Risk Adjusters

To actuaries and other financial professionals, risk adjustment means charging a premium or paying capitation in proportion to the expected costs of an individual or group. Most of the articles herein research exactly that topic. A few have a different view. Concerned with quality or performance indicators, they adjust for the variation in expected outcomes based on the variation in covered groups. We have emphasized the financial articles so only eight are non-financial. An online search for risk adjusters/adjustment does not easily produce more than 60 or so recent academic articles. Eliminating foreign articles greatly reduces this. We chose the best of the rest.

Financing Health Care

The financial articles describe linear regression models in which the dependent variable is cost and the real subject is the input variables. After development using retrospective data, the coefficients can be applied to prospective data to predict costs. A starting point for this literature is Greenwald ("Medicare..."). This article is not research but an overview of and introduction to the subject and to the special issue of *Health Care Financing Review* devoted to risk adjustment. Nine articles are from that single issue. Data is from fee for service programs but is applied to predict managed care costs. This situation will be improved as experience develops under capitation. Older models are based on only demographic variables. The current Medicare model, principal inpatient diagnosis cost group or PIPDCG (described in the articles by Ingber and Pope), includes a variable for inpatient diagnosis. Proposals for further improvement include limitations on activities of daily living (McCall and Temkin-Greener), comorbidities (Beddhu), all encounter diagnosis rather than just inpatient (Ash and Carter), or all of the above (Pope, 1998). Gilmer uses prescription records to predict medical costs. The Cumming article tests seven models, four of which use prescription records. Some more sophisticated models are mentioned such as the "two part" model (Pope, 1998) which adjusts for the typical claims distribution of many zeros and a few very large claims. The regression R^2 statistic, indicating the amount of variance in individual costs explained by the model, is low for all models. It ranges from 1% for the pure demographic model, to higher single digits for PIPDCG, to perhaps 20% with more data. The models tend to distinguish costs for large heterogeneous groups well and for small or special groups poorly.

Financing Health Care: PIPDCG and Specific Groups

Five articles (Kronick, McCall, Riley, Robinson, and Temkin-Greener) are concerned with the impact of PIPDCG on particular population groups. PIPDCG's focus on hospitalization causes it to understate costs for home health care, skilled nursing facilities, and outpatient services. Groups needing these services (frail elderly) receive too little capitation. Medicaid and dually enrolled beneficiaries are known to have higher costs that PIPDCG does not predict well. As a result of these model failures, there is still opportunity for cherry picking enrollment. And there are still perverse incentives, for instance, for hospitalization.

Financing Health Care: HMO Selection
One article (Greenwald, "Favorable Selection…") is unique. The authors attempt to
answer the old question of whether Medicare managed care plans have lower costs due to
selectively enrolling the healthier beneficiaries. They look at the average PIPDCG
coefficients for the capitated and fee for service beneficiaries in each county and
conclude that the capitated group is healthier.

Financing Health Care: Behavioral Health
Ettner tests several risk adjustment models in the behavioral health care arena.
Behavioral healthcare is often a carve-out item in capitation. The lack of a good model
means there is potential for "cherry picking." DeLiberty describes the development and
pre-testing of the Indiana risk-adjusted case rate payment system used by the Department
of Mental Health. It focuses on risk adjustment as a means to correct wrong incentives
and mitigate economic risks.

Financing Health Care: For Physicians
Three articles (Anderson, Goodson, and Iezzoni) are written for physicians. Anderson
describes how a medical practice should respond to its financial risk under capitation. He
describes the impacts of risk adjustment, carve outs, and stop loss reinsurance and gives
some advice on which types of practices (size and degree of specialization) might benefit
from each. Goodson is concerned with physicians' fiduciary duties under the perverse
incentives of managed care. Iezzoni gives an introduction to PIPDCG and asks
physicians to participate in its future development.

Quality and Outcomes Measurement
In the context of quality and performance measurement, risk adjustment is also called
case mix adjustment. However, only articles referring specifically to risk adjustment are
included here. The articles that focused on outcomes research and quality note that these
models also have direct and indirect financial impacts. Certification, contracting,
capitation, bargaining, etc. can depend on performance outcomes. Adjustment for the
populations being served makes measurement and comparison more fair and reasonable.
These models usually use logistic regression to predict some binary variable (death or
survival, complications or not, etc.). Inputs are similar to the cost models but are often
more specialized for a narrower purpose. Instead of the R^2 (percent of variance explained
measure), two articles (Myers and Romano) use c, the area under the ROC curve, to
describe overall model performance. Some compare model results by showing how
different models rank a number of providers. While the financial models were usually
prospective, the quality models are current. Variables are used to adjust the current
outcomes rather than future costs.

Quality and Outcomes Measurement: Behavioral Health Care
A good starting point is Hendryx ("Introduction…"). He introduces risk adjustment
issues in mental health care and substance abuse settings. He also introduces an issue of
the *Journal of Behavioral Health Services and Research* devoted to risk adjustment.
Three other articles are also from that journal. Risk adjustment for mental heath care is
less well developed than for other healthcare. This article gives very general guidelines

for choosing good input variables. Dow is similar but also describes research on risk adjusting behavioral health outcomes in Florida. Hendryx's third article ("Comparing…") calibrates a regression model and determines that survey data is needed to supplement demographic data. Hendryx (1999) found that severe diagnosis, age, substance abuse, functional status, and quality of life were predictive of mental health outcomes. Further, the model differentiated agency performance.

Quality and Outcomes Measurement: Various
Beddhu uses comorbidities (the Charlson Index) to predict outcomes of dialysis patients.

DiGiuseppe predicts cesarean delivery rates. Two models are developed with one based on data from medical records and the other based only on birth certificate information. Although the latter information is less reliable, a model based on certain birth certificate items predicted nearly as well as the more extensive model and could be implemented more easily and cheaply. Glantz develops risk-adjusted cesarean quality performance.

Myers develops risk-adjusted rates for complications of hysterectomies. More common surgical and more predictable medical complications are distinguished. Unusual independent variables (hospital type, insurance type, procedure) were used. The general guidance in articles above would discourage some of these variables.

Romano tries to predict myocardial infarction mortality using ICD-9-CM data at admission and at later development. The latter is more predictive but may be circular. This is and Cumming are the only articles about commercially available systems.

Anderson, G. F. and W. E. Weller (1999). "Methods of Reducing the Financial Risk of Physicians Under Capitation." Archives of Family Medicine **8**(2): 149-155.

Keywords: Risk adjustment, capitation, physicians
Purpose: This article familiarizes physicians with issues they will want to consider when they evaluate capitation options and methods that are available to reduce their financial risk.
Data and Methods: This is an opinion piece. It has some secondary data on per capita costs for illustrative purposes.
Results: Physicians, both primary care and specialty, are accepting risk for patient care by entering fully or partially capitated arrangements. In doing so, they accept risk transfer from managed care plans. Physician capitation creates incentive for physicians to provide cost-effective care, but it also places them at considerable financial risk. The authors analyze three issues: the services that are capitated, the risk sharing, and the size of the practice. Then, three methods for reducing or limiting risk are described: 1.) Reinsurance or stop-loss is described as the best method for preventing catastrophic losses to providers. It is recommended for large physician practices. 2.) Carve outs are specific services that are funded on a fee-for-service basis. It is recommended for physicians providing more limited services. 3.) Risk adjustment is described in greatest detail. The description is positive in that the physician is properly funded and the perverse incentive to avoid less healthy clients is eliminated. It is recommended for those seeing less healthy patients.
Uses: Although intended as advice for physicians, it is instructive to insurers and policy makers regarding the thinking of physicians.
Limitations: Some of this is at an introductory level.

Ash, A. S., R. P. Ellis, et al. (2000). "Using Diagnoses to Describe Populations and Predict Costs." Health Care Financing Review **21**(3): 7-28.

Keywords: Risk adjustment, DCG/HCC, Medicare, PIPDCG
Purpose: The paper describes the logic, structure, regression coefficients, and performance of three Diagnostic Cost Group Hierarchical Condition Category (DCG/HCC) models.
Data: DCG/HCC is developed and validated on three important data bases (privately insured, Medicaid, and Medicare), with more than 1 million people each.
Methods: The fit of the regression models is tested with R^2 values. Ability to predict cost for certain groups is also tested. Actual to expected costs is measured in some cases.
Results: The DCG/HCC payment models use health status to predict the future costs of groups. These models use all histories of diagnoses (ICD-9-CM, condensed) to determine medical conditions. DCG/HCC models are more sophisticated than the new, official Medicare Principal Inpatient Diagnosis standard (PIPDCG) that uses only one admission diagnosis and demographics. Costs are regressed on patient demographic and diagnostic variables. The model attempts to adjust for incentives created by the reimbursement plan. Medicare and Medicaid coefficients are similar and smaller than private insurance coefficients. R^2 for these models ranges from 8% to 23%. The authors note that R^2 for pure

demographic models is as low as 2%. More importantly, the DCG/HCC models do well at predicting costs for high and low cost groups.

Uses: It is likely that the Medicare method of risk adjusting will evolve beyond PIPDCG. This shows one favorable direction.

Limitations: The models developed are predictive rather than concurrent. They are more appropriate for determining payments to providers or managed care organizations than for evaluating performance.

Beddhu, S. M., F. J. M. Bruns, et al. (2000). "A Simple Comorbidity Scale Predicts Clinical Outcomes and Costs in Dialysis Patients." <u>American Journal of Medicine</u> **108** (8): 609-613.

Keywords: Comorbidities, dialysis, outcomes, risk adjustment

Purpose: The Charlson Comorbidity Index, a simple measure of comorbid conditions, is tested to see if it predicts clinical outcomes and costs in dialysis patients.

Data: The researchers tracked outpatients on hemodialysis or peritoneal dialysis from July 1996 to June 1998 at the University of Pittsburgh. Two hundred sixty-eight patients were observed for 293 patient-years.

Methods: Comorbidity scores and outcomes were determined by reviewing the Medical Archival Retrieval System database and outpatient records.

Results: In this university-based dialysis program, 25% of the patients accounted for 50% of the costs and 42% of the deaths. The Comorbidity Index successfully predicted admission rates, hospital stays and costs, and mortality. The Health Care Financing Administration uses age and diabetes in its model but these were found to have little predictive value.

Uses: The authors concluded that the modified Charlson Comorbidity Index is a better risk adjuster for outcomes and costs in dialysis patients. They suggest its use for determining payments for dialysis patients under capitated payment schemes and as a risk adjuster for outcomes research.

Limitations: Data was from a single center. Noncompliance and therapeutic interventions were not recorded.

Carter, G. M., R. M. Bell, et al. (2000). "A Clinically Detailed Risk Information System for Cost." <u>Health Care Financing Review</u> **21**(3): 65-92.

Keywords: CD-RISC, Medicaid, risk adjustment

Purpose: Focusing on patients under age 65, the authors discuss a burden-of-disease risk adjustment system (CD-RISC) that predicts resource use.

Data: They used data from four different payers and standardized service costs. 360,000 people were covered.

Methods: Weighted regression models provided the cost coefficients for demographic and condition codes. Split sample validation was used to assure robustness.

Results: CD-RISC uses 3 severity levels of 173 diagnoses to predict costs. It is intended to minimize perverse incentives. The models are replicable, are at least as accurate as other models in the literature (R^2), explain costs across payers, and reduce rewards for biased selection.

Uses: The authors say a prospective model with adjustment for birth episodes and complications would be an appropriate risk adjuster for Medicaid.

Limitations: A large amount of data is needed to calibrate the model. New payers can be fitted linearly by using their relative costs.

Cumming, R. B., D. Knutson, et al. (2002). "A Comparative Analysis of Claims-Based Methods of Health Risk Assessment for Commercial Populations." Society of Actuaries **www.soa.org**.

Keywords: Managed care, pharmacy, risk adjustment,

Purpose: This large study compares the predictive ability of diagnosis-based risk assessment with pharmacy-based approaches. It is the only article here focused on risk assessment for commercial plans. Employees enrolled in different plans offered through one employer may have different risk profiles. The intent is that this be reflected in the employees' or the employer's contributions. They also advocate a new measure of predictive ability, "predictive ratio", which is the inverse of the common actual-to-expected ratio.

Data: The data is all from private PPO and HMO plans covering 749,145 members during 1998 and 1999.

Methods: Seven risk adjustment models (four of which are pharmacy-based) are compared for prospective and concurrent predictive ability. They are also compared with the default weights and with weights developed by the researchers using linear regression. For the latter the data was divided into calibration and validation sets. Finally, predicted and actual claims were compared.

Results: As the other articles often note, individual prediction is poor but prediction for groups is good. Pharmacy-based models do as well as diagnosis-based models. The default weights performed only slightly worse than the newly developed weighs. The diagnosis-based models performed better than in the 1995 SOA study.

Uses: The results can help an employer or plan choose a risk adjustment methodology and system.

Limitations: The authors note that real applications of pharmacy-based systems should do relatively better than shown in this study since they can be based on more current information.

DeLiberty, R. N., F. L. Newman, et al. (2001). "Risk Adjustment in the Hoosier Assurance Plan: Impact on Providers." Journal of Behavioral Health Services & Research **28**(3): 301-318.

Keywords: Risk adjustment, mental health, provider funding,

Purpose: The authors describe the pre-implementation validation done to assure the applicability of a mental health risk adjustment methodology for the Hoosier Plan.

Data: 80% of 1,300 adult consumers completed interviews as well as 88% of 1,100 children. 268 adults and 221 children who were still in the program completed a 9 day follow-up interview. The full Hoosier Health Plan database of 60,000 was used for some aspects of the testing.

Methods: Four goals of the plan were tested: 1. Illness definitions were sound. 2. A common language would be used. 3. Outcome measures would be standardized. 4. The outcomes would be risk adjusted. Numerous statistical methods were

used.

Results: The method was implemented in Indiana in 2000. Child and adult populations are dealt with in separate models. Assessment instruments and their testing are described at length. Their implementation in the field was also tested. Resulting provider funding was pretested.

Uses: The conclusion is that the incentives are in place to provide appropriate levels of care to the individual. The funding to providers should be appropriate, stable, and predictable; monitoring is effective.

Limitations: It is very specific to the Hoosier Plan and the authors caution about external applicability.

DiGiuseppe, D. L., D. C. Aron, et al. (2001). "Risk adjusting delivery rates: A comparison of hospital profiles based on medical record and birth certificate data." Health Services Research **36**(5): 959-977.

Keywords: Cesarean delivery, performance measurement, risk adjustment

Purpose: The discrimination of two models for risk adjustment of primary Cesarean delivery are compared. The first uses variables from medical records, and the second uses variables from birth certificates. The models are used to yield hospital profiles of risk-adjusted Cesarean delivery rates. (The focus is on quality or performance measurement rather than on costs.)

Data: Twenty-nine thousand women in northeastern Ohio who had not had a previous Cesarean delivery were studied. For each to be included, both hospital records and birth certificates had to be available.

Methods: Actual and predicted Cesarean delivery rates were compared for three hospitals. Multivariate logistic regression models developed the risk-adjustment weights.

Results: The full medical record and birth certificate models were more successful than the limited common and reliable variable models. The adjusted hospital rates were similar, but the predicted outliers of the two models differed considerably.

Uses: Although the accuracy of comorbid conditions is suspect, birth certificates can be used to develop Cesarean delivery risk-adjustment models that are inexpensive and that have excellent discrimination.

Limitations: Using all birth certificate variables may yield biased comparisons. Using only reliable data elements yields rankings similar to those based on medical data.

Dow, M. G., T. L. Boaz, et al. (2001). "Risk Adjustment of Florida Mental Health Outcomes Data: Concepts, Methods, and Results." Journal of Behavioral Health Services & Research **28**(3): 258-272.

Keywords: Mental health, performance measurement, risk adjustment

Purpose: This discusses the increasingly important outcome evaluation systems for public mental health programs.

Data: All 7,929 patients at Florida state mental health facilities (excluding a few very small providers) over a three-year period are included.

Methods: The particular outcomes studied are measures of functioning and satisfaction. Risk-adjustment variables include demographics and placing patients into one of four diagnosis categories: adult disabled, crisis, forensic, and substance abuse.

Outcome comparisons for providers are made with and without risk adjustment.

Results: Overall rankings are similar but in a few cases the results are significantly different and fairer. The authors provide a good overview of risk adjustment including an unusual discussion of and response to the views of critics of risk adjustment.

Uses: This provides an introduction to risk adjustment and a discussion of satisfaction measures. Its goal is to provide a method of fairly ranking providers.

Limitations: The rankings are unreliable for smaller providers. Compliance with reporting procedures may affect the results.

Ettner, S. L., R. G. Frank, et al. (2001). "Risk Adjustment Alternatives in Paying for Behavioral Health Care Under Medicaid." Health Services Research **36**(4): 793-811.

Keywords: Capitation payments, mental health, risk adjustment, substance abuse

Purpose: The paper compares behavioral health risk-adjustment models used for mental health and substance abuse (MH/SA) capitation payments.

Data: The 1991-93 administrative data from the Michigan Medicaid program is used.

Methods: Comparisons are made between mean absolute prediction errors for several risk-adjustment models. Then comparisons are made for simulated profits and losses that behavioral health care carve outs and integrated health plans would experience under risk adjustment if they enrolled beneficiaries with a history of MH/SA problems.

Results: Models included basic demographic adjustment, Adjusted Diagnostic Groups, Hierarchical Condition Categories, and specifications designed for behavioral health. Differences in predictive ability among risk-adjustment models were small and generally insignificant. Specifications based on relatively few MH/SA diagnostic categories did as well as or better than models controlling the additional variables such as medical diagnoses at predicting MH/SA expenditures among adults. Simulation analyses revealed that among both adults and minors considerable scope remained for behavioral health care carve outs to make profits or losses after risk adjustment based on differential enrollment of severely ill patients. Similarly, integrated health plans have strong financial incentives to avoid MH/SA users even after adjustment.

Uses: One policy implication is that incentives to avoid high risk patients still exist when using these models.

Limitations: The model cannot directly set capitation rates.

Gilmer, T., R. Kronick, et al. (2001). "The Medicaid R_x Model: Pharmacy-Based Risk Adjustment for Public Programs." Medical Care **39**(11): 1188-1202.

Keywords: Medicaid, pharmacy, risk adjustment

Purpose: The authors modify a pharmacy-based risk-adjustment system for Medicaid and compare its effectiveness to the states' most popular risk adjuster, Chronic Illness and Disability Payment System (CDPS).

Data: The comparison uses extensive data from Medicaid records in four states. A third of a million disabled and a million and a half TANF or AFDC beneficiaries are in the study.

Methods: The model is developed by regressing a year's costs on the prior year's

pharmacy records. Much thought went into the drugs chosen. For the disabled, predictions for diabetes, cardiovascular illness, and mental illness are compared. Then R^2 is compared using the CDPS, the Medicaid R_x, and a combination model. Finally, predictions of group expenses for hypothetical plans are compared.

Results: Diagnostic methods are more effective for diabetes but pharmacy does better for the often underreported mental illness. Cardiovascular illness shows the advantage of using both, as either by itself misses many expensive cases. Diagnostic models are better for predicting the disabled's costs, but pharmacy comes closer to matching the predictive ability for TANF. Using both in a model does not help much with the disabled but does help predict costs for TANF beneficiaries. Results are similar when comparing predictions of plan costs.

Uses: The authors see pharmacy-based risk adjustment as a supplement to other methods due to its more complete, more timely, and more reliable data.

Limitations: Pharmacy- based systems may be easier to game than other risk adjustment. They reward plans that prescribe liberally and punish those that prescribe conservatively.

Glantz, J. C. M. (1999). "Cesarean delivery risk adjustment for regional interhospital comparisons." <u>American Journal of Obstetrics & Gynecology</u> **181**(6): 1425-1431.

Keywords: Cesarean delivery, performance measurement, risk adjustment

Purpose: The author attempts risk adjustment for Cesarean delivery rates.

Data: From a large regional perinatal database comprising 16 hospitals in New York State's Finger Lakes region, 8,229 deliveries during the first eight months of 1998 are studied.

Methods: Multiple logistic regression was used to determine and weight the significant variables. Demographic, medical, and comorbid condition variables predicted Cesarean delivery probabilities for individual patients. Probabilities were then summed over hospitals to get expected rates. Individual hospital rates were standardized by dividing the observed rate by the expected rate and multiplying by the regional rate.

Results: The regional Cesarean delivery rate was 21.9%, varying from 17.1% to 39.2% at individual hospitals. The expected range was from 18.1% to 26%. Twenty-two demographic and medical variables were associated with Cesarean delivery. Variation in observed rates for individual hospitals was less significant after risk adjustment.

Uses: Risk adjustment provides more valid hospital comparisons of hospitals' Cesarean delivery rates.

Limitations: The study has a narrow geographic scope. Unlike some similar studies, women with previous Cesarean delivery were included here with that as a variable. Some studies have focused on physicians rather than hospitals.

Goodson, J. D. M., A. S. M. Bierman, MS, et al. (2001). "The Future of Capitation: The Physician Role in Managing Change in Practice." <u>Journal of General Internal Medicine</u> **16**(4): 250-256.

Keywords: Capitation, physicians, risk adjustment

Purpose: The authors have the view that different payment plans for health care create different distributions of the patient's health risk, the physician's financial risk, and the society's financial risk, due to the different incentives for providers.

Data and Methods: Literature review. This is an opinion piece.

Results: They see physicians in a somewhat fiduciary role, especially in seeing that their patients are not harmed by perverse incentives. Capitating physician reimbursement affects the practice of medicine. Physicians must still assure that payment models do not put patients at greater health risk when accepting higher levels of personal financial risk. The authors focus on the interaction of financial incentives with medical risk.

Uses: The following checks and balances should guide physicians entering capitated contracts: (1) reimbursement must cover administrative work as well as patient encounters; (2) specialists' reimbursement is for knowledge and expertise as well as patient care encounters, and they may be reimbursed for primary care; (3) accepting risk for patient care requires managed care functions; (4) one function is coordination of the physicians treating a single patient; and (5) physicians should discuss reimbursement arrangements with patients. Accepting capitation should give physicians a greater role in the future design of health care.

Limitations: This is an opinion piece aimed at physicians.

Greenwald, L. M. (2000). "Medicare Risk-Adjusted Capitation Payments: From Research to Implementation." Health Care Financing Review **21**(3): 1-5.

Keywords: PIPDCG, provider payments, risk adjustment

Purpose: This article introduces several other articles in an issue of Health Care Financing Review devoted to risk adjustment.

Data and Methods: Literature review and introduction.

Results: All of the articles are about payment systems rather than quality or outcomes measurement. One describes the research leading to Medicare's introduction (1/1/2000) of principal inpatient diagnostic cost groups (PIPDCGs) to replace the demographic adjustment system. Several describe the implementation process. Several call for needed future or immediate improvement. All of these articles are included in this bibliography. Generally, the models all use independent variables to predict an individual's future medical costs. Older systems used only demographic variables. PIPDCG also uses the principal diagnosis for hospital admissions. Several articles point out its tendency to underestimate the costs of populations using nursing home, home health, and outpatient services. They point to its perverse incentive to reward hospital stays after others had worked to serve clients in a more cost-effective setting. Several articles discuss the improved ability of models using measures of limitations on activities of daily living (ADLs). Some of these also utilize diagnostic information but are criticized for incenting inflated diagnosis. Comorbidities are an important predictor in some models. Greenwald points out that what is left out of a model is often more key than what is included. A history of how this modeling has progressed from academic exercise to an important aspect of US health care is presented.

Uses: This is an introductory overview.

Limitations: Greenwald says there is much room for model improvement, especially with better data.

Greenwald, L. M., J. M. Levy, et al. (2000). "Favorable Selection in the Medicare+Choice Program: New Evidence." <u>Health Care Financing Review</u> **21**(3): 127-134.

Keywords: Managed care selection, Medicare, PIPDCG, ridk adjustment
Purpose: The authors contribute new evidence to the debate about whether Medicare managed care plans have selected healthier applicants.
Data: The authors analyze the initial year of Balanced Budget Act (BBA)-mandated inpatient encounter data. This is the first available data on Medicare managed care enrollees.
Methods: Average risk factors for managed care and fee-for-service enrollees are compared. The comparison is made county by county to eliminate geographic confounding.
Results: Actual managed care enrollees have lower predicted costs than fee-for-service (FFS) beneficiaries based on comparison of their average principal inpatient diagnostic cost group (PIPDCG) risk factors. This suggests bias in favor of the managed care populations. The differences are due more to health status than to demographic factors. Age, sex, and prior disability explained little of the difference. The only demographic factor that contributed to the managed care favorable selection was Medicaid eligibility. Only four of 428 counties showed lower likely costs for their FFS beneficiaries.
Uses: This is intended to resolve or at least contribute to the discussion of whether HMOs have managed to enroll more healthy members.
Limitations: The time period for studying FFS and managed care enrollees differed. Smaller managed care enrollment counties were left out of the study. The first year of data collection in a program can mislead.

Hendrex, M. S., A. Beigel, et al. (2001). "Introduction: Risk-Adjustment Issues in Mental Health Services." <u>Journal of Behavioral Health Services & Research</u> **28**(3): 225-234.

Keywords: Mental health, performance measurement, risk adjustment
Purpose: This focuses on the need for risk adjustment in quality and performance measurement in mental health services. It is an introduction to the topic and to a special issue of the <u>Journal of Behavioral Health Services and Research.</u>
Data and Methods: This is a literature introduction.
Results: Several other articles from the issue are included in this bibliography. The definition of risk adjustment (sometimes called case-mix adjustment in this context) is a means of controlling for group differences when measuring outcomes. Suggested criteria for risk variables are that they be patient-related, theoretical outcome predictors, non-treatment related, measurable, uncontrollable (non-manipulable), and discriminating. The steps necessary in developing an approach to risk-adjustment methodology are described.
Uses: Although non-financial, this article notes the financial importance of performance measurement. Certification, contracting, and payment often depend on it.
Limitations: Risk adjustment in mental health services is less developed than in other health care but is equally necessary.

Hendryx, M. S., D. G. Dyck, et al. (1999). "Risk-adjusted Outcome Models for Public

Mental Health Outpatient Programs." <u>Health Services Research</u> **34**(1).

Keywords: Risk adjustment, mental health, outpatient programs, performance
      measurement
Purpose: This focuses on nonfinancial public mental health outcomes measurement,
      but it mentions how these often have financial repercussions. Issues in risk
      adjustment are discussed, including timing issues that are particularly important
      to mental health outcomes measurement. General criteria for successful models
      are discussed.
Data: From six Washington State mental health agencies, 289 patients are interviewed.
      Provider data is also included.
Methods: The authors developed and tested risk-adjusted outcome models in publicly
      funded mental health outpatient settings using demographic, diagnostic, and
      survey variables to predict functional status, quality of life, and satisfaction.
      Model specifications were developed using linear regression analyses. Models
      were validated in a separate sample and comparative agency performance was
      examined.
Results: Variables that determine mental health outcomes are the presence of severe
      diagnoses, substance abuse, age, baseline functional status, and quality of life.
Uses: Risk-adjusted functional status and patient satisfaction outcome models could be
      developed that more accurately compare public mental health outpatient
      programs.
Limitations: Research is needed to improve the predictive accuracy of the outcome
      models developed in this study and to develop applications for field use.

Hendryx, M. S. and G. B. Teague (2001). "Comparing Alternative Risk-Adjusting
Models." <u>Journal of Behavioral Health Services & Research</u> **28**(3): 247-257.

Keywords: Mental health, performance measurement, risk adjustment
Purpose: This describes a study that compared alternative multiple regression-based
      risk-adjustment models to predict mental health outcomes.
Data: Four-hundred fifty-one adult patients from six Washington State mental health
      agencies are studied. Just 279 are in the final data- incorporating follow-up.
Methods: Demographic and diagnostic variables at the first session were used to predict
      outcomes at the follow-up. Models were calibrated. The final test was a
      comparison of agency scores and ranks by the models.
Results: In tests of six models, it is shown that the model coefficients chosen affect the
      rankings of agency performance indicating the need for much more data, theory
      development, caution, and care in model building. One particular problem
      addressed is whether survey information from the patient and caseworker should
      be included, thus getting more accuracy and more opportunity for manipulation.
      Another issue discussed and tested is whether diagnosis should be a variable or
      whether there should be different models for different diagnoses.
Uses: The need for risk adjustment is acknowledged.
Limitations: The state of the art is determined inadequate for mental health performance
      indicators. (This is less optimistic than Hendryx's 1999 article.)

Iezzoni, L., J. Ayanian, et al. (1998). "Paying more fairly for Medicare capitated care."

New England Journal of Medicine **339**(26): 1933-8.

Keywords: Capitation, PIPDCG, risk adjustment
Purpose: This is an opinion piece directed at physicians. It urges them to resist perverse incentives and to follow risk-adjustment developments.
Data and Methods: Various secondary sources are quoted.
Results: The Health Care Financing Administration (HCFA) has previously risk adjusted capitation based on demographic information (adjusted average per capita cost-AAPCC). The 1997 Balanced Budget Act (BBA) specifies future adjustments based on clinical information, as well. The authors say that capitated plans have been motivated to seek healthier clients and have thus been overpaid by 5 to 20%. The authors urge physicians to participate in the future evolution of risk adjustment and to resist the new incentive to hospitalize. They give an insider's view of the development of the principal inpatient diagnostic cost group (PIPDCG) categories. They also discuss new reporting requirements for capitated plans.
Uses: They say lack of data is the real limiting factor in developing a payment system that eliminates incentives to enroll only the healthy or to provide unnecessary care.

Ingber, M. J. (2000). "Implementation of Risk Adjustment for Medicare." Health Care Financing Review **21**(3): 119-127.

Keywords: Medicare, PIPDCG, risk adjustment
Purpose: This article describes the implementation of the Principal Inpatient Diagnostic Cost Group (PIPDCG) model.
Data and Methods: This is a historical narrative.
Results: The Health Care Financing Administration (HCFA) began phasing in risk adjustment for Medicare capitated organizations as of January 2000. The risk-adjustment system used, PIPDCG, was incorporated into the existing system and the several constraints required by the Balanced Budget Act of 1997 (BBA). BBA specified budget neutrality, for example. The description of the work of HCFA actuaries in developing new ratebooks will be especially interesting to actuarial readers.
Uses: This explains and increases appreciation of the blend of data, methods, and politics required to implement something like PIPDCG. Ingber expects better management of patients and costs to be a side benefit of PIPDCG.
Limitations: Fee-for-service data was used to set managed care rates. This will be improved in the future. The transition to PIPDCG is slow (10% for the first two years).

Kronick, R., T. Gilmer, et al. (2000). "Improving Health-Based Payment for Medicaid Beneficiaries: CDPS." Health Care Financing Review **21**(3): 29-64.

Keywords: Capitation, Medicaid, risk adjustment
Purpose: The authors describe Chronic Illness and Disability Payment System (CDPS). It is a Medicaid diagnostic classification system that is used to capitate payments for Temporary Assistance to Needy Families (TANF) and disabled Medicaid

beneficiaries.

Data: Four million records from seven states were used in the model development. 3.3 million were from TANF programs and 0.6 million were disabled.

Methods: Regression analysis was performed on a very large data set split into calibration and validation parts.

Results: The state Medicaid programs have actually been leaders in using risk-adjusted capitation rates for providers. The program is an update of the earlier DPS with many more diagnostic codes included. The authors describe the diversity of diagnoses and different burdens of illness among disabled (whose costs are more accurately predicted, $R^2 = .18$) and AFDC Medicaid beneficiaries ($R^2 = .08$ for adults and .04 for children). The authors also compare the variables and performance of CDPS with other models and find it outperforms hierarchal condition category (HCC) and adjusted clinical group (ACG) models.

Uses: Managed care payment weights are provided for the States' use.

Limitations: States may wish to modify the weights to suit their situations. Getting consistent data is still a problem both for model development and for applying factor weights.

McCall, N. and J. Korb (1998). "Risk Adjustment for Dually Eligible Beneficiaries Using Long-Term Care." Health Care Financing Review **20**(2): 71-90.

Keywords: HCC, Medicaid, Medicare, PIPDCG, risk adjustment

Purpose: This article describes data and risk adjustment on dually eligible Medicare and Medicaid recipients that are long-term care qualified but receiving home or outpatient care.

Data: Arizona and New Mexico fee-for-service beneficiaries are studied. They numbered about 5,000 with all data available.

Methods: The model predictions of total expense are compared with actual Medicare and Medicaid expense.

Results: Those dually eligible for Medicare and Medicaid have high rates of disability and higher costs (double) compared to those eligible for only one of the two programs. There is also much state legislative activity trying to coordinate the two programs and eliminate perverse incentives. This focuses on the subset that are also long-term care qualified but receiving home health care or outpatient care. It presents information on a study that looks at the use of the principal inpatient diagnostic cost group (PIPDCG) risk-adjustment methodology planned for use by the Health Care Financing Administration in implementing risk adjustment in 2000. This study compares the use of PIPDCG and an alternative hierarchical coexisting conditions (HCC) risk-adjustment methodology. The HCC considers all inpatient and outpatient diagnoses and their cumulative effect. Measures of individual predictive accuracy (for Medicare and total costs) for this population compared with the total Medicare population were similar for the PIPDCG models but somewhat smaller for the HCC models. Activities of daily living (ADLs) variables increased the R^2 values more for Medicaid than for Medicare.

Uses: This could inform opinion on program design. The target group will be underfunded by PIPDCG.

Limitations: The need for more data and more current data was acknowledged.

Myers, E. R. M., MPH and J. F. M. Steege (1999). "Risk adjustment for complications

of hysterectomy: Limitations of routinely collected administrative data." <u>American Journal of Obstetrics & Gynecology</u> **181**(3): 567-575.

Keywords: Hysterectomies, performance measurement, risk adjustment
Purpose: This article focuses on quality measurement for hysterectomies.
Data: From a large database, risk-adjusted rates of complication are developed using routinely collected administrative data (including ICD-9 codes) for risk adjustment.
Methods: Logistic regression models are constructed for the prediction of both medical and surgical complications. Only complications manifesting prior to discharge are in the dependent variable. Independent variables are demographic, diagnostic, and procedural data. The area under the ROC curve rather than the usual regression R^2 is used to summarize predictive ability.
Results: Overall, twelve percent of hysterectomies had surgical complications and seven percent had other medical problems. (About three percent had both.) Hysterectomy type, hospital type, age, and insurance status of Medicaid or no insurance were significant predictors of complications. Procedures performed for cancer or pregnancy complications had higher risk. The comorbidity models were better predictors of medical complications (C = 0.714) than surgical complications (C = 0.630). Surgical complications of hysterectomy are more common than medical complications, but are less predictable from administrative data. There is a very wide rate of complication at the different hospitals in the study.
Uses: This can inform opinion and contribute to accurate measurement of the variance.
Limitations: Lack of uniformity and accuracy in coding is discussed. There was difficulty distinguishing comorbidities from complications. The authors conclude that these risk-adjusted complication rates cannot be used to distinguish hospitals' quality. The authors note that this is focused on a younger (median age 41), healthier population than most risk-adjustment studies.

Pope, G. C., K. W. Adamache, et al. (1998). "Evaluating Alternative Risk Adjusters for Medicare." <u>Health Care Financing Review</u> **20**(2): 109-125.

Keywords: Medicare, risk adjustment
Purpose: This evaluates nine model variations using alternative demographic, survey-based and claims-based independent variables for risk adjustment for Medicare capitation payment.
Data: It is based on Medicare Current Beneficiary Survey (MCBS) data. This is a large ongoing survey of Medicare beneficiaries.
Methods: It describes but rejects a unique two part regression model designed to deal with the usual distribution of claims...many zeros and a few very large individual amounts. An alternative to a split sample design (using different years' data) is used to avoid overfitting the models.
Results: The survey health-status models explain three to four times the variation explained by the current demographic model, the adjusted average per capita cost (AAPCC). The risk-adjustment model derived from actual diagnoses explains 75 percent more variation than a comprehensive survey model. None of the models predict average expenditures well for all beneficiary subgroups described herein.

Uses: This suggests that a combined model (after eliminating redundant variables) may
   be appropriate and also suggests this is an area of future research.
Limitations: Practical and administrative concerns (manipulation, expense) with the
   alternative risk adjusters are discussed.

Pope, G. C., R. P. Ellis, et al. (2000). "Principal Inpatient Diagnostic Cost Group Model
for Medicare Risk Adjustment." Health Care Financing Review **21**(3): 93-120.

Keywords: Capitation, PIPDCG, risk adjustment
Purpose: The Balanced Budget Act (BBA) of 1997 mandated improved risk adjustment
   for Medicare capitation payments for managed care plans by January 1, 2000.
   The authors describe Principal Inpatient Diagnostic Cost Group (PIPDCG), the
   risk-adjustment model that the Health Care Financing Administration (HCFA)
   began to implement in 2000.
Data: HCFA has collected inpatient data from health plans since 1997. The data
   includes initial diagnoses and other data useful for predicting population costs.
Methods: The model variables, coefficients, application, rationale, and some of the ten-
   year history of its development are described in detail.
Results: Ten percent of capitation costs were based on PIPDCG in 2000 with the
   percent scheduled to increase. The PIPDCG model was developed as a
   practical, interim step on the way to recognition of all patient health care
   encounter data.
Uses: This provides a basic understanding of PIPDCG and its development.
Limitations: PIPDCG is due to be replaced. It is conservative, adjusting payment for the
   health status of only 12% of enrollees.

Riley, G. F. (2000). "Risk Adjustment for Health Plans Disproportionately Enrolling Frail
Medicare Beneficiaries." Health Care Financing Review **21**(3): 135-149.

Keywords: Medicare, PIPDCG, risk adjustment
Purpose: The Principal Inpatient Diagnostic Cost Group (PIPDCG) model uses
   demographics and inpatient diagnoses and was the initial Medicare risk-
   adjustment method. Inpatient diagnosis-based risk adjusters will underestimate
   the costs of health plans (such as PACE) that disproportionately enroll frail
   Medicare beneficiaries. The authors test two risk-adjustment models that could
   be used to predict Medicare costs for groups in other institutional status and by
   degree of difficulty with activities of daily living (ADLs).
Data: The extensive Medicare Current Beneficiary Survey is the source of sample
   (32,000 records) data.
Methods: Model costs are compared with actual.
Results: The second model (besides PIPDCG) uses hierarchical coexisting condition
   risk scores. Noninstitutionalized frail beneficiaries' costs were underestimated by
   each model due to missing the outliers. Incentives for plans to selectively target
   their membership exist after risk adjustment by either of these models.
Uses: This can contribute to future risk-adjustment design.
Limitations: Further development of risk adjustment is needed to fairly compensate
   plans that enroll frail beneficiaries. However, the authors caution that ADLs are
   likely not the way to improve the models as there is too much variation in costs

within ADL groupings. Further development of data on home health care use, dual enrollment (with Medicaid), and skilled nursing facility use will lead to better adjustment.

Robinson, J. and S. L. Karon (2000). "Modeling Medicare Costs of PACE Populations." Health Care Financing Review **21**(3): 149-170.

Keywords: Medicare, PACE, risk adjustment

Purpose: The article discusses proper risk adjustment for Program for All-Inclusive Care (PACE) providers. PACE enrollees are nursing home certified under local rules, but receiving other care.

Data: The Medicare Current Beneficiary Survey and the National Long-Term Care Survey are the primary sources.

Methods: Historical PACE adjustment is compared with a more appropriate frailty adjuster. Fee-for-service costs are estimated for a PACE population.

Results: Historically, PACE providers have received payment equal to 95 percent of average adjusted per capita cost (AAPCC) multiplied by a frailty adjuster of 2.39. The Balanced Budget Act of 1997 gave PACE permanent status. Its future payments are to be under the rate structure of the Medicare+Choice payment system, adjusted for frailty and other factors. The history and the calculation of the PACE frailty adjuster is reviewed here. Possible means of risk adjustment and their impacts are explored. After eliminating numerous possible predictive variables, sex, disabled v. elderly, functional/cognitive status, and recent Medicare claims were retained in the model. The appropriate frailty adjuster is shown to vary widely depending on the base to which it will be applied. It is also suggested that it vary geographically.

Uses: This can inform policy on PACE risk adjustment.

Limitations: PACE definitions vary by state.

Romano, P. S. and B. K. Chan (2000). "Risk-Adjusting Acute Myocardial Infarction Mortality: Are APR-DRGs the Right Tool?" Health Services Research **34**.

Keywords: APR-DRG, mortality, performance measurement, risk adjustment

Purpose: Focused on performance or quality measurement, this research seeks to determine the effectiveness and accuracy of a widely used proprietary risk-adjustment system, APR-DRGs, Version 12.

Data: Over 100,000 adult California patients are studied.

Methods: The study evaluated its mortality predictions. Input variables were taken from different sources of checked and corrected ICD-9-CM data. APR-DRG Risk of Mortality and Severity of Illness classes were assigned based on (1) all hospital-reported diagnoses, (2) all corrected diagnoses, and (3) corrected diagnoses present at admission.

Results: The APR-DRG Risk of Mortality class is an accurate predictor of death. c, the area under the ROC curve, = .831-.847. Adding demographic variables improved these results. The difference in c when using only initial rather than all diagnoses was large (c = .74 versus c = .93). Predicted costs for different providers was similar for the two models.

Uses: The authors conclude that the APR-DRG Risk of Mortality system is an accurate

risk-adjustment tool, largely because it includes all relevant diagnoses, not just those at admission.

Limitations: Other modeling efforts include only diagnosis at admission because later diagnosis may have been due to or a result of treatment. It is also self-referential.

Temkin-Greener, H., M. R. Meiners, et al. (2001). "PACE and the Medicare+Choice risk-adjusted payment model." <u>Inquiry</u> **38**(1): 60-72.

Keywords: Medicare, PACE, PIPDCG, risk adjusment

Purpose: This study investigates the potential impact of the Medicare principal inpatient diagnostic group (PIPDCG) payment model if applied to the Program of All-Inclusive Care for the Elderly (PACE). Almost 7,000 Medicare beneficiaries who are nursing home certifiable receive care from PACE, a national demonstration program set for provider status under the Balanced Budget Act of 1997 (BBA). PACE programs assume full financial risk for enrollees.

Data: Several data sources were used including the Medicare Current Beneficiary Survey and PACE data.

Methods: This study determines the reimbursements that twelve PACE sites would have received under PIPDCG as well as their costs.

Results: The PIPDCG payment model bases its risk adjustment on inpatient diagnosis and does not capture adequately the risk of caring for a population with functional impairments. This population requires home health care and skilled nursing facility services. BBA calls for adjustment of payments to PACE providers but does not specify the methodology. Overall PACE program payments will be reduced by 38%. The PIPDCG payments to PACE are inadequate for those with any activities of daily living (ADL) limitations and severely inadequate (60%) for those with five or more ADL limitations. The PIPDCG capitation model will encourage unnecessary hospitalization which PACE has been demonstrated to reduce.

Uses: The authors call for PIPDCG model improvement in three areas: (1) an ADL-specific adjuster, (2) home health and nursing home cost predictors, and (3) recognition of severity within current diagnostic codes. Expansion of PACE will be limited until adjustment is adequate.

Limitations: Site level analysis was hindered by the small sample size.