RESEARCH

Observational intensity bias associated with illness adjustment: cross sectional analysis of insurance claims

OPEN ACCESS

John E Wennberg *professor emeritus in the evaluative clinical sciences*[1], Douglas O Staiger *professor of economics*[2], Sandra M Sharp *research associate*[1], Daniel J Gottlieb *research associate*[1], Gwyn Bevan *professor of policy analysis, head of the department of management*[3], Klim McPherson *visiting professor of public health epidemiology and emeritus fellow*[4], H Gilbert Welch *professor of medicine, professor of The Dartmouth Institute, and professor of community and family medicine* [1]

[1]The Dartmouth Institute for Health Policy and Clinical Practice, The Audrey and Theodor Geisel School of Medicine at Dartmouth, Hanover, NH 03755, USA; [2]Department of Economics, Dartmouth College, Hanover, NH, USA; [3]London School of Economics and Political Science, London, UK; [4]Nuffield Department of Obstetrics and Gynaecology, New College, Oxford, UK

Abstract

Objective To determine the bias associated with frequency of visits by physicians in adjusting for illness, using diagnoses recorded in administrative databases.

Setting Claims data from the US Medicare program for services provided in 2007 among 306 US hospital referral regions.

Design Cross sectional analysis.

Participants 20% sample of fee for service Medicare beneficiaries residing in the United States in 2007 (n=5 153 877).

Main outcome measures The effect of illness adjustment on regional mortality and spending rates using standard and visit corrected illness methods for adjustment. The standard method adjusts using comorbidity measures based on diagnoses listed in administrative databases; the modified method corrects these measures for the frequency of visits by physicians. Three conventions for measuring comorbidity are used: the Charlson comorbidity index, Iezzoni chronic conditions, and hierarchical condition categories risk scores.

Results The visit corrected Charlson comorbidity index explained more of the variation in age, sex, and race mortality across the 306 hospital referral regions than did the standard index (R^2=0.21 *v* 0.11, P<0.001) and, compared with sex and race adjusted mortality, reduced regional variation, whereas adjustment using the standard Charlson comorbidity index increased it. Although visit corrected and age, sex, and race adjusted mortality rates were similar in hospital referral regions with the highest and lowest fifths of visits, adjustment using the standard index resulted in a rate that was 18% lower in the highest fifth (46.4 *v* 56.3 deaths per 1000, P<0.001). Age, sex, and race adjusted spending as well as visit corrected spending was more than 30% greater in the highest

fifth of visits than in the lowest fifth, but only 12% greater after adjustment using the standard index. Similar results were obtained using the Iezzoni and the hierarchical condition categories conventions for measuring comorbidity.

Conclusion The rates of visits by physicians introduce substantial bias when regional mortality and spending rates are adjusted for illness using comorbidity measures based on the observed number of diagnoses recorded in Medicare's administrative database. Adjusting without correction for regional variation in visit rates tends to make regions with high rates of visits seem to have lower mortality and lower costs, and vice versa. Visit corrected comorbidity measures better explain variation in age, sex, and race mortality than observed measures, and reduce observational intensity bias.

Introduction

Good methods of risk adjustment are essential to make sense of observed regional variations in utilization, expenditure, and mortality rates between healthcare regions. Unwarranted variation in the use of resources and in outcomes can thus be understood more precisely and the information targeted to improve quality and respond to fiscal pressures. Studies over three decades have shown that the key driver of regional variations in per capita spending for hospitals across areas are variations in admission rates (not in costs per case), and these exceed what would be expected given known differences in morbidity.[1-5] These studies raise the question of the extent to which variation in resource use and healthcare outcomes, after the conventional adjustments in the United States for age, sex, and race, are justified because of differences in risk. Although

Correspondence to: john.e.wennberg@Dartmouth.edu

it is well recognized that data on utilization of healthcare cannot be used directly as a sound proxy for risk, as these data also reflect differences in access to supply of healthcare resources,[6-9] methods of risk adjustment have been developed based on diagnoses recorded in medical records and insurance claims. Such methods are routinely used in calculating standardized mortality rates in the public reporting of hospital mortality[10-15] and in observational studies of the relation between healthcare use and outcomes.[16-19] They are also used to adjust for comorbidity in making payments to physicians[20] and to third party payers: insurance companies in the Netherlands,[21] the US Medicare's Advantage program,[22] and the formula used to determine target allocations of resources to clinical commissioning groups in England.[23]

The validity of methods that use data on diagnoses to estimate risk depends on the assumption that recorded diagnoses are independent of supply, and hence closely reflect the true underlying burden of illness; or that adequate controls have been developed to control for the effects of supply. One study[9] emphasized the importance of adequate controls in designing methods of risk rating to counter incentives for "cream skimming." The assumption that diagnostic data do offer objective measures of need is known to be suspect as differences in coding practices have already been shown to introduce an "up-coding" bias in comparing mortality across hospitals.[24] But these studies did not consider the effects of differences in intensity of patient observation. Recent studies that have done so have demonstrated that the intensity of patient observation, measured by the frequency of visits by physicians and the number of laboratory tests and imaging exams is associated with the frequency of diagnosis recorded in Medicare claims data, independent of underlying illness, as measured by mortality adjusted for age, sex, and race. In a natural experiment, Medicare beneficiaries who migrated to regions with a higher intensity of care (measured by spending on end of life care) experienced more physician visits, diagnostic tests, and imaging exams and accrued more comorbid diagnoses as measured by hierarchical condition categories risk score than did those who migrated to regions with a lower intensity of care, even though the mortality rates over a three year follow-up period after migration were similar.[25] A second study showed that as the regional intensity of patient observation increased (measured by the average numbers of physician visits, different physicians seen, imaging exams, and laboratory tests), the mean number of chronic illnesses diagnosed using the Iezzoni convention for measuring comorbidity also increased, even though the age, sex, and race mortality rates among regions were similar.[26]

We examined the magnitude of the bias associated with the intensity of patient observation using three conventions for measuring comorbidity: the Charlson comorbidity index,[27] the Iezzoni chronic condition count,[28] and the hierarchical condition categories risk score.[22] We compared two methods for risk adjustment: the standard method, which adjusts for comorbidity based on recorded diagnoses, and a modified method that reduces the impact of supply by correcting comorbidity measures to remove the component associated with intensity of patient observation (visits by physicians). To make this correction we used the rate of visits by a physician during the last six months of a patient's life, a measure of observational intensity that is unrelated to disease burden as measured by age, sex, and race adjusted mortality. We tested the validity of the two methods by examining how well they explain and reduce regional variation in age, sex, and race adjusted mortality and their effects on illness adjusted mortality and price adjusted spending rates in regions with high and low rates of physician visits. We discuss

the implications of our findings for the ways in which these strategies are used for risk adjustment.

Methods
Conceptual framework

Figure 1⇓ illustrates our conceptual framework and study design. We consider the observed number of different diagnoses to be a product of two factors: the actual burden of disease (the true or intrinsic amount of disease) and the intensity of observation (which conditions the likelihood of having a diagnosis for a given level of disease burden). Conceptually, the intensity of observation represents the combined effect of a spectrum of factors: the resolution of diagnostic exams (including the physical exam, imaging, and laboratory exams), the thresholds used to label exam results as abnormal, the frequency with which exams are done, and the number of observers who have an opportunity to make a diagnosis.

The intensity of observation is related to the frequency of contact with a physician: the number of visits by physicians, specialist referrals, and admissions to hospital. For this analysis we made use of our prior work[29] and focused on the rate of physician visits in the last six months of life as a proxy for intensity of observation. We then used this proxy to produce a visit corrected measure of illness—essentially subtracting out the intensity of observation component from the observed Charlson comorbidity index, the Iezzoni chronic illness count, and the hierarchical condition categories risk score.

To test whether this new method is a better approximation of actual disease burden, we compared observed comorbidity measures with visit corrected comorbidity measures for their ability to explain and reduce regional variation in the least ambiguous measure of disease burden: age, sex, and race adjusted mortality rates among the sampled Medicare population. We then compared the effect of these two methods of risk adjustment on mortality and spending in individual hospital referral regions and in hospital referral regions aggregated into fifths according to the frequency of physician visits.

Data and measures

We obtained from the Centers of Medicare and Medicaid Services the insurance claims for a 20% sample of Medicare beneficiaries. We restricted the analysis to those beneficiaries who were either fully enrolled in part A and part B of Medicare throughout 2007 and were 65-99 years old on 31 December 2007 or fully enrolled beginning 1 January 2007 until their death that year and were 65-99 years old at their time of death. Because the claims data for beneficiaries enrolled in risk contract health maintenance organizations were incomplete, they were excluded. Analysis was also restricted to beneficiaries who were resident in one of the 306 hospital referral regions as defined in the *Dartmouth Atlas of Health Care*.[29] (The hospital referral regions were empirically defined based on patient origin studies to define the geographic region served by tertiary hospitals.) The sampled Medicare population in hospital referral regions ranged from 2549 to 91 020 beneficiaries. The final sample, covering all 306 hospital referral regions, totaled 5 153 877 beneficiaries.

Our proxy measures for the number of different diagnoses were the illness measures used by each convention. The Charlson comorbidity index[27] comprises 19 acute and chronic conditions with assigned risk scores calibrated to predict one year mortality. The Iezzoni chronic condition[28] is a count of up to 12 chronic

conditions selected on ability to predict one year mortality. The hierarchical condition categories risk score[22] is a composite measure based on an individual's age, sex, and Medicaid and disability status. It is based on the coding algorithms used by Medicare to adjust payments for Medicare Advantage plans[30]; it incorporates an array of diagnoses grouped according to the hierarchical condition categories classification system—70 diagnostic groups calibrated to predict spending and also used to adjust mortality rates for illness in public reporting of hospital mortality.

For a beneficiary to be counted as having a condition, the diagnosis had to be coded either on at least one hospital discharge abstract after an inpatient stay or on at least two claims involving physician contact that were at least seven days apart. This was done to reduce the likelihood that a diagnosis was recorded for a condition that was to be "ruled out" by further study.

The proxy for intensity of observation was the frequency of visits by physicians. To ensure that no direct relation could exist between our proxy for intensity of observation and comorbidity score or mortality we measured physician visits in the prior year (2006). We have adapted this measure for the *Dartmouth Atlas of Health Care*[29] in our hospital referral regions, by reducing it to the average of the total number of visits for evaluation and management (both inpatient and outpatient) made by physicians per beneficiary in the last six months of life, a subset of patient records for which the severity of illness is unlikely to vary importantly across regions. (The regional variation in visit rates is not explained by differences in illness related factors such as age, sex, race, poverty, or type of medical condition.)[31]

Visit corrected comorbidity score

We calculated a visit corrected measure of illness using simple linear regression in which the dependent variable was the observed illness measure (as calculated from the claims data) and the independent variable was the measure of frequency of visits by physicians. The dependent variable was an individual person level comorbidity measure and the independent was the physician visit rate at hospital referral region level. For example, for the Charlson comorbidity index, the residual from this regression—the difference between observed index and predicted index based on the frequency of physician visits—is the visit corrected index. In other words, the residual represents the component of illness not explained by the frequency of physician visits—that is, our proxy for the intensity of observation. Comparable regression models were used to generate person level visit corrected indices for both the hierarchical condition categories risk score and the Iezzoni chronic condition count.

Ability to explain and reduce regional variation in age, sex, and race mortality

We then compared observed comorbidity with visit corrected comorbidity for each of the three measurement conventions for their ability to explain and reduce the regional variation in age, sex, and race adjusted mortality for the sampled populations. Regional adjusted mortality rates were calculated using a linear regression model at the patient level (SAS GENMOD procedure) with 20 age, sex, and race indicator variables, along with the 306 hospital referral regions included as classification variables, and no intercept. We used the resulting region level coefficients to construct age, sex, and race adjusted mortality rates. Regional "standard method" illness adjusted rates were calculated by adding the patient level illness measure (for example, Charlson

score) to the age, sex, and race regression model, using the region level coefficients to construct the rates. We calculated the regional "modified method" illness adjusted rates by adding the patient level visit corrected comorbidity score to the age, sex, and race regression model, again using the region level coefficients to construct the rates.

The percentage of the variation in age, sex, and race adjusted mortality explained by observed compared with visit adjusted comorbidity was analyzed at the regional level using the coefficient of determination (R^2). Results are reported for weighted as well as unweighted regressions for hospital referral region sample size. We measured the effect on regional variation among the 306 hospital referral regions by using standard statistics: ratio of highest to lowest hospital referral region; ratio of 75th to 25th centile; and the coefficient of variation (standard deviation divided by the mean, expressed as a percentage).

Effect of illness adjustment on mortality and Medicare spending

Finally, we compared the effect of risk adjustment using observed and visit corrected comorbidity measures on mortality and price adjusted Medicare spending.[32] Adjustment of spending according to age, sex, and race was accomplished using the same methods as for mortality. The resulting hospital referral region variable estimates represent direct adjusted hospital referral region level mortality and spending rates. Similarly, we also computed adjusted rates for hospital referral regions aggregated into fifths based on the rates of visits by physicians. A series of three regression models was run similar to the regional models but incorporated the fifths as classification variable instead of hospital referral regions.

Results

Observed comorbidity, physician visits, and mortality

Among the 306 hospital referral regions, the mean number of physician visits per decedent during the last six months of life varied across the regions from 10 to 59 and was not correlated with age, sex, and race adjusted mortality (R^2=0.000, P=0.88). The rate of visits, however, was strongly correlated with the number of diagnoses observed in the claims data. The correlation with the mean hierarchical condition categories score was R^2=0.53 (P<0.001) and with the mean number of Iezzoni chronic conditions and Charlson comorbidity index was 0.46 (P<0.001) each. This combination of findings suggests that the frequency of physician visits serves as a good proxy for the intensity of observation: associated with the frequency of diagnosis but not confounded by the actual burden of disease.

Ability to explain and reduce variation in age, sex, and race adjusted mortality

The ability of observed and visit corrected comorbidity measures to explain the variation in age, sex, and race adjusted mortality was shown using unweighted regressions (fig 2⇓). For the hierarchical condition categories, the Iezzoni and the Charlson conventions, comorbidity measures based on the diagnoses in the claims data explained 10-12% of the variation in age, sex, and race adjusted mortality among the 306 hospital referral regions; however, when the observed comorbidity measures were corrected for visits they explained 21-24% of the variation, or about twice as much. Regressions weighted for the sample size of the hospital referral regions explained less: 2-5% of the

variation in age, sex, and race adjusted mortality using observed comorbidity and 17-21% for visit corrected comorbidity.

Compared with age, sex, and race adjusted mortality, adjustment using observed comorbidity resulted in an increase in regional variation in mortality rates, whereas adjustments using visit corrected comorbidity resulted in a decrease in variation for each of the three conventions (fig 3⇓).

Effect of adjustment on hospital referral region specific mortality

Table 1⇓ illustrates the effect of risk adjustment using the Charlson comorbidity index on apparent mortality among hospital referral regions aggregated into fifths based on rates of visits by physicians. The rate of visits in the highest fifth was 2.4 times that of the lowest fifth. However, the age, sex, and race adjusted mortality rates in these two fifths were similar (51.0 *v* 50.0 per 1000). Adjustment for illness using the observed Charlson comorbidity index had a strong impact on illness adjusted mortality: it became highest in the lowest fifth for visits and decreased in a stepwise fashion across fifths of increasing visit rates. Compared with adjustment by age, sex, and race alone, mortality increased by 10.3% in the lowest fifth of visits and decreased by 7.1% in the highest fifth. Thus, after adjustment using the observed Charlson comorbidity index, the illness adjusted mortality rate in the highest fifth of visits was 17.6% lower than in the lowest fifth, despite a nearly identical age, sex, and race adjusted mortality rate.

The effect on mortality of adjustment using the visit corrected Charlson comorbidity index was different. Compared with age, sex, and race adjusted mortality rates, the mortality rates among the fifths showed little change; the stepwise inverse association with rate of visits seen with adjustment using the observed index disappeared; and there was little difference in mortality between the highest and lowest fifths.

Table 1 also provides information on the effect of adjustment for illness in six hospital referral regions with large urban populations; three with high and three with low mean rates of visits by physicians. Compared with adjustment for age, sex, and race, adjustment using the observed Charlson comorbidity index decreased mortality by 18.3% in Miami, 3.2% in Los Angeles, and 13.3% in Manhattan; mortality rates in Minneapolis, Seattle, and Salt Lake City increased by 23.5%, 15.1%, and 20.6%, respectively. Adjustment using visit corrected Charlson comorbidity index resulted in less change in each region except for Los Angeles.

Similar findings were seen for the fifths and the selected hospital referral regions, when adjustments were made using Iezzoni chronic conditions and hierarchical condition categories risk scores (table 2⇓) For example, using the standard method, the illness adjusted morality rate in the highest fifth of visits was 19.0% lower than in the lowest fifth using Iezzoni chronic conditions and 21.8% lower using hierarchical condition categories risk scores.

Effect of adjustment on hospital referral region specific price adjusted spending

Age, sex, and race adjusted and price adjusted Medicare spending in our 20% sample varied among regions, from $5323 (£3376; €3936) to $15 706 per beneficiary and, in contrast with age, sex, and race mortality, was highly correlated with supply as measured by physicians visit rates for end of life care (R²=0.46). Table 3⇓ illustrates the effect of adjustment for illness on Medicare spending using observed hierarchical condition categories risk scores (the standard adjustment used by the

Medicare program to pay insurance companies) compared with visit corrected risk scores. In the fifth with the highest rate of visits, age, sex, and race-price adjusted spending was 32.4% greater than in the lowest fifth. Adjusting for observed hierarchical condition categories risk scores increased spending by 12.8% in the lowest fifth of visits and decreased it by 8.3% in the highest fifth of visits; as a result, illness adjusted spending in the highest fifth was only 7.7% greater than in the lowest fifth. Risk adjustment using visit corrected hierarchical condition categories risk scores resulted in a much smaller change.

Table 3 also illustrates the impact of adjustment on the selected hospital referral regions. Based on age, sex, and race-price adjustment alone, the mean Medicare spending per beneficiary in Miami, Los Angeles, and Manhattan was substantially greater than in Minneapolis, Seattle, and Salt Lake City. After adjustment for illness using observed hierarchical condition categories risk scores, spending in Manhattan was less than in Minneapolis, Seattle, and Salt Lake City; and spending in Minneapolis and Salt Lake City was more than in Los Angeles. Adjustment using visit corrected hierarchical condition categories score resulted in important less change in spending in each region.

Similar findings were obtained for fifths and the selected hospital referral regions using the Charlson comorbidity index and Iezzoni chronic conditions (table 4⇓).

Data for each hospital referral region for visit rates and effects of adjustment on age, sex, and race mortality and price adjusted spending can be downloaded from the Dartmouth atlas website (www.dartmouthatlas.org/).

Discussion

Methods of risk adjustment lack face validity if they fail to reduce variation and result in implausible changes in age, sex, and race mortality rates. The standard method of illness adjustment based on observed comorbidity measures obtained from Medicare's administrative databases explained little of the variation in underlying disease burden as measured by age, sex, and race adjusted mortality and increased rather than decreased regional variation. Moreover, it resulted in implausible changes in mortality rates in regions with high and low rates of visits by physicians. Although age, sex, and race adjusted mortality rates were virtually the same in the highest and lowest fifths of visits, the standard method resulted in illness adjusted mortality that was 18% lower in the highest fifth of visits using the Charlson comorbidity index, 19% lower using Iezzoni chronic conditions, and 22% lower using hierarchical condition categories risk scores.

The visit corrected method for illness adjustment proved to be a better predictor of the underlying burden of illness. For each convention for defining comorbidity, the visit corrected method explained much more of the variation in age, sex, and race adjusted mortality; it reduced rather than increased regional variation compared with the standard method of adjustment and did not result in unlikely changes in illness adjusted mortality rates in regions with high and low rates of visits by physicians.

Limitations of the study

Our analysis has several limitations. Firstly, it was restricted to fee for service Medicare and thus captures only part of the experience of the insured population. Since our data did not include Medicare Advantage (Medicare's capitated insurance plan), we were unable to examine the association between financial incentives involving capitation and the number of

diagnoses nor evaluate the impact of Medicare's policy to use hierarchical condition categories risk scores to reduce "cream skimming" by health insurance companies serving patients enrolled in Medicare Advantage. Secondly, our analysis cannot distinguish between the intensity of observation and intentional up-coding. Coding practices could vary among regions and are not addressed by the risk adjustment methods discussed here. However, to the extent that such behavior was correlated with frequency of visits by physicians, it would be controlled for by our adjustment method. Thirdly, although we use physician visits during the last six months of life as a proxy for intensity of patient observation, the results are not sensitive to this measure: Among the 306 hospital referral regions, the rates of visits in 2006 for those alive at the end of the year were highly correlated with visits for those who died and, like the end of life visit rate, were not correlated with age, sex, and race adjusted mortality rates. Using visit rates in 2006 for those alive at the end of 2006 to adjust hierarchical condition categories scores gave results similar to those reported here. Fourthly, the method for correcting for observational intensity using visit rates as a proxy depends on the availability of this information in administrative databases, limiting the applicability of the method elsewhere. Fifthly, our method for counting diagnoses excludes likely rule-out diagnoses and is thus more conservative than the method used by the Centers for Medicare & Medicaid Services in adjusting Medicare Advantage payments, which counts such diagnoses. In a sensitivity analysis, we found that including rule-out diagnoses in calculating risk scores resulted in even greater changes from age, sex, and race adjusted mortality and spending, particularly in regions at the extreme high and low end of the spectrum of variation in rates of physician visits. Finally, this study did not evaluate risk adjustment methods that, in addition to comorbidity measures based on administrative databases, incorporated measures of supply[23] or prior healthcare use.[33] Further research would be needed to evaluate the ability of such methods to reduce the observational intensity bias observed when risk adjustment is based on counts of diagnoses recorded in claims data.

Our study emphasizes the importance of recognizing the differences between data generated in different ways, as Taleb argued.[34] Those that are produced socially are problematic to interpret because they have meaning to those that generate them and tend to have highly skewed distributions. Those that are produced naturally are straightforward to interpret and tend to follow Gaussian distributions. It is thus a serious categorical error to treat the former data like the latter. In healthcare this distinction has also to be made between social data, such as per capita supply of medical expenditure, admission rates, physician visits, and medical diagnoses; and biologic data, such as height and mortality, which tend to follow a Gaussian distribution.

Adjustment methods that depend on the number of different diagnoses recorded in medical records and administrative databases are subject to observational bias rooted in the varying capacity of the local healthcare system: the more physicians involved in care, the more visits, the more diagnostic tests, the more imaging exams, and the more diagnoses and comorbidities uncovered and recorded in the claims data. Adjustment methodologies that assume that a diagnosis made in Miami, Manhattan, or Los Angeles contains the same information as one made in Minneapolis, Seattle, or Salt Lake City are guilty of what one author labeled the constant risk fallacy: the assumption that the underlying relations between case mix variables and outcomes are constant across populations or over time.[35] In their study of bias in the Dr Foster Unit's standardized mortality ratios, Mohammed et al[36] illustrate this phenomenon

by documenting inconsistency in the relation between the Charlson comorbidity index and adjusted mortality rates across four UK hospitals. For the same methodological reasons this work identifies the same problem on an international scale and offers a plausible reason that can be corrected through a less biased case mix adjustment.

In short, the more one looks, the more one finds. This means that using data on diagnoses to adjust for risk produces potential problems in three areas:

Bias in research and evaluation—Observational studies that use case mix adjustments based on claims data are suspect. For example, if, as for Medicare, spending or utilization per capita and rates of diagnosis are highly correlated, then research that seeks to evaluate the relation between use of medical care and mortality while controlling for illness using recorded diagnoses will produce biased estimates.[30-32]

Biased performance measures—Adjusting performance measures using several different diagnoses makes providers who frequently make diagnoses look better than those who manage their patients more conservatively. Thus, as in the Dr Foster Unit example, adjusting quality and efficiency measures for observed number of diagnoses appears to distort rather than improve the reliability of performance measures. The use of these adjustment methods in public reporting of hospital mortality is widespread, including England,[10] the United States,[11 12] Canada,[13] the Netherlands,[14] and Sweden.[15]

Biased payment to third party payers—Adjusting payments to third party payers on the basis of the frequency of diagnoses recorded in administrative databases may result in higher per capita payments in regions with more physicians, hospital beds, and visits per capita, independent of underlying disease burden. This problem has long been recognized in designing methods of risk adjustment to counter the tendency to avoid sicker patients in order to improve apparent outcomes and profitability, the so called "cream skimming" [9] Hence, methods of risk adjustment that use data of diagnoses try to control for the effects of supply by including socioeconomic data, and data on supply and utilization by individuals. The question raised by the evidence reported here is to what extent is a given method susceptible to observational intensity bias?

Conclusions

We have shown that a method of risk adjustment that used data on diagnoses and controlled for the effects of supply, by using data on the frequency of visits by physicians in the year prior to a patient's death, was more efficient than the standard method; but that still accounted for less than 25% of geographic variation in age, sex, and race adjusted mortality among fee for service Medicare beneficiaries. Thus, our study points to the importance of developing risk adjustment methods that better explain variation in age, sex, and race mortality rates and suggests that these will be found by using data that are clearly independent of the effects of supply.

What is already known on this topic

Among US regions, a strong association exists between the mean number of diagnoses in Medicare's database for insurance claims and the intensity of patient observation measured by physician visits per capita

However, rates of visits by physicians are not correlated with age, sex, and race adjusted mortality

Illness adjustment methods that use the number of diagnoses recorded in administrative databases may be biased by the intensity of observation

What this study adds

After risk adjustment using Medicare's usual method, regions in the highest fifth for visits by physicians had a mean mortality rate that was about 20% lower than the regions in the lowest fifth (observational intensity bias)

When the usual Medicare risk adjustment method was corrected to remove the effect of visits, the mortality rates were similar in all fifths

The visit corrected method better explained and reduced variation among regions in burden of illness measured by age, sex, and race mortality and resulted in plausible mortality and spending rates in regions with high and low visit rates

1 Wennberg JE, Barnes B, Zubkoff M. Professional uncertainty and the problem of supplier-induced demand. *Soc Sci Med* 1982;16:811-24.
2 Wennberg JE, McPherson K, Caper P. Will payment based on diagnosis-related groups control hospital costs? *N Engl J Med* 1984;311:295-300.
3 Wennberg JE, Freeman JL, Culp WJ. Are hospital services rationed in New Haven or over-utilized in Boston? *Lancet* 1987;1:1185-8.
4 Welch WP, Miller ME, Welch HG, Fisher ES, Wennberg JE. Geographic variation in expenditures for physicians' services in the United States. *N Engl J Med* 1993;328:621-7.
5 Wennberg, JE. Tracking medicine: a researcher's quest to understand health care. Oxford University Press; 2010:191-3.
6 Department of Health and Social Security. Sharing resources for health in England. Report of the Resource Allocation Working Party (The RAWP Report) [RARP 1] HMSO, 1976. 2013. www.dh.gov.uk/en/Managingyourorganisation/Financeandplanning/Allocations/DH_4108515.
7 Morgan M, Mays N, Holland WW. Can hospital use be a measure of need for health care? *J Epidemiol Community Health* 1987;41:269-74.
8 Bevan G. The search for a proportionate care law by formula funding in the English NHS. *Financial Accountability and Management* 2009;25:391-410.
9 Van de Ven W, Ellis R. Risk Adjustment in competitive health care plans. Handbook of health economics. North-Holland Elsevier, 2000;755-845.
10 Jarman B, Gault S, Alves B, Hider A, Dolan S, Cook A, et al. Explaining differences in English hospital death rates using routinely collected data. *BMJ* 1999;318:1515-20.
11 Move your dot™: measuring, evaluating, and reducing hospital mortality rates (part 1). IHI Innovation Series white paper. Institute for Healthcare Improvement; 2003. 2013. www.IHI.org.
12 Whittington J, Simmonds T, Jacobsen D. Reducing hospital mortality rates (part 2). IHI Innovation Series white paper. Institute for Healthcare Improvement; 2005. 2013. www.IHI.org.
13 Canadian Institute for Health Information. HSMR: a new approach for measuring hospital mortality trends in Canada. CIHI, 2007.
14 Heijink R, Koolman X, Pieter D, van der Veen A, Jarman B, Westert G. Measuring and explaining mortality in Dutch hospitals; the hospital standardized mortality rate between 2003 and 2005. *BMC Health Serv Res* 2008;8:73.
15 Koster M, Jurgensen U, Spetz C, Rutberg H. [Standardized hospital mortality as quality measurement in healthcare centres and hospitals] [Swedish]. *Lakartidningen* 2008:191391-6.
16 Silber JH, Kaestner R, Even-Shoshan O, Wany Y, Bressler LJ. Aggressive treatment style and surgical outcomes. *Health Serv Res* 2010;45(6p2):1872-92.
17 Ong MK, Mangione CM, Romano PS, Zhou Q, Auerbach AD, Chun A, et al. Looking forward, looking back: assessing variations in hospital resource use and outcomes for elderly patients with heart failure. *Circ Cardiovasc Qual Outcomes* 2009;2:548-57.
18 Romley JA, Jena AB, Goldman DP. Hospital spending and inpatient mortality: evidence from California. *Ann Intern Med* 2011;154:160-7.
19 Schmidt M, Jacobsen JB, Lash TL, Bøtker HE, Sørensen HT. 25 year trends in first time hospitalisation for acute myocardial infarction, subsequent short and long term mortality, and the prognostic impact of sex and comorbidity: a Danish nationwide cohort study. *BMJ* 2012;344:17;e356.
20 GAO Report. Medicare physician payment: care coordination programs used in demonstration show promise, but wider use of payment approach may be limited, (GAO 08-65, Feb 2008). 2013. www.gao.gov/new.items/d0865.pdf.
21 Van de Ven WPMM, Schut FT. Universal mandatory health insurance in the Netherlands: a model for the United States? *Health Aff (Millwood)* 2008;27:771-81.
22 Pope GC, Kautter J, Ellis RP, Ash AS, Ayanian JZ, Iezzoni LI, et al. Risk adjustment of Medicare capitation payments using the CMS-Hierarchical Condition Categories model. *Health Care Financ Rev* 2004;25:119-41.
23 Dixon J, Smith P, Gravelle H, Martin S, Bardsley M, Rice N, et al. A person based formula for allocating commissioning funds to general practices in England: development of a statistical model. *BMJ* 2011;343:1085;d6608.
24 Lilford R, Pronovost P. Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *BMJ* 2010;340:955-7;c2016.
25 Song Y, Skinner J, Bynum J, Sutherland J, Wennberg JE, Fisher ES. Regional variations in diagnostic practices. *N Engl J Med* 2010;363:45-53.
26 Welch HG, Sharp SM, Gottlieb DJ, Skinner JS, Wennberg JE. Geographic variation in diagnosis frequency and risk of death among Medicare beneficiaries. *JAMA* 2011;305:1113-8.
27 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373-83.
28 Iezzoni LI, Heeren T, Foley SM, Daley J, Hughes J, Coffman GA. Chronic conditions and risk of in-hospital death. *Health Serv Res* 1994;29:435-60.
29 Wennberg JE, Fisher ES, Goodman DC, Skinner JS. Tracking the care of patients with severe chronic illness: the Dartmouth atlas of health care. The Dartmouth Institute for Health Policy and Clinical Practice, Dartmouth Atlas Project, 2008.
30 Centers for Medicare & Medicaid Services. Risk adjustment. 2013. www.cms.gov/MedicareAdvtgSpecRateStats/06_Risk_adjustment.asp.
31 Wennberg JE. Chapter 9. Tracking medicine: a researcher's quest to understand health care. New York, NY: Oxford University Press; 2010.
32 Gottlieb DJ, Zhou W, Song Y, Andrews KG, Skinner JS, Sutherland JM. Prices don't drive regional Medicare spending variations. *Health Aff (Millwood)* 2010;29:537-43.
33 Baker LC, Johnson SJ, Macaulay D, Birnbaum H. Integrated telehealth and care management program for Medicare beneficiaries with chronic disease linked to savings. *Health Aff (Millwood)* 2011;30:1689-97. doi:10.1377/hlthaff.2011.0216.
34 Taleb N. The black swan: the impact of the highly improbable. Penguin, 2007.
35 Nicholl J. Case-mix adjustment in non-randomised 4 observational evaluations: the constant risk fallacy. *J Epidemiol Community Health* 2007;61:1010-3.
36 Mohammed MA, Deeks JJ, Girling A, Rudge G, Carmalt M, Stevens AJ, et al. Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals. *BMJ* 2009;338:817-20;b780.

Tables

Table 1| Effect on apparent mortality of standard and visit corrected methods of adjustment for illness using Charlson comorbidity index

Variables	Visits/decedent	ASR adjusted mortality: deaths/1000 (95% CI)	Standard Charlson comorbidity index		Visit corrected Charlson comorbidity index	
			Deaths/1000 (95% CI)	% change from ASR adjustment	Deaths/1000 (95% CI)	% change from ASR adjustment
Visit fifth*:						
1st (lowest)	18.0	51.0 (50.6 to 51.4)	56.3 (55.9 to 56.7)	10.3	52.6 (52.2 to 53.0)	3.1
2nd	23.6	54.0 (53.6 to 54.4)	54.9 (54.5 to 55.3)	1.7	53.1 (52.7 to 53.5)	−1.6
3rd	26.8	53.1 (52.7 to 53.6)	52.8 (52.4 to 53.2)	−0.6	52.1 (51.7 to 52.5)	−1.9
4th	31.2	53.1 (52.7 to 53.5)	51.0 (50.6 to 51.4)	−4.1	51.7 (51.3 to 52.2)	−2.6
5th (highest)	43.9	50.0 (49.5 to 50.4)	46.4 (46.0 to 46.8)	−7.1	51.5 (51.1 to 51.9)	3.2
Ratio (high-low)	2.44	0.98	0.82		0.98	
Selected regions:						
Salt Lake City	13.7	49.0 (46.3 to 51.8)	59.1 (56.5 to 61.7)	20.6	53.9 (51.3 to 56.6)	10.0
Seattle	17.9	49.6 (47.5 to 51.7)	57.1 (55.1 to 59.1)	15.1	53.3 (51.3 to 55.3)	7.5
Minneapolis	20.3	43.4 (41.3 to 45.5)	53.6 (51.6 to 55.6)	23.5	50.6 (48.6 to 52.6)	16.7
Manhattan	43.2	46.3 (44.6 to 48.0)	40.2 (38.5 to 41.8)	−13.3	45.1 (43.4 to 46.7)	−2.7
Los Angeles	55.9	47.3 (45.9 to 48.8)	45.8 (44.4 to 47.2)	−3.2	55.0 (53.7 to 56.4)	16.3
Miami	58.3	51.6 (49.3 to 53.9)	42.2 (39.9 to 44.4)	−18.3	52.2 (50.0 to 54.5)	1.2

ASR=age, sex, and race.

*Data are for 306 hospital referral regions aggregated into fifths according to mean number of visits by physicians in last six months of life for deaths in 2006.

Table 2| Effect on apparent mortality of standard and visit corrected methods of adjustment for illness using Iezzoni chronic conditions and hierarchical condition categories (HCC) risk scores

Variables	Standard Iezzoni chronic conditions		Visit corrected Iezzoni chronic conditions		Standard HCC risk scores		Visit corrected HCC risk scores	
	Deaths/1000 (95% CI)	% change from ASR adjustment	Deaths/1000 (95% CI)	% change from ASR adjustment	Deaths/1000 (95% CI)	% change from ASR adjustment	Deaths/1000 (95% CI)	% change from ASR adjustment
Visit fifth*:								
1st (lowest)	56.4 (56.0-56.8)	10.5	52.4 (52.0 to 52.8)	2.8	57.2 (56.8 to 57.6)	12.1	51.8 (51.4 to 52.2)	1.5
2nd	55.0 (54.6 to 55.4)	2.0	53.1 (52.7 to 53.5)	−1.6	55.3 (54.9 to 55.6)	2.4	52.6 (52.2 to 53.0)	−2.5
3rd	53.1 (52.7 to 53.5)	−0.1	52.3 (51.9 to 52.8)	−1.5	53.1 (52.7 to 53.5)	−0.1	52.1 (51.7 to 52.5)	−2.0
4th	51.2 (50.8 to 51.6)	−3.6	52.1 (51.7 to 52.5)	−2.0	51.2 (50.8 to 51.6)	−3.6	52.4 (52.0 to 52.8)	−1.4
5th (highest)	45.7 (45.3 to 46.1)	−8.6	51.2 (50.8 to 51.6)	2.5	44.7 (44.3 to 45.1)	-10.5	52.2 (51.8 to 52.6)	4.5
Ratio (high-low)	0.81		0.98		0.78		1.01	
Selected regions:								
Salt Lake City	57.9 (55.2 to 60.7)	18.1	52.3 (49.7 to 55.0)	6.7	59.3 (56.8 to 61.9)	21.0	51.8 (49.2 to 54.3)	5.6
Seattle	57.8 (55.7 to 59.8)	16.5	53.7 (51.7 to 55.8)	8.3	57.4 (55.5 to 59.4)	15.8	51.9 (49.9 to 53.9)	4.6
Minneapolis	54.7 (52.7 to 56.7)	26.1	51.6 (49.5 to 53.6)	18.9	55.4 (53.5 to 57.4)	27.8	51.1 (49.2 to 53.1)	17.9
Manhattan	37.2 (35.5 to 38.8)	−19.8	42.4 (40.8 to 44.1)	−8.4	35.1 (33.5 to 36.7)	−24.1	42.3 (40.7 to 43.9)	−8.7
Los Angeles	45.4 (44.0 to 46.7)	−4.2	55.3 (53.9 to 56.7)	16.8	37.8 (36.5 to 39.1)	−20.2	51.3 (50.0 to 52.6)	8.4
Miami	38.5 (36.3 to 40.8)	−25.3	49.3 (47.1 to 51.6)	−4.4	32.6 (30.5 to 34.8)	−36.7	47.3 (45.2 to 49.5)	−8.2

*Data are for 306 hospital referral regions aggregated into quintiles according to mean number of physician visits last 6 months of life for deaths in 2006.

Table 3| Effect on apparent price adjusted Medicare spending of standard and visit corrected methods of adjustment for illness using hierarchical condition categories (HCC) risk scores

Variables	Visits/ decedent	Age, sex, and race adjusted spending/person (95% CI)	Illness adjusted			
			Standard HCC risk scores		Visit corrected HCC risk scores	
			Spending/person (95% CI)	% change from ASR adjustment	Spending/person(95% CI)	% change from ASR adjustment
Visit fifth*:						
1st (lowest)	18.0	7228 (7195 to 7262)	8153 (8129 to 8177)	12.8	7342 (7318 to 7366)	1.6
2nd	23.6	8227 (8193 to 8260)	8419 (8395 to 8443)	2.3	8027 (8003 to 8051)	−2.4
3rd	26.8	8498 (8464 to 8531)	8492 (8468 to 8516)	−0.1	8337 (8314 to 8361)	−1.9
4th	31.2	8878 (8844 to 8912)	8595 (8571 to 8619)	−3.2	8767 (8743 to 8791)	−1.3
5th (highest)	43.9	9572 (9539 to 9605)	8782 (8759 to 8805)	−8.3	9910 (9886 to 9933)	3.5
Ratio (high-low)	2.44	1.32	1.08		1.35	
Selected regions:						
Salt Lake City	13.7	7399 (7184 to 7615)	8944 (8791 to 9097)	20.9	7808 (7655 to 7961)	5.5
Seattle	17.9	6677 (6511 to 6844)	7848 (7729 to 7966)	17.5	7022 (6904 to 7141)	5.2
Minneapolis	20.3	6986 (6822 to 7150)	8792 (8675 to 8909)	25.8	8149 (8032 to 8266)	16.6
Manhattan	43.2	9300 (9165 to 9435)	7624 (7528 to 7720)	−18.0	8699 (8603 to 8795)	−6.5
Los Angeles	55.9	9571 (9459 to 9683)	8141 (8061 to 8221)	−14.9	10 169 (10 089 to 10 249)	6.3
Miami	58.3	15 706 (15 522 to 15 891)	12 866 (12 735 to 12 997)	−18.1	15 070 (14 939 to 15 201)	−4.1

*Data are for 306 hospital referral regions aggregated into fifths according to mean number of visits by physicians in last six months of life for deaths in 2006.

Table 4| Effect on apparent price adjusted Medicare spending of standard and visit corrected methods of adjustment for illness using Charlson comorbidity index and Iezzoni chronic conditions

Variables	Standard Charlson comorbidity index		Visit corrected Charlson comorbidity index		Standard Iezzoni chronic conditions		Visit corrected Iezzoni chronic conditions	
	Spending/person (95% CI)	% change from ASR adjustment	Spending/person (95% CI)	% Change from ASR adjustment	Spending/person (95% CI)	% change from ASR adjustment	Spending/person (95% CI)	% change from ASR adjustment
Visit fifth*:								
1st (lowest)	8051 (8024 to 8079)	11.4	7475 (7447 to 7503)	3.4	8226 (8179 to 8255)	13.8	7489 (7460 to 7528)	3.6
2nd	8370 (8342 to 8397)	1.7	8091 (8063 to 8118)	−1.7	8424 (8395 to 8453)	2.4	8067 (8039 to 8096)	−1.9
3rd	8449 (8421 to 8476)	−0.6	8339 (8311 to 8367)	−1.9	8489 (8460 to 8517)	−0.1	8348 (8320 to 8377)	−1.8
4th	8540 (8512 to 8568)	−3.8	8663 (8635 to 8691)	−2.4	8527 (8498 to 8556)	−3.8	8683 (8654 to 8712)	−2.2
5th (highest)	9017 (8990 to 9044)	−5.8	9819 (9792 to 9846)	2.6	8775 (8747 to 8803)	−8.3	9800 (9772 to 9828)	2.4
Ratio (high-low)	1.12		1.31		1.07		1.31	
Selected regions:								
Salt Lake City	8974 (8796 to 9152)	21.3	8166 (7988 to 8344)	10.4	9046 (8860 to 9231)	22.2	8011 (7826 to 8197)	8.3
Seattle	7846 (7709 to 7984)	17.5	7259 (7121 to 7396)	8.7	8194 (8050 to 8337)	22.7	7442 (7299 to 7585)	11.5
Minneapolis	8575 (8440 to 8711)	22.7	8118 (7982 to 8254)	16.2	9091 (8949 to 9232)	30.1	8506 (8364 to 8647)	21.7
Manhattan	8340 (8228 to 8451)	−10.3	9105 (8993 to 9216)	−2.1	7599 (7483 to 7715)	−18.3	8578 (8461 to 8694)	−7.8
Los Angeles	9331 (9238 to 9423)	−2.5	10 774 (10 681 to 10 867)	12.6	9203 (9107 to 9300)	−3.8	11 050 (10 954 to 11 147)	15.5
Miami	14 235 (14 083 to 14 387)	−9.4	15 803 (15 651 to 15 955)	0.6	13 283 (13 124 to 13 441)	−15.4	15 289 (15 131 to 15 448)	−2.7

*Data are for 306 hospital referral regions aggregated into fifths according to mean number of visits by physician in last six months of life for deaths in 2006.
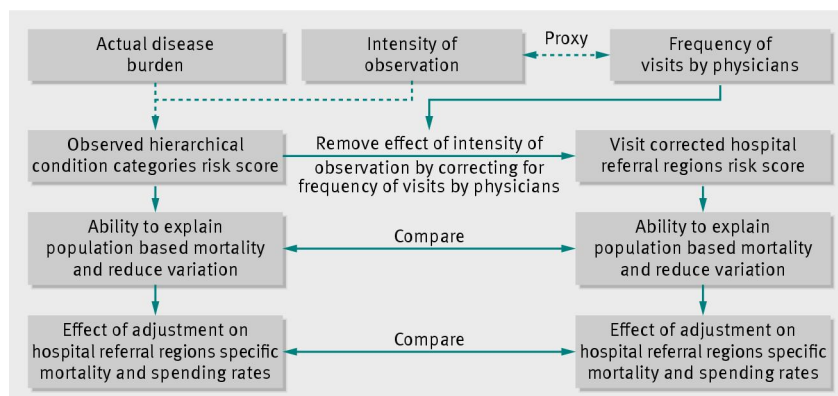
Figures

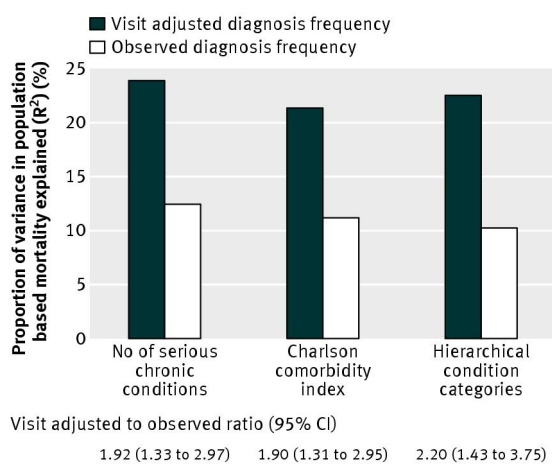**Fig 1** Conceptual framework and study design

Fig 2 Effect of adjusting for visits by physicians on the ability of diagnosis frequency to explain variance in mortality across 306 US hospital referral regions, using three measures for frequency of diagnosis. Regression equation was unweighted for hospital referral region sample size
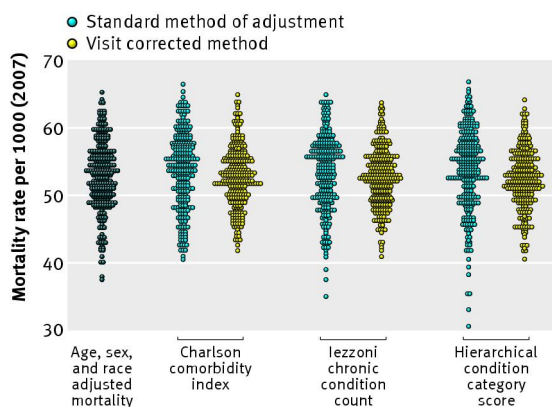

Fig 3 Mortality rates for each US hospital referral region (n=306) according to method of adjustment