeline was that we were unable to access Medicare charge data.

Analysis Methods

Data were first cleaned to correct or remove out-of-range values. The resulting descriptive statistics of the clinical descriptors and outcome measures were then reviewed by the project's physicians for validity and generalizability.

Patient characteristics were compared among the MR and control groups using chi-square and t-tests as appropriate for categorical and continuous variables, respectively. Kaplan-Meier survival curves were plotted for times to death, first hospital admission, first ER visit, re-admission, and repeat ER visit, using the date of MR diagnosis as the reference index date for the MR cases, the date of the echocardiogram as the reference index date for the echocardiogram controls, and reference index date for the non-echo controls. Cox proportional hazards models were used to compare the survival curves. Poisson

regression was used to model the hospital admission counts, ER visit counts, and outpatient visit counts.

A subset of patients was used in order to calculate the charge data. These patients must have had an insurance identifier and must have had the potential to contribute a full year's worth of claims data to analysis. These patients' records were then pulled and charges for subsequent encounters, procedures, and diagnosis calculated for the one year follow-up period. For this portion of the analysis, the cohort was partitioned to those patients who were either on Medicaid and/or Anthem and had valid charges after their index date (Table 3).

Medicare charges were not available for this analysis. Also, majority of the patients with data were from Medicaid and very few from commercial insurance. Due to this limitation, the analyses of the charge data subset are not representative of the full cohort. While the average charges were found to be higher for the MR patients compared to the two sets of controls, the analysis is not representative of the whole population and therefore it is difficult to draw conclusions.

Patients with charge data represented the group which was the distribution of the charges is highly skewed. Analyses were performed 3 ways: a) nonparametric Wilcoxon Rank Sum tests, b) Wald tests for log-normal data with zeros, c) linear regression for log (charges) with patient characteristics included as covariates. Associations of patient characteristics with log (charges) were analysed for each cohort using linear regression.

Discussion

This is a unique study highlighting the speed and efficiency with which data can be analysed from clinical data repositories as well as the complex interplay of various institutions that capitalize on their individual expertise to provide the answers to the proposed questions. It shows a reproducible, systematic route that focuses on a distinct disease process to delineate the clinical and financial burden in the studied population. It also highlights the ability of academia and industry to work together ethically without ignoring the inherent conflicts of interest [29].

The fundamental goal was to define the overall clinical and financial burden of mitral regurgitation. There are a myriad of options

for diagnoses and treatment of MR, and yet evidence suggests that the appropriate intervention is often not performed when indicated due to perceived adverse clinical and financial impacts of such surgery [30]. Within an 8 month time-frame clinicians and biostatisticians met bimonthly and with representatives from Abbott Vascular monthly to discuss the overall goals, identify patients, verify their inclusion, review and analyse the descriptive statistics, and interpret the overall findings (Figure 2). This retrospective study of 1,067 patients with moderate or worse MR took 37 weeks from initial data retrieval to generation of the final analysis dataset. IU's cardiovascular service reads 21,000 echocardiograms per year at five hospitals and identifies on average 320 new patients with echocardiographic evidence of moderate or worse MR per year with no confounding cardiac factors. Therefore, a prospective study of this size would take at least 3.3 years or need to include 3 times as many incident echoes for prospective data collection to be completed in one year.

MR is a prevalent disease that if discovered promptly and treated appropriately, would have significant positive impact for the quality and quantity of life for individual patients as well as the population as a whole due to improved resource utilization. In an upcoming manuscript as well as an abstract presented at the American College of Cardiology, we show that patients with moderate or worse MR have significantly lower survival probability at three years compared to controls. Patients with moderate MR have similar survival to patients with moderate to severe MR and severe MR. The moderate or worse MR cohort was responsible for significantly more inpatient encounters of which the majority was due to heart failure or heart failure plus other heart disease. Finally, based on the limited subset of patients with charge data available, the MR group had a 2.5 times higher rate of healthcare utilization than controls along with 52% higher cost of care. We hope the findings of this study will highlight to clinicians the complexity of patients with mitral regurgitation. Currently, most patients with MR are managed conservatively unless it is severe and they are symptomatic. This study may help steer best practices to intervene sooner which may help improve a patient's quality of life and overall mortality. This study helps to define that burden to guide clinicians and public health officials in an era where there is increased scrutiny on cost and evidence based practice. We propose that this method is ideal for disease entities that are readily identifiable have validated treatment options, and yet where the appropriate intervention is irregularly employed.

One significant limitation was the dearth of charge data that was captured within the population. Although the charge data provided some evidence as to the costs associated with mitral regurgitation, it is incomplete. However, all data networks will suffer from this constraint and we can only make inferences based on the data provided while recognizing that this is not proof [31].

SUBSET	PROVIDER	
	Medicaid	Anthem
MR Cases (n=251)	89.2%	10.8%
Echo Control (n=212)	83.5%	16.5%
No Echo Control (n=115)	45.2%	54.8%

Table 3: Breakdown of insurance provider.

Activity	Deliverable	Estimated schedule							
		Months							
		1	2	3	4	5	6	7	8
Signed contract by Abbott, Regenstrief and IU	Signed contract	X							
Kickoff meeting, review abott materials	Final protocol	X							
Analysis dataset created and validated	Report with descriptive statistics		X	X	X				
Outcome and descriptive variables defined	Report with descriptive statistics					X	X		
Economic and clinical burden of MR analyses completed	Manuscripts submitted to journals							X	
Final report and power point presentation	Final report and slides to Abbott.								X

Figure 2: Timeline.

Acknowledgments

We would like to thank our colleagues Cynthia Lim and Jennifer Gatz at the Regenstrief Institute and Zach Richardson with the IU Health echocardiography lab for their invaluable insight and expertise.

Funding

This study was funded in part by a grant through Abbott Vascular, Inc., IL, USA.

References

1. Schenthal JE, Sweeney JW, Nettleton W (1960) Clinical application of large-scale electronic data processing apparatus. *JAMA* 173: 6-11.
2. Angus DC (2016) Electronic health records as sources of research data-reply. *JAMA* 315: 202.
3. Madden JM, Lakoma MD, Rusinak D, Lu CY, Soumerai SB (2014) Missing Clinical and Behavioral Health Data in a Large Electronic Health Record System. *J Am Med Informat Ass.*
4. The American Recovery and Reinvestment Act (2009) Pub L. No. 111-5, Stat. 123.
5. Adler-Milstein J, DesRoches CM, Kralovec P, Foster G, Worzala C (2015) Electronic Health Record Adoption in US Hospitals: Progress Continues, But Challenges Persist. *Health Aff* 34: 2174-2180.
6. DesRoches C (2015) Progress and Challenges in Electronic Health Record Adoption: Findings from a National Survey of Physicians. *Ann Intern Med* 162: 396.
7. Feldman B, Martin EM, Skotnes T (2012) Big data in healthcare: hype and hope.
8. NIH Big Data to Knowledge (BD2K) Biomedical data science training coordination center.
9. Abouzahra M, Sartipi K, Armstrong D, Tan J (2014) Integrating data from ehrs to enhance clinical decision making: the inflammatory bowel disease case. *Comp-Bas Med Sys* 1: 531-532.
10. Weber GM, Kohane IS (2013) Extracting physician group intelligence from electronic health records to support evidence based medicine. *PLoS ONE* 8: e64933.
11. Brilakis ES, O'Donnell CI, Penny W (2016) Percutaneous coronary intervention in native coronary arteries versus bypass grafts in patients with prior coronary artery bypass graft surgery: insights from the veterans affairs clinical assessment, reporting, and tracking program. *J Am Coll Cardiol Intv* 9: 884-893.
12. Nkomo V, Gardin M, Skelton T (2006) Burden of valvular heart diseases: a population-based study. *Lancet* 2:1005-1011.
13. Borger MA, Alam A, Murphy PM (2006) Chronic ischemic mitral regurgitation: Repair, replace or rethink? *Ann Thorac Surg* 81: 1153-1161.
14. Jones EC, Devereux RB, Roman MJ, Liu JE, Fishman D, et al. (2001) Prevalence and correlates of mitral regurgitation in a population based sample. *Am J Cardiol* 87: 298-304.
15. Singh JP, Evans JC, Levy D, Larson MG, Freed LA, et al. (1999) Prevalence and clinical determinants of mitral, tricuspid, and aortic regurgitation (the Framingham Heart Study). *Am J Cardiol* 86: 897-902.
16. Kang DH, Kim JH, Rim JH, Kim MJ, Yun SC, et al. (2009) Comparison of early surgery versus conventional treatment in asymptomatic severe mitral regurgitation. *Circulation* 119: 797-804.
17. Meslin EM, Rager J, Schwartz PH (2015) Benchmarks for ethically credible partnerships between industry and academic health centres: beyond disclosure of financial conflicts of interest. *Clin Transl Med* 4: 36.
18. Zoghbi WA, Enriques-Sarano M, Foster E, Grayburn PA, Kraft CD, et al. (2003) Recommendations for evaluation of the severity of native valvular regurgitation with two-dimensional and doppler echocardiography. *J Am Soc Echocardiogra* 16: 777-802.
19. <https://www.random.org/integers/>
20. Biondich PG, Grannis SJ (2004) The Indiana network for patient care: an integrated clinical information system informed by over thirty years of experience. *J Public Health Manag Pract Suppl*: pp: S81-S86.
21. McDonald CJ, Overhage JM, Barnes M, Schadow G, Blevins L, et al. (2014) The Indiana network for patient care: a working local health information infrastructure. *Health Aff* 24: 1214-1220.
22. <http://www.ihie.org/docs4docs>
23. <http://www.regenstrief.org/resources/regenstrief-data-core/>
24. <http://www.regenstrief.org/resources/ndepth/>
25. Schwartz M, Yound DW, Siegrist R (1995) The ratio of costs to charges: how good a basis for estimating costs? *Inquiry* 32: 476-481.
26. Schimmel VE, Alley C, Heath AM (1987) Measuring costs: Product line accounting versus ratio of cost to charges. *Top Health Care Financ* 13: 76-86.
27. Tierney WM, Miller ME, Overhage JM, McDonald CJ (1993) Physician inpatient order writing on microcomputer workstations: Effects on resource utilization. *JAMA* 269: 379-383.
28. Tierney WM, Fitzgerald JF, Miller ME, James MK, McDonald CJ (1995) Predicting inpatient costs with admitting clinical data. *Med Care* 33: 1-14.
29. Tierney WM, Meslin EM, Kroenke K (2015) Industry support of medical research: Important opportunity or treacherous pitfall? *J Gen Int Med* 31: 228-233.
30. Mirabel M, lung B, Baron G (2007) What are the characteristics of patients with severe, symptomatic, mitral regurgitation who are denied surgery? *Eur Heart J* 28: 1358-1365.
31. Gradus-Pizlo I, Khemka A, Kovacs R, Tu W, Hayden R, et al. (2017) Understanding the health burden of mitral regurgitation through clinical data repositories. Presentation at American College of Cardiology Annual Convention, Washington, DC, USA.