



ADB Working Paper Series

**LONG-TERM CARE AND PAY-FOR-
PERFORMANCE PROGRAMS**

Edward C. Norton

No. 680
March 2017

Asian Development Bank Institute

Edward C. Norton is a professor at the University of Michigan and a research associate at the National Bureau of Economic Research.

The views expressed in this paper are the views of the author and do not necessarily reflect the views or policies of ADBI, ADB, its Board of Directors, or the governments they represent. ADBI does not guarantee the accuracy of the data included in this paper and accepts no responsibility for any consequences of their use. Terminology used may not necessarily be consistent with ADB official terms.

Working papers are subject to formal revision and correction before they are finalized and considered published.

The Working Paper series is a continuation of the formerly named Discussion Paper series; the numbering of the papers continued without interruption or change. ADBI's working papers reflect initial ideas on a topic and are posted online for discussion. ADBI encourages readers to post their comments on the main page for each working paper (given in the citation below). Some working papers may develop into other forms of publication.

ADB recognizes "China" as the People's Republic of China, and "Korea" as the Republic of Korea. In this publication, "\$" refers to US dollars. Unless otherwise stated, boxes, figures, and tables without explicit sources were prepared by the authors.

Suggested citation:

Norton, E. C. 2017. Long-term Care and Pay-for-Performance Programs. ADBI Working Paper 680. Tokyo: Asian Development Bank Institute. Available: <https://www.adb.org/publications/long-term-care-and-pay-performance-programs>

Please contact the authors for information about this paper.

Email: ecnorton@umich.edu

Yubraj Acharya, Young Do, Eric Finkelstein, Chie Hanaoka, Toshi Iizuka, Joyce Koh, Ying-Chun Li, Jui-fen Rachel Lu, Marcus Ong, Hua Wang, and Zhou Yang provided information on pay-for-performance programs in Asia. John Syrjamaki cleaned the Medicare data set. Yoshihiko Kadoya provided valuable comments at the ADBI-AGRI Workshop on Aging in Kitakyushu. Their help is greatly appreciated.

Asian Development Bank Institute
Kasumigaseki Building, 8th Floor
3-2-5 Kasumigaseki, Chiyoda-ku
Tokyo 100-6008, Japan

Tel: +81-3-3593-5500
Fax: +81-3-3593-5571
URL: www.adbi.org
E-mail: info@adbi.org

© 2017 Asian Development Bank Institute

Abstract

The Center for Medicare and Medicaid Services has introduced several pay-for-performance programs in the last few years to encourage hospitals to improve quality of care and reduce costs. Some state Medicaid programs have also introduced pay-for-performance for nursing homes. Long-term care providers play an important role in hospital pay-for-performance programs because they can affect the readmission rate and also total episode payments. A good pay-for-performance program will focus on improving quality of care that affects health outcomes. In addition, that quality must vary across providers and be measurable. Furthermore, it is important that the measures be reported in a timely way, that both demand and supply respond to the measures, and that the measures be risk adjusted. Empirical data from Medicare beneficiaries in the state of Michigan show that mean episode payments and readmission rates in skilled nursing facilities vary widely and are sensitive to the number of observations. These practical matters create challenges for implementing pay-for-performance in practice. There is an extensive literature review of pay-for-performance in long-term care in the United States and in Asia.

JEL Classification: I11, I13, I18

Contents

1.	INTRODUCTION	1
2.	CONCEPTUAL FRAMEWORK	2
3.	LITERATURE REVIEW	4
4.	EMPIRICAL WORK	7
5.	PAY-FOR-PERFORMANCE IN ASIA.....	12
5.1	Japan.....	12
5.2	The Republic of Korea.....	13
5.3	Taipei,China	13
5.4	Singapore	14
5.5	The People’s Republic of China	14
6.	CHALLENGES	14
7.	CONCLUSION	15
	REFERENCES	16

1. INTRODUCTION

The Medicare insurance program, which covers elderly persons and some non-elderly disabled in the United States, is undergoing a profound change in how it pays for health care. Traditionally, Medicare has paid for most health care based on quantity of services performed. More services—visits, tests, days, drugs, and consults—means more reimbursement. Under this system, the incentives for providers are to provide more services. Quality of care is not rewarded directly. Keeping costs down, on a per-visit or per-episode basis, is not rewarded directly. Without global budgets or caps, and with managed care playing only a modest role in the entire system, the overall health care system rewards greater quantity, not greater quality of care. It is perhaps not surprising that the United States health care system is by far the most expensive in the world and, by many measures, has poor health outcomes relative to other developed nations. Nor is it surprising that recently policy makers have sought ways to change this system to reward quality of care and cost control.

In recent years, hospitals in the United States have seen several new initiatives to reward both high quality and low episode payments. These programs are generally known as pay-for-performance programs, or P4P. The Center for Medicare and Medicaid Services (CMS) now has five main pay-for-performance programs for the Medicare program. The Hospital Readmission Reduction Program (HRRP) aims to reduce the number of readmissions, which are costly and often indicate poor quality of care. The Hospital Value-Based Purchasing (HVBP) program encourages both higher quality of care and lower episode payments, where episodes of care include 30 days post discharge (Norton et al. 2016). The Comprehensive Care for Joint Replacement (CRJ) model provides bundled payments to hospitals, physicians, and post-acute care providers for treatment of hip and knee replacement, one of the most common surgical procedures among elderly patients. The Hospital-Acquired Condition Reduction (HAC) Program penalizes the lowest performing quartile of hospitals 1% of their Medicare revenue. The Bundled Payments for Care Improvement Initiative (BPCI) encourages care coordination by paying a fixed amount for a collection of services to treat an episode of illness.

These programs all share three important features. First, they take a more comprehensive view of treatment, encompassing not only the index hospital admission, but also the period (usually 30 days) post discharge. This creates incentives to improve coordination of care among different providers, traditionally a weakness of the United States health care system. Second, they reward better quality of care. Quality can be measured in different ways, including lower mortality, fewer readmissions, and better patient satisfaction, but the important thing is that quality of care and outcomes now matter. Third, they reward lower episode payments. The financial incentives for lower episode payments are different (financial penalties for readmissions in HRRP, percentage bonus for lower episode payments in HVBP, and capitation in CJR), but all in some way reward lower Medicare payments and penalize higher Medicare payments.

Taken together, these pay-for-performance programs represent a sea change in how Medicare pays for major health care episodes. No longer will providers be able to earn more money through high readmission rates, not suffer financially for enduring high mortality rates, or not be responsible in part for expensive post-acute care treatment. Hospitals must now think carefully about how to manage other providers, especially skilled nursing facilities, rehabilitation centers, and home health agencies, to improve the entire episode of care.

Furthermore, although the amount of money at state is currently modest, CMS plans to ratchet up the incentives and the number of services covered by pay-for-performance. The trend is for more pay-for-performance in the future, not less. It is imperative to understand the economic incentives in these pay-for-performance programs.

While most of the emphasis on pay-for-performance seems to be directed towards hospitals, long-term care providers are also an extremely important part of pay-for-performance programs. Long-term care providers matter both directly and indirectly. Some states have started their own pay-for-performance programs for Medicaid, which is the state and federal insurance program for people who fall below certain income thresholds, for nursing home care. In those states, nursing homes are directly affected by pay-for-performance. In addition, for all of the hospital pay-for-performance programs, nursing homes and home health agencies are indirectly affected because the quality and cost measures are collected over the 30 days post discharge. When hospitals are held accountable for what happens during the post-discharge period, they will find a way to hold those other providers accountable too.

This paper provides an overview of long-term care and pay-for-performance. It starts with a conceptual framework of how pay-for-performance matters for long-term care providers. The conceptual framework addresses both direct and indirect incentives. There is a literature review of how pay-for-performance has worked thus far in long-term care settings. In addition, there is a close look at readmissions in the state of Michigan, where the private insurer Blue Cross Blue Shield of Michigan has started its own pay-for-performance program for all hospitals. This study of Michigan reveals the many issues and challenges in trying to measure quality of care in long-term care facilities, report the information to hospitals, and have hospitals judged in part on whether skilled nursing facilities have high quality and low cost. There is also a review of pay-for-performance programs in Asian countries.

2. CONCEPTUAL FRAMEWORK

This conceptual framework begins with a discussion of how pay-for-performance directly affects nursing homes, when the pay-for-performance program is targeted directly at the nursing homes. First, it is necessary that quality of care matters. While this may seem obvious, or tautological, it is essential that actual quality of care affects outcomes (health and satisfaction and cost) in a meaningful way. One reason why this relationship in nursing homes is not always obvious is that in nursing homes, death is an expected outcome for many residents. Even high-quality nursing homes have high mortality rates. Furthermore, satisfaction of patients who have dementia may not be as responsive to actual quality as patients with no cognitive problems. But setting aside those issues, we take it as given that improved quality of care will also improve patient outcomes, including health, satisfaction, and episode spending.

Second, there needs to be variation in quality of care. That is, some nursing homes need to be better than others, and nursing homes can improve over time if they put in the effort. Overall quality of care in nursing homes is considered poor, but also quite variable (Grabowski and Norton 2012). Again, this may seem obvious, but it is essential for an incentive program for there to be variation and for nursing homes to be able to improve.

Third, regulators need to be able to measure quality of care. Measuring quality of care in nursing homes is hard. As already mentioned, focusing solely on outcomes like mortality and patient satisfaction provides misleading measures of quality of care. Process measures, such as regulatory fines or staffing, are fairly easy to measure but provide indirect measures of quality that may not be highly correlated with actual quality of care. Other measures such as quality of nurses are notoriously difficult to measure. But, if quality of care is to be rewarded, it must be measured.

Fourth, quality of care must not only be measured, but measured in a timely way and reported back to the regulator and the provider quickly. If it takes years and years to collect and report data, then there cannot be timely improvements. Using claims data usually requires a one- to two-year lag. The CMS HVBP program has a two-year lag.

Fifth, there must be response on both the demand and the supply sides to the information provided about quality of care. On the supply side, nursing homes must have incentives that are strong enough to want to improve quality of care. For example, in the CMS HVBP program, hospitals can gain performance points either by doing well on an absolute scale, or by improving, thus giving incentives to all hospitals to either improve or maintain high quality of care (Das et al. 2016). Konetzka and colleagues (forthcoming 2016) found that different weights for nursing home quality measures mattered, with larger weights not surprisingly leading to larger improvements in clinical outcomes. They also found that a simple rule for deficiencies (threshold for few or no deficiencies) was more effective than a complicated one. On the demand side, patients must respond to improved quality of care by voting with their feet and going to high-quality nursing homes. If there is no demand response, then there is no reason to create a program to try to improve quality of care.

Sixth, the measures of quality of care need to be risk adjusted to allow appropriate comparisons conditional on the case mix severity of patients. Without risk adjustment, nursing homes that treat sicker patients may do worse on measured quality of care, not because they are worse quality of care, but because the patient population has higher morbidity and worse expected health outcomes. Without risk adjustment, there will always be incentives for nursing homes to select healthy patients to achieve a good score through patient selection instead of through high performance.

If all those conditions are met, then a pay-for-performance program aimed at improving quality of care in nursing homes could work.

The payment side of the equation has somewhat different issues. Payments are relatively easy to measure because there is a paper trail. Administrative data keeps close track of all payments. The demand side response is not really relevant for Medicare or Medicaid episode payments, because consumers only respond to out-of-pocket payments (although out-of-pocket payments are usually highly correlated with insurer payments). The insurer (Medicaid or Medicare) wants lower total payments. Furthermore, payments tend to be the same for the same service, eliminating basic variation in per unit prices. However, to the extent that there is price variation (e.g., due to cost of living adjustments, teaching adjustments, inflation across years) those should be eliminated by price standardization, which assigns the same average price to the same service. The real issue with trying to reduce payments is how to do it in a way that does not compromise quality of care. If reducing payments means discharging too early from a nursing home, then that could lead to higher morbidity and mortality.

The economic issues and incentives for pay-for-performance programs aimed at hospitals are different. In these programs, the direct incentives are for hospitals, and nursing homes are only affected indirectly. However, the same list of six issues from above still applies. Assume that quality of care matters, that there is variation, that quality can be measured in a timely way, that there is response on both the supply and demand sides of the market, and that measures are risk adjusted. For long-term care providers, the issues are mostly about post-acute care. Hospitals now care about quality and episode payments, so they want to steer patients to the kind of treatment that will minimize the chance of readmissions and episode payments (at least up to 30 days).

Hospitals can achieve some of this during the index hospitalization. The better the initial surgery and the better the information provided to the patient, the better the post-acute care experience.

However, there are two ways that hospitals can influence what happens during the post-acute care period. First, they can steer patients to different post-acute care providers. Hospital discharge planners have enormous influence over what happens to patients upon discharge. They can nudge patients towards different types of care (e.g., home health vs. skilled nursing facility) or to different specific providers within a type of care. Second, they can try to change the way that post-acute care providers provide treatment and care. This could be through information or threats of withholding patient referrals. The goal would be to change the practice of the provider if a patient is sent there.

These two ways of influencing post-acute care—changing the probability of the provider and changing the type of care conditional on the provider—could be done informally or done formally through contracts. At the extreme, hospitals can buy or build a nursing home, creating a hospital-based nursing home. Or they can contract with existing non-hospital-based nursing homes to try to ensure quality of care.

Having laid out the main issues facing pay-for-performance programs for long-term care providers, we next turn to a literature review of the experience of pay-for-performance. In particular, the next section addresses what pay-for-performance programs look like in terms of measures of quality, incentives, and reporting. The section also summarizes what has happened to nursing homes in practice.

3. LITERATURE REVIEW

The United States is not only a pioneer in developing pay-for-performance programs for inpatient care, but it is also one of the only countries that has created pay-for-performance programs specifically for long-term care providers (Briesacher et al. 2009). As of 2010, 14 (out of 50) state Medicaid programs had enacted pay-for-performance programs or were planning them. Medicaid provides means-tested health insurance, with combined state and federal funding (Werner et al. 2010). Vermont was the first state to have a pay-for-performance program, in 2000. The nine states with existing pay-for-performance programs have about 20% of all nursing homes. The remaining five states that were planning to implement as of 2010 have about 15% of all nursing homes. Therefore, a significant fraction of nursing home residents who are covered by Medicaid are in nursing homes subject to pay-for-performance. Medicaid pays for roughly half of all nursing home days, but its reimbursement is lower than private payers by roughly 10% to 30% (Norton 2000).

The details of the pay-for-performance programs vary widely, yet all the programs share certain core features (Werner et al. 2010). Each program employs a variety of measures to try to capture quality of care. Each state uses at least three, most at least four. Nearly all use staffing, regulatory deficiencies, and patient satisfaction. Staffing and regulatory deficiencies are easy to measure as they are collected for other purposes. Other measures used include clinical process measures, occupancy, efficiency, and Medicaid use.

Most states assign points to nursing homes for achieving a certain threshold or if it is among the highest ranked (Werner et al. 2010). Points are only awarded for top performers, not for low performers who nonetheless improve (as is often done in hospital pay-for-performance programs). One problem of this system is that it gives little incentive for a low performing nursing home to improve at the margin (Norton et al., 2016). The amount of money spent on the financial bonuses is modest, ranging from 0.1% to 1.8% (Werner et al., 2010). An important policy question is whether the incentives have spurred improvements in quality of care or lowered overall payments.

In a large-scale study of the effects of pay-for-performance for nursing homes by eight state Medicaid agencies during the decade of the 2000s, Werner and colleagues (2013) found only a few measures improved. They found that three clinical quality measures improved. They showed that the use of restraints, development of pressure sores, and number patients in moderate to severe pain all declined. However, neither of the two structural measures—nurse staffing and number of deficiencies—that were directly linked to points improved. Other health measures either did not improve or got worse. In sum, the results were decidedly mixed and somewhat disappointing. Werner and colleagues (2013) speculate that the incentives may have been too small, the results of measures not timely enough, or that it could be necessary to provide financial incentives directly to staff instead of to the organization.

Evaluation of Medicaid pay-for-performance programs by Werner and colleagues (forthcoming 2016) has revealed another interesting design issue. One challenge with creating incentives is to make positive incentives for all nursing homes, including those that are starting at a high level of quality of care and those that are starting at a low level. Take a threshold system, where nursing homes that achieve a certain threshold get a bonus. If the bonus is hard to attain, so that only a small fraction of nursing homes can realistically surpass the threshold, then nursing homes that start at a low level of quality may feel that they have no chance and will not try. In economic terms, the return on investment is too great to bother investing. On the other hand, if the threshold is modest, so that some nursing homes easily surpass it, then those at the top have no incentive to improve. Werner and colleagues (forthcoming 2016) found exactly this last problem when comparing results across six states. Nursing homes well above the threshold got worse, although those well below the threshold seemed to have enough incentive to try to improve.

There is only one prior peer-reviewed study that has rigorously evaluated a pay-for-performance program in nursing homes, using treatment and control nursing homes (Norton 1992). The nursing homes were given monetary incentives to improve the health of nursing home residents and lower Medicaid expenditures. Results showed that the incentives improved the quality of care and reduced Medicaid expenditures on nursing homes. Furthermore, the study found that nursing homes admitted more people with severe disabilities and that their average length of their stay was shortened. This sort of pay-for-performance program is consistent with the goal of transferring more people out of hospitals and into nursing homes, which serve as a lower-cost substitute.

While most nursing home pay-for-performance programs are run by state Medicaid programs, a recent study by Grabowski and colleagues (forthcoming 2016) evaluated the Medicare Value-Based Purchasing Program demonstration in three states (Arizona, New York, and Wisconsin). The purpose of the program was to provide incentives to improve quality of care and lower Medicare spending. The study found no change in measured quality of care. There were declines in Medicare spending in the first year for treatment nursing homes in Arizona, and for the first two years in Wisconsin, but no change in New York or in the long run in any state. The authors also interviewed administrators and concluded that few nursing homes made any appreciable effort to change, so finding so significant results should not be a surprise. The program as designed was not effective.

As hospitals are increasingly held accountable for patients' post-discharge outcomes under pay-for-performance programs, hospitals may want to direct patients to certain skilled nursing facilities (SNFs) at discharge to better manage these outcomes. This raises the question of whether patients discharged to hospital-based SNFs have better outcomes. Rahman and colleagues (2016) found that hospital-based skilled nursing facility patients spent about 5 more days in the community and 6 fewer days in the SNF in the 180 days following their original hospital discharge. They found no significant effect on mortality or hospital readmission, and yet Medicare spent almost \$2,900 less on a hospital-based SNF patient in the 30 days following their original hospital discharge. The analysis controlled for selection into hospital-based skilled nursing facilities.

The research question of whether hospital-based SNFs produce better outcomes with lower Medicare payments is important given recent pay-for-performance policies that hold hospitals accountable for post-hospital discharge outcomes and spending. The results of Rahman and colleagues (2016) provide some support for vertical integration of hospitals and SNFs. Hospitals can consider several ways to partner with SNFs, including ownership or other contractual arrangements with freestanding SNFs. Hospital networks are currently developing strategies to develop networks of SNF partners (Maly et al. 2012; Lage et al. 2015).

CMS is committed to tying Medicare payments to the value of care delivered. One of CMS's major initiatives is the Hospital Value-based Purchasing Program (HVBP), which pays financial bonuses to hospitals based on their quality of care and the episode-based payments of care. Norton and colleagues (2016) argue that every Medicare patient affects their hospital's performance on a variety of quality and spending measures. The change in these measures translates directly to changes in program points and eventually to dollars of Medicare reimbursement. Every Medicare patient affects the hospital reimbursement through that patient's marginal future reimbursement. Norton and colleagues (2016) estimate the magnitude of the marginal future reimbursement for individual patients for each quality and efficiency measure. Their method can be generalized to any other pay-for-performance program.

One concern of all pay-for-performance programs is whether the incentives are large enough to cause hospitals to change their behavior, as was discussed in the section on conceptual framework. It is necessary that providers respond to the incentives. Norton and colleagues (2016) found some evidence that hospitals improved their performance over time in the areas where they have the highest marginal incentives to improve care.

The results of Norton and colleagues (2016) align with several of the issues presented in the conceptual framework. There is great variation across providers in the measures, and these risk-adjusted measures are strongly linked to quality of care. The measures are measurable in a fairly timely way. Finally, there is some evidence of supply-side response.

Two of the implications of rewarding quality of care over an episode of care, which requires coordination across multiple providers, are that hospitals that have already integrated with post-acute care providers will perform better, and that some hospitals will have strong incentives to integrate more with post-acute care providers (Norton et al. 2016). For example, hospitals could strengthen integration with SNFs through mergers, acquisitions, or informal contracts. The HVBP program is especially important for the considering coordination with long-term care providers because the HVBP applies to almost all general hospitals.

4. EMPIRICAL WORK

To investigate some of these ideas further, it is instructive to look at Medicare data on nursing home use by elderly persons. SNFs play an important role in the post-acute care of elderly persons because any Medicare beneficiary who has a hospital stay of at least three days is fully covered by Medicare for the first 20 days, and then the next 80 days are partly subsidized. Specifically, this analysis looks at two outcomes: episode payments and readmission rates, each at the level of the skilled nursing facility (SNF). From a hospital's perspective, these are the two most important outcomes that affect pay-for-performance measures that are plausibly affected by SNF quality. There is evidence that observed risk-adjusted readmission rates from a SNF back to a hospital signal differences in quality, not just patient severity (Rahman et al. forthcoming 2016). Hospitals want to send discharged patients to nursing homes that will have short lengths of stay and that will not readmit the patients.

Consider a hospital's discharge planning decision. If the hospital is considering discharging a patient to one of two nursing homes, and if that hospital knows that one nursing home has lower average payments for post-acute care, then they would want their patient to go to that nursing home, *ceteris paribus*. Similarly, if the hospital knew that one nursing home had lower average readmission rates, then again they would want their patient to go to that nursing home. Of course, it would be important for the measures of payments and readmission rates to be price standardized and risk adjusted. A hospital with accurate information about those important outcomes could presumably improve their pay-for-performance measures of episode payments and lower their readmission rates.

Therefore, the analysis will look at those two outcomes averaged at the SNF level. Using these data it will answer four questions. First, how much variation is there in these measures across SNFs? Second, if all SNFs with episode payments, or with readmission rates, above the overall average reduced their numbers to the overall average, how much improvement would there be? Third, how much of the variation can be explained by small sample sizes? Fourth, how correlated are the results from one year to the next?

These questions get at the essential theoretical issues discussed in the conceptual framework. It is necessary that there is variation across SNFs in measures, and that the variation be largely systematic instead of random.

The Medicare data are for elderly persons the state of Michigan, which is located among the Great Lakes in the upper Midwestern region of the United States. Michigan is the 10th largest state by population and the 11th largest by land area (see Figure 1). I have access to Medicare claims data for Michigan Medicare beneficiaries, from July 2010 through June 2015, for a total of 5 years.

Figure 1: Map of the State of Michigan (in red)



The data set consists of all Medicare SNF episodes following a hospital stay for one of seven major conditions: joint replacement, coronary artery bypass graft (CABG), acute myocardial infarction (AMI), pneumonia, congestive heart failure, colectomy, and spine surgery. These conditions were chosen because they are common, expensive, include both surgical and medical conditions, and are used in the Michigan Value Collaborative as the main conditions used in the Blue Cross Blue Shield of Michigan pay-for-performance program. In short, these are conditions that matter from both a clinical perspective and from an economic perspective.

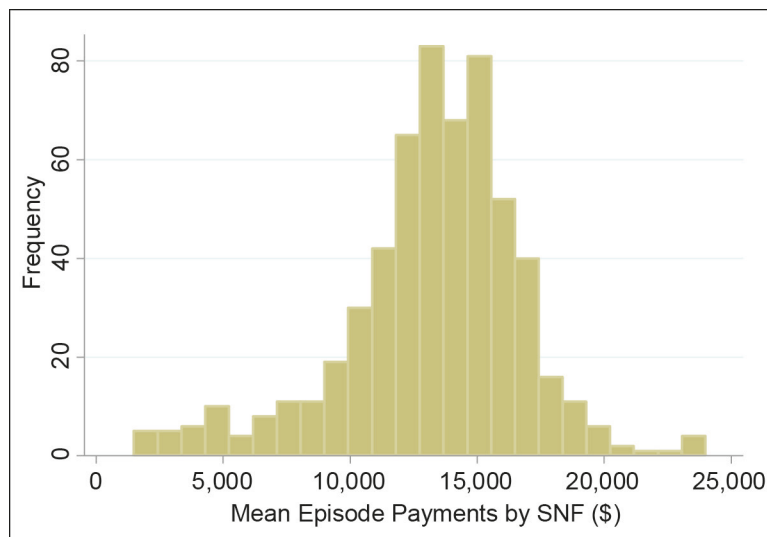
The data are aggregated up to the SNF level (or in some cases the SNF-year level). Due to the potential problem of small sample sizes, I eliminated SNFs that had fewer than 10 admissions during this time period. Because the data are originally collected at the episode level, it is possible for some individuals to appear in the data more than once. The original sample includes 194,213 SNF episodes; after excluding observations from small SNFs (many of which are not located in Michigan); the final sample has 190,174 episodes in 581 unique SNFs.

The results start with average episode SNF payments. Because the payments are price standardized, the amount of payment is basically a linear function of the number of days in the SNF. Price standardization uses the same price per day for the same service. There are no differences in payments due to cost of living adjustments, rural adjustments, teaching adjustments, or any other adjustments. However, the results are not risk adjusted. Risk adjustment would attempt to adjust for case mix, essentially rewarding SNFs that admit sicker patients by adjusting their episode payments to

reflect greater expected payments. While it is possible that the results would differ with case mix adjustment, in my experience case mix adjusting nursing home use is difficult, with more severely ill patients both staying longer (greater acuity) and staying shorter (due to death and readmission). Even without risk adjustments, the main points of the analysis remain.

The variation in mean episode payments for SNFs is wide. The interquartile range is about \$3,725 (from \$11,693 to \$15,418) (see Figure 2). The mean Medicare payments are about \$13,233, for an average of 28 days. If the SNF-level mean accurately reflects differences in SNFs in how they treat patients—and are not due to case mix differences or random variation (issues that we will explore further)—then this observed variation would be extremely important for hospitals that are trying to reduce post-acute care episode payments on SNFs. As an example, suppose that hospitals refused to send patients to SNFs with above-the-mean payments, and instead sent them to SNFs that are exactly at the mean. This is equivalent to truncating the entire distribution in Figure 2 at the mean. In that case the new mean would be only \$11,934, for a reduction of about \$1,300 or about 10%. That is, reducing the highest spending down to the mean would reduce overall spending on SNFs by about 10%.

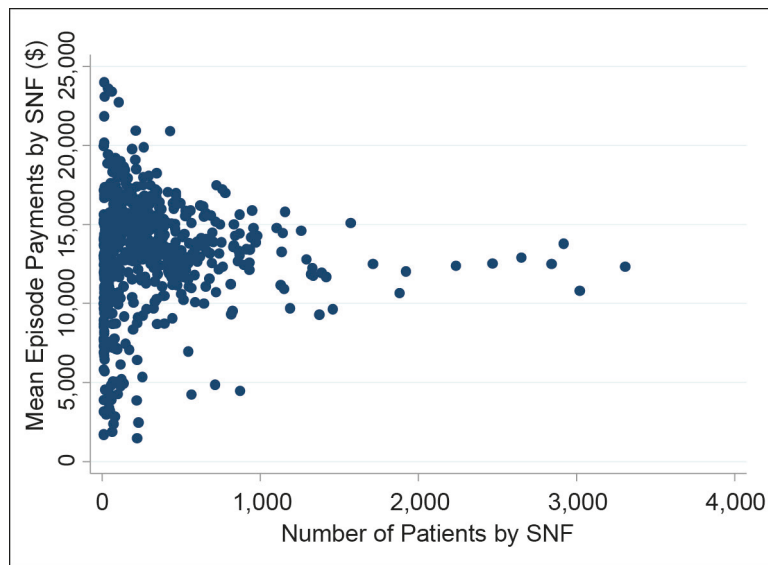
Figure 2: Histogram of the Mean Episode Payments, by SNF



SNF = skilled nursing facilities.

Could some of that variation be due to small sample size? One way to explore this is to graph average SNF payments by the number of patients. We know that small samples have higher variance than large samples, but it is not obvious how large a sample one needs before the SNF-level means stabilize. The answer is that one needs several hundred observations, as shown in Figure 3. SNFs with fewer than 200 observations have wide variation, too wide to have any confidence that the estimates represent only SNF-specific quality of care. SNFs with more than 1,000 observations have small variation, but relatively few SNFs have that many. Further compounding the problem is that many pay-for-performance measures are measured on annual data, whereas the measures presented in this analysis are from a full five years. Data based on a single year would be far more variable.

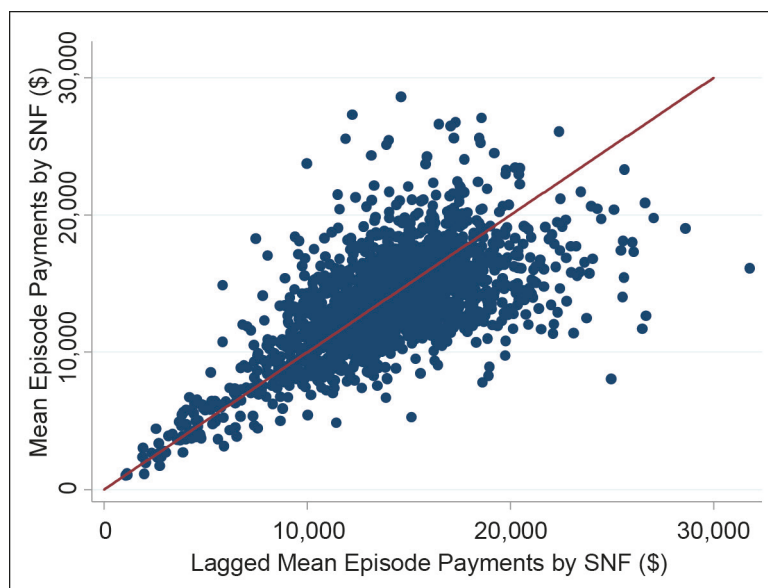
Figure 3: Mean Episode Payments by SNF Plotted against Number of Patients



SNF = skilled nursing facilities.

Another simple way to investigate the stability of the SNF-level estimates is to plot one year’s estimate against the prior year’s estimate. If the estimates are a perfect signal of immutable SNF quality, then the estimates will align along the 45 degree line. If the estimates are pure random noise, with no SNF-specific information, then the estimates will form a round ball of points. The results show strong correlation (about 0.648) and with wider variation for higher values (see Figure 4). That is, the signal is fairly strong, but there is still much year-to-year variation in the measures. When limited to just hospitals with a high number of observations, the plotted points are much more tightly grouped around the 45 degree line (results not shown).

Figure 4: Mean Episode Payments by SNF Plotted against Lagged Values

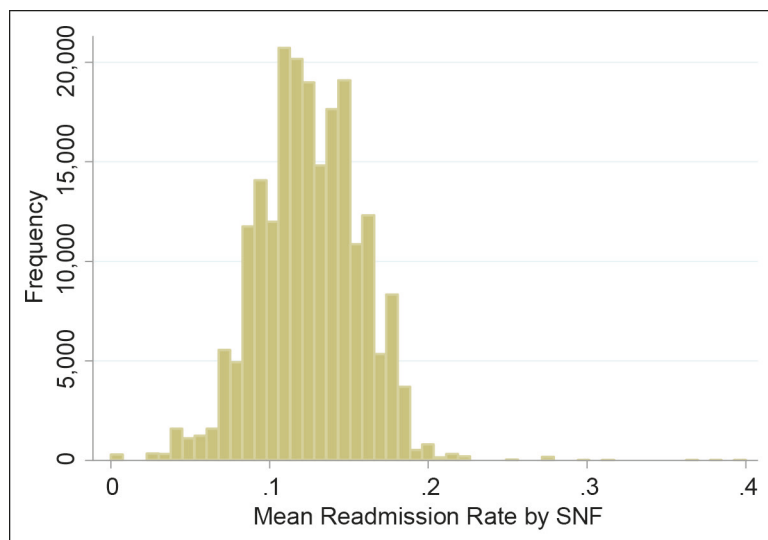


SNF = skilled nursing facilities.

The story is similar for the SNF-level readmission rates. Overall 30-day readmission rates are about 14.6%, meaning that about one in seven Medicare patients who go to a SNF for post-acute care following a hospitalization for one of the seven conditions listed above will be readmitted to the hospital within 30 days. To be clear, these are readmissions that occur after going to the SNF.

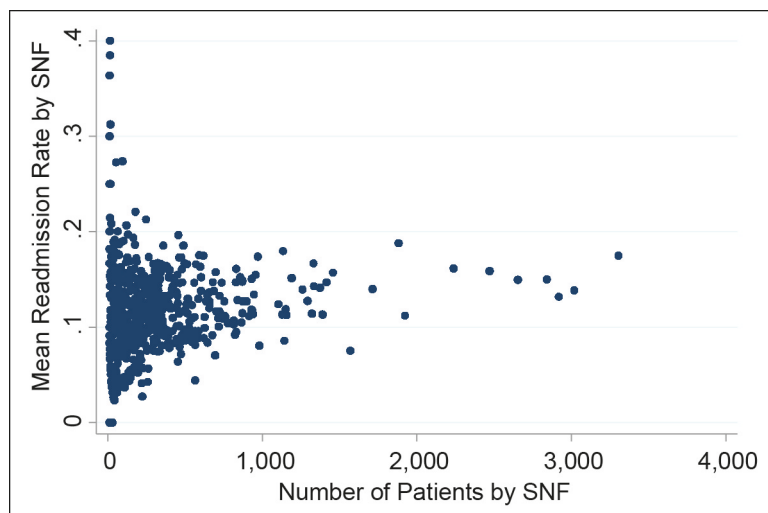
Again, the variation in SNF-level readmission rates is wide, with most of the variation in the range of 10% to 20% (see Figure 5). This substantial variation is based on five years of aggregated data, and could be due to both differences in SNF quality of care in avoiding readmissions, as well as unmeasured case mix and random noise. If all patients who went to a SNF with above-average readmission rates instead went to SNFs with exactly average rates, then the mean readmission rates would fall from 14.6% to 13.2%, for a decline of 1.4%, or again about a 10% reduction.

Figure 5: Histogram of Mean Readmission Rates, by SNF



SNF = skilled nursing facilities.

Figure 6: Mean Readmission Rates by SNF Plotted against Number of Patients



SNF = skilled nursing facilities.

The variance of the estimated SNF-level readmission rates is inversely related to the number of patients, as shown in Figure 6. Again, the variance is extremely high in the many SNFs with fewer than 200 observations.

The results highlight some of the concerns and problems with hospitals trying to use SNF-level information to make decisions on post-acute care so as to improve their own pay-for-performance measures. Data on quality of care is more reliable when based on several hundred observations instead of only a few dozen. However, these measures do have variation across SNFs and the variation is fairly highly correlated from year to year.

5. PAY-FOR-PERFORMANCE IN ASIA

Pay-for-performance programs are still relatively new in Asia, if they are to be found at all. This is not surprising because the political and economic forces for pay-for-performance programs to exist are much stronger in developed countries. The goal of reducing health care spending is more compelling when spending per capita is high. Because health care is a normal good, spending on health care as a percentage of GNP tends to rise with national wealth. In addition, an effective pay-for-performance program is hard to design and monitor. A country must have advanced methods to measure health outcomes in a timely way and must have thought through how to combat selection and problems to avoid providers gaming the system. It is not a surprise that within Asia, the Republic of Korea and Japan have used pay-for-performance programs to a greater extent than other countries.

5.1 Japan

In 2008, Japan began a nationwide pay-for-performance program to improve the health outcomes of stroke patients (Milstein and Schreyoegg, in press). Hospitals are rewarded on the basis of improved clinical process measures and on patient health outcomes. Specifically, hospitals were expected to discharge at least 60% of participating stroke survivors to the community, to hospitalize at least 15% of severe stroke patients, and have at least 30% of patients show improvement in activities of daily living or functional recovery by the time of discharge (Jeong et al. 2010). When compared to other pay-for-performance programs, the Japanese stroke program has relatively small financial incentive, with potential bonuses less than 1% and no penalties.

One study showed positive effects of Japan's pay-for-performance program on clinical process measures, but no effect on health outcomes (Inoue et al. 2011). That study also found evidence that providers selected into the program patients who are good risks for improved health outcomes.

Although Japan does now have long-term care insurance (Hanaoka and Norton 2008), and most stroke victims are elderly, the Japanese stroke pay-for-performance program is not specifically targeted at long-term care providers. However, Japan introduced two pay-for-performance programs for public long-term care services in April 2006. One program pays long-term care facilities an increased rate if the fraction of patients who have stroke rehabilitation exceeds a certain threshold. The other program rewards high-performing long-term care institutions that provide prevention programs through day care services. Another program was implemented just in Shiga prefecture in 2012, where providers received a bonus for better outcomes. However, Iizuka and colleagues (2016) found no overall effect on health outcomes and expenditures, but only small distributional changes towards providers with larger expected incentives.

5.2 The Republic of Korea

The Republic of Korea began a limited pay-for-performance program in 2007 (Lee et al. 2012). The Ministry of Health and the Health Insurance Review & Assessment Service (HIRA) focused on two specific conditions: acute myocardial infarctions (AMI) and caesarian sections (C-sections). Those conditions were chosen because the Republic of Korea ranked relatively poorly for those two conditions among OECD countries (Milstein and Schreyoegg, in press). The goal was to improve health outcomes and to reduce the variation across tertiary teaching hospitals. The amount of revenue at stake was up to 2%, with up to 1% bonus for high-performing hospitals and up to 1% penalty for poor-performing hospitals (Lee et al. 2012; Milstein and Schreyoegg, in press). Only tertiary teaching hospitals were involved.

Based on the initial success with improving care for AMI and reducing the rate of C-sections in teaching hospitals, the Republic of Korea decided to expand the pay-for-performance programs to include general hospitals (Lee et al. 2012). The next conditions added to the program were stroke and prophylactic use of antibiotics, but eventually the intent is to include many more conditions. Over time, there are plans to expand to more conditions and to more providers. The number of health care providers included will also expand beyond hospitals.

5.3 Taipei, China

Taipei, China has several pay-for-performance programs that are targeted at specific diseases. Starting in 2001, the Bureau of National Health Insurance implemented pay-for-performance programs for diabetes mellitus, tuberculosis, breast cancer, cervical cancer, and asthma (Lee et al. 2010). Later these programs were expanded to other diseases, including hypertension and hepatitis. The patients with these diseases are mostly treated as outpatients. The providers are paid a bonus if there is significant improvement in the patient's health outcome. The bonus payments are paid to the medical institutions, not to the medical professionals directly.

The Taipei, China experience, however, exposes one major challenge in designing appropriate incentives in pay-for-performance programs. Patients are recruited into the pay-for-performance program at the request of the medical professional. Therefore, physicians can select those patients who are most likely to earn them a bonus, and exclude those patients who will not. Chang and colleagues (2012) studied the first 5 years of the pay-for-performance program (2001–2005) in Taipei, China for patients with diabetes. Essentially, all patients who were enrolled in the pay-for-performance program adhered to all process measures. This result is not that impressive, however, considering that less than half of all diabetic patients were actually enrolled in the program. Favorable selection means that the incentives to improve care were only applied to a minority of diabetes patients.

The Taipei, China tuberculosis pay-for-performance program went national in 2004, although hospitals could choose whether to participate or not. Li and colleagues (2010) used that source of variation in participation to compare health outcomes of patients with tuberculosis, even though within participating hospitals only some patients were selected (non-randomly) for participation. They found significant improvement in the cure rate and the average rate of treatment for patients in participating hospitals.

Currently the National Health Insurance program in Taipei, China does not cover nursing home care. Therefore, any of the pay-for-performance programs would not affect long-term care directly.

5.4 Singapore

The Singapore Ministry of Health is planning to try pay-for-performance for a few select procedures, initially as pilot projects. For example, starting in November, 2016, they will run a bundled payment program for hip fractures as a pilot project. Therefore, it is too early to have any results from this pilot project.

5.5 The People's Republic of China

Providers in the People's Republic of China (PRC) are paid almost exclusively on a fee-for-service basis. This has led to predictable concerns about providers recommending unnecessary diagnostic tests and prescribing unnecessary prescription drugs to maximize revenue (Sun et al. 2016; Yip et al. 2014). Prescription drugs account for a large fraction of revenues for hospitals and providers. There is belief that fee-for-service reimbursement has caused providers to over-prescribe. Recently, the PRC government has encouraged local governments to pilot test alternative payment methods, including pay-for-performance (Yip et al. 2014).

There are two recent studies of pilot programs in rural PRC. While these studies are fairly rigorous, both point out concerns about selection and political interference that may still contaminate results. Sun and colleagues (2016) found that a combination of a global budget and pay-for-performance could reduce what had been extremely high prescription drug rates under fee-for-service. These results were for Shandong Province, a relatively wealthy part of eastern PRC in 2011–2012. Yip and colleagues (2014) studied a different pilot program in rural Ningxia Province. That pilot program also included global budgets and pay-for-performance on a wide variety of measures, including prescribing, spending, visits, and patient satisfaction. While the authors found a significant reduction in antibiotic prescriptions, there was only a small reduction in total spending and no change in other measures.

Like most Asian countries, there is a long-standing tradition in the PRC of adult children caring for their aging parents. Long-term care is relatively new and still uncommon in the PRC. There is no long-term care insurance program.

6. CHALLENGES

Having presented a conceptual framework for thinking about pay-for-performance programs, reviewed the literature in the United States and in Asia, and analyzed Medicare data on readmission rates and spending at the SNF level, it is clear that there are many challenges to creating an effective pay-for-performance program. There are many examples of pay-for-performance programs that have little or no demonstrated effect on quality or on health outcomes. There appear to be four main challenges.

First, it is important to choose measures of quality of care that affect health outcomes and to measure them well. Although quality of individual nurses certainly matters, measuring individual quality is nearly impossible. Instead, pay-for-performance programs have used other measures such as staffing, which is easy to measure and has been shown to be related to health outcomes, or structural deficiencies, which are also easy to measure but less correlated directly with health outcomes. Using actual rates of health outcomes, such as readmission rates to hospital or decubitus ulcer rates, raise concern about favorable selection of patients and small number variation.

Second, it is important to risk adjust the measures so that providers that care for sicker patients are not unduly penalized. Sicker patients generally have worse health outcomes. Using structural or process measures avoids some of this problem.

Third, it is important to report the results in a timely way to policy makers, providers, and consumers. If it takes many years for the results to be reported back to the providers, then it is impossible to have timely continuous quality improvement. All stakeholders need transparent access to the data to be able to understand it and act upon it. As discussed in the conceptual framework, there needs to be a supply and demand response to better quality of care, so it makes sense that CMS reports most of the measures publicly on the Hospital Compare and Nursing Home Compare websites.

Finally, it is important to get the incentives right. Too small and the providers will not respond because the return on investment will not be worth the effort. Too large and there is concern that providers will go to excessive lengths to game the system and to cherry-pick patients through favorable selection. Most pay-for-performance programs start with small financial incentives, perhaps as a conservative approach while learning how best to measure and report results, but overall it seems as though incentives are too small.

7. CONCLUSION

Pay-for-performance will continue to grow in importance as a way for insurers to provide incentives for providers to improve quality of care and to lower episode payments. Over the first few years of the pay-for-performance programs, CMS has gradually increased the percentage of Medicare payments to hospitals that are at stake. CMS has stated that their goal is to tie 85% of fee-for-service Medicare payments to either quality or value by 2016 (Burwell 2015). Many states now have pay-for-performance for nursing homes under their Medicaid programs. Blue Cross Blue Shield of Michigan has a required pay-for-performance program for all Michigan hospitals.

For long-term care providers in the United States, pay-for-performance has become important both directly and indirectly. In addition to the direct incentive effects by Medicaid in some states, Medicare's pay-for-performance programs for hospitals affect long-term care providers indirectly. For example, SNFs provide important post-acute care for many Medicare patients. Many outcomes during the first 30 days post discharge are subject to measure and performance bonuses or penalties. More readmissions and higher 30-day episode payments can adversely affect hospitals' bottom line.

However, many challenges remain. To be effective, the measures must be related to true quality of care or cost, have variation across providers, be measureable in a timely way, not be too noisy, and be appropriately adjusted for risk. There are problems with all of these requirements, especially for long-term care because of the nature of the quality measures. State Medicaid programs have tried to overcome these by using a mix of certain measures that have higher validity. Hospitals are starting to merge or contract with long-term care providers to have tighter control over the treatment during the post-acute care period.

Asian countries currently rely less on pay-for-performance programs than the United States, but some are moving in that direction. Hopefully, over time we can all learn from the collective experiences and find ways to improve the measures and incentives in pay-for-performance programs in long-term care.

REFERENCES

- Briesacher, B. A., T. S. Field, J. Baril, and J. H. Gurwitz. 2009. Pay-for-performance in Nursing Homes. *Health Care Financing Review* 30(3): 1–13.
- Burwell, S. M. 2015. Setting value-based payment goals—HHS efforts to improve U.S. health care. *New England Journal of Medicine* 372(10): 897–899.
- Chang, R.-E., S.-P. Li, and D. C. Aron. 2012. *Health Affairs* 31(1): 93–102. <https://www.ncbi.nlm.nih.gov/pubmed/22232099> (accessed 22 February 2017).
- Das, A., E. C. Norton, D. C. Miller, A. M. Ryan, J. D. Birkmeyer, and L. M. Chen. 2016. Adding a Spending Metric to Medicare's Value-based Purchasing Rewarded Low-quality Hospitals. *Health Affairs* 35(5): 898–906.
- Grabowski, D. C., and E. C. Norton. 2012. Nursing Home Quality of Care. In *The Elgar Companion to Health Economics, Second Edition*, edited by A.M. Jones. Northampton, MA: Edward Elgar, pp.307–317.
- Grabowski, D. C., D. G. Stephenson, D. J. Caudry, A. J. O'Malley, L. H. Green, J. A. Doherty, and R. G. Frank. 2016 forthcoming. The Impact of Nursing Home Pay-for-Performance on Quality and Medicare Spending: Results from Nursing Home Value-Based Purchasing Demonstration. *Health Services Research*.
- Hanaoka, C., and E. C. Norton. 2008. Informal and Formal care for Elderly Persons: How Adult Children's Characteristics Affect the Use of Formal Care in Japan. *Social Science & Medicine* 67(6): 1002–1008.
- Iizuka, T., H. Noguchi, and S. Sugawara. 2016. Pay-for-Performance for Long-term Care. *University of Tokyo working paper*.
- Inoue, Y., S. Jeong, and K. Konto. 2011. Changes in Hospital Performance after Implementation of Pay-for-performance in Japan. *Clinical Audit* 3: 17–23.
- Jeong, S., K. Katsunori, N. Shiraishi, and I. Yusuke. 2010. An Evaluation of the Quality of Post-stroke Rehabilitation in Japan. *Clinical Audit* 2: 59–66.
- Konetzka, R. T., M. Skira, and R. M. Werner. Forthcoming 2016. Incentive Design and Quality Improvements: Evidence from State Medicaid Nursing Home Pay-for-Performance Programs. *American Journal of Health Economics*.
- Lage, D. E., D. Rusinak, D. Carr, D. C. Grabowski, and D. C. Ackerly. 2015. Creating a Network of High-quality Skilled Nursing Facilities: Preliminary Data on the Post-acute Care Quality Improvement Experiences of an Accountable Care Organization. *Journal of the American Geriatric Society* 63(4): 804–808.
- Lee, J. Y., S.-I. Lee, N.-S. Kim, S.-H. Kim, W.-S. Son, and M.-W. Jo. 2012. Healthcare Organizations' Attitudes toward Pay-for-performance in Korea. *Health Policy* 108: 277–285.
- Lee, T.-T., S.-H. Cheng, C.-C. Chen, and M.-S. Lai. 2010. *American Journal of Managed Care* 16(1): 65–69. <https://www.ncbi.nlm.nih.gov/pubmed/20148607> (accessed 22 February 2017).
- Li, Y.-H., W.-C. Tsai, M. Khan, W.-T. Yang, T.-F. Lee, Y.-C. Wu, and P. T. Kung. 2010. *Health Policy and Planning* 25(4): 334–341. <https://www.ncbi.nlm.nih.gov/pubmed/20207703> (accessed 22 February 2017).

- Maly, M. B., S. Lawrence, M. K. Jordan, W. J. Davies, M. J. Weiss, L. Deitrick, and D. Salas-Lopez. 2012. Prioritizing Partners across the Continuum. *Journal of the American Medical Directors Association* 13(9): 811–816.
- Milstein, R., and J. Schreyoegg. Forthcoming 2016. Pay for Performance in the Inpatient Sector: A Review of 34 P4P Programs in 14 OECD Countries. *Health Policy*.
- Norton, E. C. 1992. Incentive Regulation of Nursing Homes. *Journal of Health Economics* 11(2): 105–128.
- . 2000. Long-term Care. In *Handbook of Health Economics, Volume 1B*, edited by A. J. Culyer and J. P. Newhouse. New York, NY: Elsevier Science, pp.956–994.
- Norton, E. C., J. Kim, A. Das, and L. M. Chen. 2016. Moneyball in Medicare. *NBER Working Paper* No. 22371. Cambridge, MA: NBER.
- Rahman, M., D. C. Grabowski, V. Mor, and E. C. Norton. Forthcoming 2016. Is a Skilled Nursing Facility's Rehospitalization Rate a Valid Quality Measure? *Health Services Research*.
- Rahman, M., E. C. Norton, and D. C. Grabowski. 2016. Do Hospital-owned Skilled Nursing Facilities Provide better Post-acute Care Quality? *Journal of Health Economics* 50: 36–46.
- Sun, X., X. Liu, Q. Sun, W. Yip, A. Wagstaff, and Q. Meng. 2016. The Impact of a Pay-for-performance Scheme on Prescription Quality in Rural China. *Health Economics* 25(6): 706–722.
- Werner, R. M., R. T. Konetzka, and D. Polsky. 2013. The Effect of Pay-for-Performance in Nursing Homes: Evidence from State Medicaid Programs. *Health Services Research* 48(4): 1393-1414.
- . Forthcoming 2016. The Effect of Pay-for-Performance in Nursing Homes: Evidence from State Medicaid Programs. *Health Services Research*.
- Werner, R. M., R. T. Konetzka, and K. Liang. 2010. State Adoption of Nursing Home Pay-for-Performance. *Medical Care Research and Review* 67(3): 364–377.
- Yip, W. et al. 2014. Capitation Combined With Pay-For-Performance Improves Antibiotic Prescribing Practices In Rural China. *Health Affairs* 33(3): 502–510.