American Academy of Actuaries
Risk Adjustor Work Group

ACTUARIAL REVIEW OF THE HEALTH STATUS RISK ADJUSTOR METHODOLOGY

Risk Adjustor Work Group

William F. Bluhm, MAAA, FSA, FCA, Chair
Robert C. Benedict, MAAA, FSA
John M. Bertko, MAAA, FSA, FCA
Robert B. Cumming, MAAA, FSA
Patrick J. Dunks, MAAA, FSA
Kathleen E. Ely, MAAA, FSA
Alan D. Ford, MAAA, FSA, EA
P. Anthony Hammond, MAAA, ASA
William R. Lane, MAAA, FSA
Lucinda M. Lewis, MAAA, FSA
David R. Nelson, MAAA, FSA
Donna C. Novak, MAAA, ASA
Leslie F. Peters, MAAA, FSA
Martin E. Staehlin, MAAA, FSA
Jill A. Stockard, MAAA, FSA

The American Academy of Actuaries is the public policy organization for actuaries of all specialties within the United States. In addition to setting qualification and practice standards, a major purpose of the Academy is to act as the public information organization for the profession. The Academy is nonpartisan and assists the public policy process through the presentation of objective analysis. The Academy regularly prepares comments on proposed federal regulations, and works closely with state officials on issues related to insurance. The Academy also develops and upholds actuarial standards of conduct, qualification and practice, and the Code of Professional Conduct for all actuaries practicing in the United States.

January 14, 1999

Table of Contents

I. Executive Summary

This report presents the analysis by the American Academy of Actuaries' Risk Adjustor Work Group of the health status risk adjustment methodology proposed for implementation by the Health Care Financing Administration (HCFA). By law, a risk adjustment methodology must be used by HCFA for determining payments to Medicare+Choice health plans starting in the year 2000. The Risk Adjustor Work Group was formed in response to a request from HCFA for an actuarial review of their work on a risk adjustment methodology. This review is required by the Balanced Budget Act of 1997 which mandated HCFA's development of a risk adjustment payment system.

The adoption of a new risk adjustment system for Medicare reimbursements based on health status factors represents a significant change for health plans, contracting providers and health plan members. ***While the Academy Work Group believes the conceptual basis of the risk adjustment method proposed by HCFA is "actuarially sound" as defined in our report, we have serious concerns about the method's implementation, operation and impact.***

The new methodology for making health status risk adjustments to Medicare payments appears to meet the requirements of the Balanced Budget Act of 1997, provided the system is implemented carefully. On balance, and with a phase-in, the proposed risk adjustment method appears to be a reasonable first step in what should be a long-term evolutionary process. HCFA is to be commended for the progress to date and for recognizing the limitations of the proposal arising from the available data, timing requirements and areas for future improvements.

While HCFA has done much work in a short time period to develop the new methodology and design implementation strategies, additional work remains to fully define HCFA's risk adjustment method and test application of the method to make sure it achieves the intended results. The Work Group recommends that HCFA further modify the risk adjustment model with the knowledge gained during the first year of operation.

Based on our review of the information and data provided by HCFA, the Work Group has serious concerns about the actual implementation of the new payment system and its impact on the Medicare health plan market. These issues include:

- *Decisions to exclude or limit the use of certain types of diagnosis categories from the risk adjustment methodology, such as one-day hospital stays, which may penalize health plans that effectively manage the delivery of health care.*

- *Lack of adequate testing of the potential impact of the new methodology on health plans and Medicare+Choice beneficiaries.*

- *Administrative feasibility of the implementation of the new system because of timing and data collection issues.*

- *The processing of extraordinary amounts of newly collected data and completing a series*

of complex calculations introduces an element of uncertainty that cannot be anticipated until health plans and HCFA have full opportunity to understand the implications.

- *Use of only fee-for-service data as the basis for the development of risk adjustment weights.*

During the review process, HCFA provided the Work Group with preliminary results of the potential payment impact of the risk adjustment methodology on Medicare+Choice plans. However, the Work Group was not able to verify the accuracy of the data collected by HCFA or the calculations used by HCFA to determine the impact on health plans. In addition, HCFA did not provide the Work Group with an assessment of the impact of the risk adjustment methodology on beneficiaries.

There is a substantial risk for the Medicare system if the risk adjustment methodology does not work as intended. The negative consequences could include withdrawal of Medicare+Choice health plans from the market, financial problems or insolvency for health plans and the potential for a reduction in benefits provided to beneficiaries. Because of these concerns, the Work Group believes HCFA's decision to implement the new methodology under a phased-in approach is a sound one and will limit changes from the current payment system while HCFA and the health plans assess the impact of the new methodology.

The Work Group was unable to fully analyze the proposed risk adjustment method due to: (1) incomplete available data and information, and (2) the continuing development of the new risk adjustment methodology by HCFA. The Work Group was not able to undertake a detailed analysis of the mathematical formulas used to develop the risk adjustment methodology, but rather focused its review on the conceptual and theoretical basis of the system. Because HCFA is still working on the proposed methodology and there are a number of unresolved implementation issues, this report is a qualified review of the actuarial soundness of the proposal. The Work Group would like the opportunity to provide further comments on the new system as it is completed.

II. Introduction

The American Academy of Actuaries (Academy) has been asked by the Health Care Financing Administration (HCFA) to evaluate its proposed method for using the health status of Medicare beneficiaries to risk adjust Medicare payment rates. The Academy formed a Risk Adjustor Work Group (Work Group) consisting of health actuaries who are consultants to health plans and health insurers and staff actuaries for health plans or health insurers to review HCFA's proposal. This report presents the Work Group's analysis, conclusions and recommendations.

The Balanced Budget Act of 1997 (BBA) requires HCFA to incorporate health status risk adjustment

in the agency's payments to Medicare+Choice health plans.[1] The law also provides that HCFA will report to Congress on its proposed method for risk adjustment. The purpose of this report is to assist HCFA in satisfying Section 1853(a)(3)(A) of the Act, which states that HCFA's Report to Congress shall include, "*an evaluation of such method by an outside, independent actuary of the actuarial soundness of the proposal.*"

HCFA plans to implement the initial health status risk adjustment method on January 1, 2000 and then replace it with a more comprehensive method at a later date. The scope of the Academy's report includes both the initial Principal Inpatient Diagnostic Cost Group (PIP-DCG) method and the possible subsequent modifications to the methodology proposed by HCFA.

The Work Group analyzed the actuarial soundness of HCFA's risk adjustment proposal in terms of (1) established actuarial goals and criteria for risk adjustment, (2) Actuarial Standards of Practice, and (3) the general principles and practices of actuarial science. Actuarial Standards of Practice are guidelines developed by the Actuarial Standards Board to help actuaries in their work. Specific actuarial goals and criteria for risk adjustment are described in the Academy's May 1993 monograph titled, "Health Risk Assessment and Health Risk Adjustment: Crucial Elements in Effective Health Care Reform" (Health Risk Adjustment Monograph).

The Academy understands that this report will be used by both Congress and HCFA and will become part of the public record. The Academy also understands the report may be provided to other interested parties. This report should only be distributed in its complete form.

III. Report Background and Methodology

A. Overview of Risk Adjustment

Health risk adjustment is a means of modifying or redistributing payments received by risk bearing entities within a health insurance system to more equitably compensate those entities for the risks they have assumed relative to one another. A major purpose of a health risk adjustment system is to make the basis of competition among carriers their administrative and medical efficiency rather than the health plans' ability to select healthy people.

The risk adjustment process uses the results of health risk assessment to determine the appropriate magnitude of revenue adjustments. Health risk assessment is a method for objectively determining the relative health risks (or expected relative costs) of individuals or groups of individuals relative to an average. Risk is assigned as a simple numerical value or score reflecting the relative cost of health care resources required to meet the total health care needs of that individual or group. A key goal of the risk adjustment process is to more equitably match financial reimbursement with financial liability within an insurance system.

[1]Section 1853 of the Balanced Budget Act of 1997 (PL 105-33).

B. HCFA's Proposal

Currently, HCFA's published local payment rates for Medicare+Choice health plans are adjusted to reflect the risk characteristics of the plans' participants in a particular county as defined by demographic factors: age, gender, status (institutionalized or non-institutionalized, Medicaid or non-Medicaid and Working Aged). Beginning in the year 2000, HCFA is required by the Balanced Budget Act of 1997 to supplement demographic adjustments with a health status risk adjuster.

The PIP-DCG risk adjuster methodology was developed for HCFA by researchers at Health Economics Research, Inc. (HER), Boston University, and Harvard Medical School. The PIP-DCG risk adjuster will be used to assign each Medicare beneficiary a risk score based on diagnosis information from hospital inpatient stays. These risk scores, along with county of residence, age, gender, and other factors, will be used on a prospective basis to determine the Medicare payment rate for each beneficiary in a Medicare+Choice health plan.

As part of HCFA's proposed risk adjustment method, HCFA will be "rescaling" or adjusting the base payment rates. The purpose is to ensure that for each county, the new reimbursement rates utilizing health status risk adjusters should produce the same total payments for the fee-for-service populations as the current approach if every Medicare FFS member in a county were enrolled in a Medicare+Choice organization. However, implementing risk adjusters could increase or decrease total Medicare payments to health plans depending on whether Medicare+Choice organizations currently enroll a higher or lower than average share of the less healthy Medicare beneficiaries.

C. Work Group's Method of Review

The Academy was asked by HCFA to evaluate the "actuarial soundness" of its proposal. Although there is no widely recognized definition of "actuarial soundness," the Work Group analyzed HCFA's proposal according to the standards for risk assessment and risk adjustment outlined in the Academy's Health Risk Adjustment Monograph. These criteria are:

Accuracy: Since payments to health plans will be determined based on the risk adjustment mechanism, accuracy and avoidance of statistical bias is critical.

Practicality and Reasonable Cost: The risk adjustment mechanism should not be so complex that implementation is extremely cumbersome, thereby adding significant cost to the system.

Timeliness and Predictability: Carriers setting premium rates should be able to predict the impact of risk adjustment on their premiums with a fair degree of accuracy and in a timely manner, in order to avoid solvency concerns and disruption to members.

Resistance to Manipulation: The risk adjustment mechanism should aim to make it impossible for

specific carriers to benefit financially by "gaming" the mechanism.

In addition, the Academy has assessed the effectiveness of the proposed methods in achieving the *goals of risk adjustment* as outlined in the Health Risk Adjustment Monograph. These goals are:

- *Reducing the effects of either inadvertent or intentional risk selection*, so carriers in a competitive market can compete on the basis of medical and administrative efficiency and the quality of service and care, rather than on the ability to select risk;

- *Compensating carriers fairly and equitably for risks they assume*;

- *Maintaining consumer choice* between multiple health plans based on rates or employee contributions that reflect relative medical and administrative efficiencies; and

- *Protecting the financial soundness of the health care system*.

The Academy's review takes into account all aspects of the proposed methodologies that impact on its "actuarial soundness," including, but not limited to:

- The proposed formulas;

- The availability, quality, and relevance of the data required; and

- The ability to be implemented as intended.

In addition, the Academy has evaluated the appropriateness of the proposed methods in relation to available alternatives (including non-administrative data models such as surveys, enhanced age/gender/status, and the status quo) and in light of the modifications being made to the underlying base rates by county over the same time period.

D. Limits of the Work Group's Analysis

In preparation for this analysis, the Academy's Work Group met with representatives of HCFA to establish the purpose and scope of the evaluation to be provided. During the meeting, HCFA staff provided an overview of their proposed methodology that they indicated was still in draft form.[2]

[2] During the meeting with HCFA staff, the Work Group was provided with materials outlining the current Medicare payment system and the proposed PIP-DCG methodology including a technical paper titled, "Risk Adjustment for the Medicare Program: Lessons Learned from Research and Demonstrations" by Leslie M. Greenwald, PhD, Al Esposito, MS, Melvin J. Ingber, PhD and Jesse M. Levy, PhD. All of the authors are with HCFA's Office of Strategic

The agency's staff also stated they would be receptive to the Academy's suggestions for modifications to the methodology that would improve its soundness or effectiveness. Shortly before this review was finalized, HCFA also introduced three modifications to the proposal that the Work Group had reviewed. While the Work Group sees no immediate major implications of those three changes (other than as already discussed in this report), we have not had the opportunity to analyze the changes in depth.

A preliminary draft of this report was provided to HCFA for review and comment on December 8, 1998. The agency responded and submitted additional materials and data to the Work Group along with specific comments concerning some of the issues raised in the report. The Work Group considered HCFA's response and materials as it completed its work on this report.

In performing this review, the Work Group relied upon information and data provided by HCFA. This information included descriptions of the statistical methodologies summarized by Health Economics Research,[3] reports and other summary materials, answers to written inquiries submitted by the Academy,[4] and an initial version of the methodology established by HCFA during the course of this review. In addition, the Academy relied upon descriptions of the methodologies published in the Federal Register.[5]

It is important to note that the analysis and conclusions in this report are dependent on the information supplied to the Work Group by HCFA. As of the date of this report, HCFA has not provided the final version of the PIP-DCG risk adjustment formula. In addition, only preliminary revisions of the comprehensive data methods for the risk adjustment methodology have been discussed. No formal methods have been released. **Changes in the methodology or adjustments to the data and information provided by HCFA to the Academy could dramatically impact the findings of this report.** The development of a risk adjusted payment system by HCFA is still a "work in progress" and **this report reflects a qualified opinion by the Academy on the "actuarial**

Planning (Research and Evaluation Group).

[3]The Work Group reviewed several reports from HER: (a) Diagnostic Cost Group (DCG) and Hierarchical Coexisting Conditions (HCC) Models for Medicare Risk Adjustments (Volumes I and II, April 26, 1996); (b) Revised Diagnostic Cost Group (DCG)/Hierarchical Coexisting Conditions (HCC) Models for Medicare Risk Adjustment (February 6, 1998) and © Updated and Revised Principal Inpatient Diagnostic Cost Group Models (Draft Report, July 17, 1998).

[4]HCFA provided three submissions to the Work Group (dated October 8, 1998, November 3, 1998 and November 19, 1998) in response to requests for additional information. HCFA also submitted additional materials to the Work Group on December 14, 1998, December 21, 1998, December 29, 1998 and January 5, 1999, as part of its response to a draft copy of the report reviewed by HCFA.

[5]Federal Register, Vol. 63, No. 173 (September 8, 1998).

soundness" of HCFA's proposals based on the information available to the Work Group at the time it performed its review. The limitations on the Work Group's analysis and findings are applied throughout the report.

In addition, the Work Group did not undertake an analysis of the specific mathematical formulas used by HCFA in the development of risk scores and was not able to determine the accuracy of HCFA's application of the risk adjustment methodology to the data collected from Medicare health plans. As a result, this report should not be considered a "peer review" of the risk adjustment formula under which the Work Group would have examined the mathematical processes used to develop the health status risk scores. The Work Group's analysis is limited to the conceptual framework of the risk adjustment methodology developed by HCFA.

IV. PIP-DCG Risk Assessment Model

A. Key Components of Model

Use of Only Inpatient Data

As previously discussed, the initial model developed by HCFA will use inpatient diagnostic data to develop a risk score for each Medicare beneficiary.[6] This information is based on hospital inpatient stays over one day for certain diagnosis groups. The risk score will be combined with revised demographic factors to develop the payment rates.

A significant component of the PIP-DCG model is the restriction of the risk adjustment method to conditions identified by inpatient hospital claims. This feature has both advantages and disadvantages. As one positive factor, this requirement matches well with the information currently available to the Medicare program. Currently, hospital claim information is more accessible and easier to audit than ambulatory care data and requires a lower amount of additional work by health plans to report to HCFA.

However, there are several drawbacks to a system that uses only inpatient data. First, it is possible that a system relying on inpatient data may penalize plans which more efficiently manage health care. A major feature of managed care has been the measurable reduction in use of inpatient care, the shifting of that care to other, more cost-effective sites of service and the substitution of less invasive therapies to treat a given condition. When the risk assessment system is restricted to inpatient claims, the members enrolled in managed care can appear healthier than their actual risk level because of limits on what is measured.

---

[6]Individuals who are newly eligible for Medicare will be assigned a risk score based on HCFA's analysis of existing Medicare fee-for-service data. HCFA will construct a special set of risk scores for these individuals which estimates their predicted medical expenditures since they will not have any inpatient claims experience under the Medicare system.

A possible ramification of the PIP-DCG method is that, by using only inpatient data, the risk adjustment method is less effective than one that also includes ambulatory data because the PIP-DCG formula measures less risk (based on benchmarks such as the R-Squared statistic and predictive ratios). If ambulatory data is added to the inpatient claims information, a better picture of the potential "risk" of each individual Medicare beneficiary is obtained. The PIP-DCG methodology may result in a smaller variation in risk-adjusted payments made to Medicare+Choice health plans than would occur if a more comprehensive method were used. In the initial phases of the program this may be desirable, to the extent it causes less disruption to plans which participate in Medicare+Choice.

Principal Diagnosis

The PIP-DCG model measures conditions by capturing the principal diagnosis recorded on each inpatient claim. The use of the principal diagnosis for the PIP-DCG model is based on existing coding practices for inpatient claims used by hospitals. Since only the principal diagnosis is generally used, it is possible that not all appropriate information is collected or used. A qualifying condition could be listed as the secondary (or other) diagnosis which could be a contributing factor leading to the need for hospitalization.

For example, a hospital admission for an acute condition caused or exacerbated by hypertension or diabetes may identify that acute condition as the principal diagnosis even though it could be argued the patient would not have been hospitalized had it not been for the underlying chronic condition. There is also the possibility that restricting the data source to principal diagnosis could lead to listing a qualifying event as a "principal diagnosis" on a claim in order to receive "credit" for that more serious inpatient condition.

Alternately, there is a common belief that many secondary conditions currently reported are not as reliable and should not be included in the measurement system. Since the initial stages of the risk assessment system will be using data that was recorded without the presence of direct coding incentives, it may be reasonable to use only principal diagnosis information. However, as the PIP-DCG system is implemented, the restriction to using only principal diagnostic groups should be re-evaluated.

Number and Development of PIP-DCG Groups

Health Economics Research constructed the diagnostic groupings using HCFA's survey of Medicare FFS data (a sample of 5% of Medicare beneficiaries). The claims and eligibility information for this analysis fell in the two-year interval from January 1, 1995 through December 31, 1996. Beneficiaries who were not alive and enrolled in Medicare for the entire period from March 1, 1995 through December 31, 1995 and enrolled on January 1, 1996 were excluded from the sample. Beneficiaries were removed from the data sample if they would not have been eligible

for coverage by a Medicare+Choice program for various reasons.[7]

The 5% sample is itself a bit of a misnomer. There were approximately 37.3 million Medicare beneficiaries on July 1, 1995. A 5% sample should yield around 1.9 million lives. However, after excluding beneficiaries based on length of eligibility or future managed care plan participation requirements, the sample used in setting the PIP-DCG risk adjusters is 1.4 million, which is 25% smaller. Therefore, the 5% sample is really roughly a 3.5% sample.

HER used diagnostic codes to form the diagnostic groups (DxGroups) which are used in the PIP-DCG methodology. In order to qualify for a DxGroup used in the PIP-DCG formula, there must have been at least 1,000 individuals with that diagnosis in the Medicare 5% sample. When the PIP-DCG methodology is used starting in 2000, those Medicare beneficiaries who do not fall into one of the DxGroups will be classified into a "base" group and they will be scored only on the demographic risk factors.[8]

There are questions with respect to data credibility and the design of diagnostic groups based on the HCFA 5% sample. For example, is it appropriate to use 1,000 individuals as the "cut-off" for forming DxGroups? How different would the resulting DxGroups look if the sample had been 10% of Medicare beneficiaries, or if a different 5% sample had been selected?

A second sample could be drawn to test the variability of PIP-DCGs. One possible approach would be to use a stratified sample which examines a higher percentage of beneficiaries with claims in the diagnostic categories that make up the PIP-DCG formula. While the overall number of Medicare beneficiaries in the sample might be less than 5%, the survey could sample a greater number of health plan members who fall into one of the claims categories that make up the diagnostic groups and greatly increase the effectiveness of the process. Since the goal of the sample would be to examine the cost of claims and not necessarily claim frequency, a stratified sampling would seem to be a useful tool.

In addition, the requirement to utilize DxGroups with at least 1,000 members may be overly conservative. While this level of robustness certainly contributes to the credibility of each DxGroup, it may result in risk assessment values that are not as widely dispersed as the underlying distribution of health risk. Relaxing the restraint that DxGroups have at least 1,000 members may result in use of more DxGroups, a likely higher maximum value and a more continuous distribution of resulting risk assessment values (i.e., there would be more DxGroups with smaller increments between categories).

---

[7] For a complete description of the sampling technique, see Chapter 2 of HER's draft report, "Updated and Revised Principal Inpatient Diagnostic Cost Group Models" (July 17, 1998).

[8] The development of the diagnostic groups is discussed in Chapter 4 of the HER draft report dated July 17, 1998.

Because of this requirement, diagnoses could be paid at radically different levels depending upon the adequacy of the sample size when the risk adjusters are established. For example, under the current PIP-DCG modeling with no discretionary diagnosis exclusions, the highest DxGroup is formed at expenditure level 32 (approximately $32,000). There were an insufficient number of beneficiaries to form PIP-DCGs until level 23 (approximately $23,000). The information available to the Work Group did not state how many beneficiaries had PIP DxGroups at levels 24 through 31. Similarly, reversing the exclusion of PIP diagnoses with under 50 people could have the effect of restoring some bona fide conditions to the list of risk adjuster diagnoses.

The Work Group recommends that HCFA reexamine the decision to construct DxGroups using the 1,000 member cut-off once they have started collecting information from Medicare+Choice plans and have implemented the new risk adjustment payment system. The validity of this size criteria can then be based on more current inpatient data on Medicare beneficiaries.

Exclusion of "Discretionary" Conditions

The base cost group also includes Medicare beneficiaries with diagnoses that were determined by HER to be discretionary, vague, or which only occasionally resulted in inpatient admissions. This exclusion of those "discretionary" conditions has the beneficial effect of reducing potential bias in the formula against Medicare+Choice health plans with well managed care delivery systems by not giving credit for discretionary admissions and by removing the incentives to hospitalize a patient for minor illness. The diagnoses included in this restriction should be reviewed in the future as coding practices change under the PIP-DCG system. If hospitals become more aggressive in their coding in the future, the percentage of claims falling into a PIP-DCG may change and weights would need to be recalibrated, particularly if the PIP-DCG method is used beyond the currently planned three-year period.

Exclusion of 1-Day Hospitalizations

The HER report recommends excluding one-day hospitalizations from the risk assessment system to avoid giving credit for very short stays, under the assumption that including them may result in "gaming" of the system by health plans. Plans could "game" the system by ordering one-day stays for minor medical conditions in order to include beneficiaries in the health status risk adjustment process.

The prohibition against using one-day stays may result in lower risk scores being given to members in plans which efficiently manage the delivery of health care because such plans generally have shorter lengths of hospital stays. The HER report asserts that this exclusion results in only 5% of otherwise qualifying diagnoses being removed from the health status risk adjustment formula. However, this measurement was made using fee-for-service data; the impact on managed care plans may be significantly different.

HER has proposed an alternative method to excluding one-day hospitalizations which assigns

varying weights to stays that are one day long versus two or more days. While this alternative does at least partially address the gaming issues, it may cause other problems. First, crediting various values to one-day hospital stays could add to the complexity of administering and understanding the system. Second, if variable DCG weights were constructed for one-day versus longer stays using fee-for-service data, these weights might not reflect the higher intensity and cost of one-day stays in a managed care plan. Third, if the same credibility and robustness requirements were applied to these hospital stays (i.e., DxGroups must contain at least 50 people), this calculation could result in more diagnoses falling into the "base" category due to insufficient people in the category and clinical judgment would be required to "reclassify" those DxGroups.

The underlying concept of excluding one-day admissions does have merit. It can reduce gaming of the system by requiring each hospitalization to be of a certain severity (measured by a length of two days or more) and plans would not have an incentive to hospitalize a patient overnight just to receive "credit," thus, the majority of PIP-DCG diagnoses are severe enough to have an average length of stay in excess of this two-day minimum. However, there are several disadvantages that should be considered.

First, according to the HER report, excluding one-day stays reduces the predictive power of the health status risk adjustment methodology.[9] If data from Medicare+Choice organizations is used, there may be more than 5% of otherwise qualifying conditions excluded from the formula. Second, this exclusion may penalize plans which more efficiently manage care, since data generally indicates that these plans have a lower average length of stay. This indicates that a plan which manages care may be paid less than a plan which does not manage care for exactly the same type of patients. Finally, it should be considered if excluding one-day hospitalizations shifts the issue of "gaming" from whether to hospitalize someone at all to a question of whether to keep the patient for an extra day. It would be appropriate to analyze the risk adjustment methodology based on whether it is easier to "game" admissions or to "game" length of stay and any resulting adverse incentives for health plans.

The Work Group suspects that the disadvantages of excluding one-day hospitalizations may outweigh any possible gain. One of HCFA's goals in designing the risk assessment methodology should be to negate bias against the prudent management of health care costs by Medicare+Choice plans. HCFA may want to consider either using one-day stays as part of the risk adjustment formula or giving a partial credit or other adjustments for those hospitalizations in structuring payments to health plans.

Chemotherapy

HCFA has indicated that beneficiaries who are undergoing chemotherapy will be placed in a diagnosis category based on the patient's secondary diagnosis (most likely cancer). Since the

[9] See Chapter 4 of the draft HER report dated July 17, 1998.

medical conditions underlying the need for chemotherapy represent high-cost, ongoing conditions that are predictive of future medical expenses, it is appropriate that they be included in the risk assessment model. The Work Group believes including chemotherapy as part of the diagnosis groups will increase the ability of the methodology to predict future health care costs.

Demographic Factors

The health status risk adjustment methodology includes a number of demographic factors that will be used to measure the baseline predicted cost for each person. Medicare+Choice health plan members with PIP-DCG conditions will be assigned the extra cost of the diagnosis in addition to their underlying demographic costs. In general, it is not possible to "game" demographic factors such as age and gender. A brief discussion of the other demographic factors follows:

- **One Rate Book for Aged and Disabled.** While not a demographic factor per se, it is an important feature of this system because it creates a unified and self-contained methodology that includes all Medicare members (except for beneficiaries with End Stage Renal Disease, which is handled through a separate payment system).

- **Ever Disabled.** Having this add-on factor for Aged members (i.e., beneficiaries who qualify for Medicare because of their age) who were previously covered by Medicare due to disability maintains the internal consistency of the model and appears to appropriately measure the additional cost of these Aged members.

- **Medicaid.** HER indicated in its report that Medicare beneficiaries who were eligible for Medicaid one month or more during the 12-month data collection period typically had higher medical expenses in the future. Medicaid status information is generally available and does not require additional plan reporting so this coverage may be an appropriate factor to add to this model. Before implementation of the PIP-DCG system, it would be desirable to sample and verify Medicaid status, particularly in small enrollment counties where this factor could make a significant impact in the risk assessment values and the resulting payments to health plans.

- **Institutional Status.** The HER analysis indicated that available Medicare data on institutional status of beneficiaries includes two different groups, those individuals in skilled nursing facilities, which have high current medical costs, and individuals in other types of sub-acute long term nursing home care, which generally have lower current costs. However, the HER report indicated HCFA does not collect this information on a routine basis and thus it is difficult to accurately distinguish between the two types of care using currently available Medicare data. Since institutional status information may not be uniformly collected by health plans and is subject to potential gaming, it is appropriate that it not be included as a demographic factor.

There is a concern however, that certain institutional demonstration projects (such as Programs of All-Inclusive Care for the Elderly or "PACE" and Social Health Maintenance Organizations) may have other, more expensive subgroups of institutionalized individuals; individuals who are

significantly underpaid by the exclusion of this status, especially if the programs reduce the number of acute admissions which form the start of an institutional stay. HCFA should consider the development of specific health status categories for these individuals.

- **Working Aged.** At the time the draft HER report was provided to the Academy, there was an indication that HCFA was considering use of a factor for those Medicare beneficiaries who are still employed (Working Aged), since a large portion of the medical costs for those plan members may be paid by their employers. However, the use of this demographic factor must be carefully considered, since the employer who is the primary payer for medical services may not report to HCFA that a Medicare beneficiary is covered through an employer health plan.

Exclusion of Indirect Medical Education Costs

The model developed by HER excludes indirect medical education (IME) costs from the Medicare FFS data used to calculate the relative weights used in this system. The IME costs are approximately two-thirds of the total graduate medical education costs currently paid through Medicare (the FFS data does include direct medical education (DME) expenses). While it is technically incorrect to include any graduate medical education costs (since medical education costs will be paid outside of the capitation rate in the future), any distortion is likely to be small. However, it is possible there will be some internal inconsistencies in the model since high-cost conditions captured in the PIP-DCGs may more likely be treated in a tertiary care or teaching hospital.

Factors for Newly Enrolled Medicare Members

In addition to currently eligible Medicare beneficiaries (either in the FFS program or in health plans), the risk adjustment method will have "neutral" factors for new Medicare members without any diagnostic history. HCFA has decided to develop a special set of risk scores for those individuals who are eligible for Medicare for the first time and do not have any prior encounter data in the Medicare system.

HCFA has used FFS data to construct average expenditures for categories of newly eligible members (beneficiaries who become eligible for Medicare because of age or disability or members who were previously eligible for coverage but deferred entry into the Medicare system). Newly eligible members will be assigned an estimated risk score based on HCFA's estimate of their predicted medical expenditures. The validity of these risk scores is unclear. We therefore suggest HCFA review its risk scores for the newly eligible once current data is available.

Application

In developing risk assessment scores for each person, HCFA intends to examine all fee-for-service and encounter data to produce each person's score. It is important to combine data from all sources to account for movement between plans and between the fee-for-service and managed

care systems. The impact of the new system on individual Medicare+Choice contractors is unclear. HCFA has not completed substantial testing of individual contractors.

B. Predictive Power of the Model

According to HER, approximately 87% of Medicare+Choice health plan members will receive a score based on demographic factors alone. The other 13% will also be assigned a score based on their PIP-DCG diagnosis. A primary question is the extent to which the proposed health status risk adjustment methodology is useful in predicting future medical expenses and therefore is an appropriate formula for Medicare reimbursement to health plans.

Risk Assessment - A "Work in Progress"

The goal of a risk adjustment method is to match the payment to a health plan or provider with the need for health care services of members. The goal of risk assessment for the Medicare population is to appropriately pay health plans for chronically ill members. Defining the "need" for health care services is a less than exact science. Most risk assessment methods which have been developed to date have tended to share certain inherent assumptions. While these assumptions represent the best available current mechanisms for measuring health care needs, they are also generally recognized as being less than perfect.

One assumption is that the "need" for health care services can be measured by prior use of services. Some observers argue that many of the health care services which are performed today are not medically necessary. However, it must also be recognized that there is no general agreement about which services are appropriate and which services are, in fact, unnecessary. To the extent that certain disease categories have a greater or lesser number of "unnecessary" services than other disease categories, current risk assessment methods may overstate or understate the true "need" for services.

Another premise is that people with similar physical conditions will need similar amounts of medical services. The difficulty lies in defining "similar physical conditions." No current administrative method of diagnostic coding captures all of the aspects of illness which relate to a patient's "need" for services. Even if such a method were devised, it is not currently administratively feasible to collect all of the specific details in order to completely define any given patient's medical needs. Risk assessment methods will always put some patients with varying needs for services into identical risk stratification categories. At the same time, no risk assessment mechanism should be expected to completely replicate prior costs.

Concurrent Versus Prospective Risk Assessment

Another important factor in predictive power is whether the risk assessment method is prospective or concurrent. A concurrent method matches the current year's risk factors with the

current year's health care need. A prospective method uses the current year's risk factors to predict the following year's need. Predicting future costs based on current conditions will always be less accurate than predicting the costs which were incurred during the time frame when the risk assessment factors were assigned.

For a variety of reasons, however, prospective approaches for a risk assessment method are still preferred by most observers. Concurrent methods tend to compensate for the treatment of acute conditions. In addition, concurrent methods compensate for accidental conditions which are otherwise unpredictable and do not require risk adjustment. A prospective risk adjustment method, therefore, should not be rejected simply because it does not predict costs as well as a concurrent method.

HCFA has indicated it will use the prospective system in its risk adjustment methodology. This choice is likely to reduce gaming of the system by placing the emphasis on diagnoses with high ongoing costs, which are typically chronic in nature, thus reducing the emphasis on traumatic, acute conditions which may be self-limiting. The disadvantage of a prospective system is that it has significantly lower predictive power than a concurrent system. The advantage of the prospective methodology is that it provides health plans with an incentive to manage care because they stand to gain if future costs are lower than prospective payments. A prospective risk adjustment methodology also provides health plans with a greater ability to predict future payments which improves the solvency of the system.

Measures of Predictive Power

There are several measures of predictive power in general use today, including the R-Squared statistic and the Predictive Ratio. The R-Squared statistic measures the variance between the predicted use of services and the actual use of services on an individual by individual basis and compares the result to the variance for the entire population. The resulting score is expressed as a percentage of variance and the highest (i.e., most predictive) score is 100%. The best prospective risk assessment methodologies currently range in the area of 10% for individual Medicare beneficiaries. Using demographic information alone (similar to the current Medicare AAPCC model) will usually produce an R-Squared score of about 1%. A number of researchers estimate the best possible individual score is around 20% for risk assessment systems.

The Individual R-Squared value is less useful to health plans since the nature of insurance is to spread risks of random fluctuations over larger populations. Identifying actual values for each participant is less important.

Another measure of predictive power is the "predictive ratio." This indicator measures the ratio of the actual use of services to the predicted use of services for a group of individuals. Predictive ratio scores may range from 0 to many multiples of 1. Typically, risk assessment methods for large random groups will typically generate predictive ratios very near or at 1.00. On the other hand, application of risk assessment to non-random groups consisting of all individuals with

certain diagnosis categories such as diabetes or asthma can sometimes achieve scores between .85 and 1.15 if better risk adjustment methods are applied.[10] In general, predictive ratios are more meaningful than R-Squared scores to health plans, particularly if a health plan contracts with specific providers with expertise in certain high cost conditions and therefore attracts more than its share of individuals with these conditions.

Risk Adjustment - Not a Remedy for Inefficiency

If a risk adjustment method has a high degree of predictive power, it will better allocate available funds according to the underlying need. This does not mean, however, that all health plans will have similar financial results. Even if the predictive power of a risk assessment scheme is very accurate, more efficient health plans will either be more profitable or will be able to charge lower premiums than less efficient health plans.

How Much Predictive Power is Enough

One test of a risk assessment mechanism is whether a majority of health plans receive sufficient income to cover their costs on an on-going basis. A successful method will encourage health plans to continue contracting to provide Medicare coverage and to expand their marketing to higher risk beneficiaries.

For Medicare beneficiaries, a successful method will minimize disruption in the market and possibly increase the number of health plan choices available. For policy makers and financing organizations, a successful method will provide an indicator that funding is appropriate (i.e., minimizing excess profits/surplus), that enrollment is expanding, and that beneficiaries with higher risk status are not avoided.

C. Implications of Indefinite Use of the PIP-DCG System

The proposed PIP/DCG system is expected to be in place only for a few years at most and then replaced with an enhanced system. The Work Group believes that there are significant negative implications if the proposed PIP-DCG system is used more than a few years. One of these is that health plans which use outpatient alternatives to hospitalization would be financially penalized by a risk adjustment system that uses a formula based only on inpatient diagnostic data. Potentially, this limitation could penalize the more efficient plans enough to make them leave the Medicare market. In addition, it might create an incentive for plans to promote less efficient care modalities in order to increase Medicare payments.

In addition, the proposed system does not increase risk factors when an individual falls into two or more diagnostic groups; rather, the individual is scored only at the highest severity group. This

[10] See: "Risk Adjustment For The Medicare Program," Greenwald, et al.

lack of increased risk adjustment for combined conditions may unduly reduce the effectiveness of the risk adjustment system, as health plans with the most severe cases of a given DxGroup will have a greater incidence of multiple conditions.

This restriction could underpay health plans which include providers that typically treat more patients with multiple chronic conditions, provide incentives for those plans to drop those providers from their provider panels, or provide incentives for health plans to market themselves to only the most healthy potential enrollees. On the other hand, any recognition of the higher cost of these chronically ill members is an improvement over the current demographic-only method of capitation rate development.

V. Consideration of Comprehensive Data Models

The currently proposed risk adjustment methodology only uses diagnostic data from inpatient hospitalizations. HCFA has indicated that one future enhancement to the system will be to include comprehensive data, starting with ambulatory diagnostic data.

There are a number of advantages to including ambulatory data:

- Using ambulatory data can capture high risk situations even where hospitalization did not occur. Some high cost treatments, such as chemotherapy, can be performed on an outpatient basis. Even though such patients require a high level of medical services, they are not recognized as high risk by an inpatient-only system. In addition, efficient health plans which have implemented disease management processes to avoid the need for hospitalization would be treated more appropriately under a system which includes ambulatory care.

- Some risk assessment systems using ambulatory data have significantly better ability to predict risk levels. Higher predictive power means more equitable payments to health plans.

- Even if not used for risk adjustment purposes, it would be helpful for health plans to capture this data to measure the efficiency and quality of health care delivery.

- Using only inpatient diagnoses produces a financial advantage to health plans to admit patients even if hospitalization is otherwise questionable. Public policy probably argues that health plans should not be rewarded for choosing more expensive treatment than is medically necessary.

On the other hand, there are several disadvantages:

- Complete ambulatory data is not currently captured by many health plans via an electronic mechanism. Requiring ambulatory data will require the expenditure of

significant capital expenses for certain plans. This requirement will add to the overall cost of the system.

- Including ambulatory encounters will significantly increase the amount of data required from health plans. This situation increases the opportunity for error. Auditing an inpatient stay requires a relatively small expenditure relative to the cost of hospitalization. Auditing an ambulatory encounter requires an expenditure which represents a significantly larger percentage of the cost of the services provided.

- Many diagnoses are originally coded by physicians more as possibilities rather than conclusions. Hospital diagnoses are the diagnosis as of discharge, which is much more likely to reflect the best conclusion of the medical staff. Ambulatory diagnoses do not distinguish between admitting and discharge diagnoses and are, therefore, somewhat less likely to be accurate.

- Ambulatory data will need to be captured from a wide variety of sources and will likely require significantly more time to accumulate. A small fraction of the population is hospitalized in any given year, but the majority of people have some contact with the medical community. The increase in the input data to the risk assessment system would be very significant.

- There is more likelihood of inconsistent coding, because while some physician contracts may include incentives for better coding, other capitation contracts may reduce physician incentives to code thoroughly.

Including ambulatory diagnoses in the risk assessment system should increase its predictive power and therefore the equity of payments. It will reduce bias toward compensating health plans for more intensive treatment than is strictly necessary. On the other hand, including ambulatory diagnoses will increase the cost of any risk assessment system, will decrease the timeliness of the system, and may provide more opportunity for incorrect or inconsistent coding.

In general, the advantages of including ambulatory diagnoses seem to justify their use. A more thorough cost-benefit analysis should be performed. Currently, the Work Group is unaware of any other means of risk assessment which would meet the stated goals of a risk assessment system in a more efficient manner than a proper use of ambulatory diagnostic data.

VI. Implementation Issues

One of the keys to any risk adjustment methodology is the ease or difficulty with which it is implemented by HCFA and the resulting impact on Medicare+Choice health plans and their members. The recent decision by a number of Medicare+Choice organizations to withdraw from the Medicare market underscores the need for developing a system which fits within the operational needs of managed care plans.

A. Timing of Payments

HCFA has indicated that data are likely to be collected from July 1, 1998 through June 30, 1999 for application to capitation rates effective January 1, 2000. This schedule, combined with prospective risk adjustment regression factors, will result in capitation rates that lag behind the theoretical prediction period. For example, for a person hospitalized on July 1, 1998, the regression factors predict that person's cost for the year beginning July 1, 1999. However, the capitation rates using these regression factors will not be paid until January 1, 2000. This lag between the data reporting and application, while lengthy, is significantly shorter than the lag in other systems currently used, such as the Health Insurance Plan of California or the planned implementation by the Washington State Health Care Authority. Other systems, such as Business Health Care Action Group in Minneapolis, use more frequent quarterly updates.

Another timing issue concerns the determination of year 2000 benefits. The proposed timing of risk adjustment causes some serious problems for Medicare+Choice health plans. By March 1, 1999, these carriers will only know the preliminary estimated risk scores, calculated on a market wide average basis. However, the health plans must also commit to the year 2000 premium rates and benefits by May 1, 1999 when they file the Adjusted Community Rate (ACR) proposal. Health plans can not make such commitments with any degree of certainty while their income in the year 2000 is so uncertain. As a result, carriers may take various measures to limit this risk, including reducing benefits or even leaving the program altogether.

The Work Group suggests that HCFA consider addressing this timing by giving each health plan their own member-level risk scores by March 1, 1999 or earlier, or by allowing health plans to submit rate and benefit revisions after May 1, 1999. HCFA may also consider limiting the amount by which the risk adjustment factor will be used to calculate the reimbursements.

The Work Group recognizes that some of the timing issues are based on legislative requirements set out in the BBA. To the extent that any of the changes suggested by the Work Group in this report require legislative action, HCFA may want to work with Congress to modify the existing law.

B. Recalibration of the PIP-DCG Model

As discussed, HCFA intends to start using additional data in its risk adjustment methodology within the next few years. Therefore, risk adjustments are likely to be based upon updated and more comprehensive data in the near future. However, the proposed time frame for establishing a more comprehensive risk adjustment system may be aggressive. If the PIP-DCG system is used going forward, re-measurement and recalibration of the current PIP-DCG weights should be considered.

The Balanced Budget Act of 1997 enacted several changes in provider payments. These changes

are effective at different points in time, but all are effective starting after 1997. Therefore, the relative values of the diagnostic cost groupings based upon future claims will most likely differ from measurements based upon 1995 and 1996 Medicare payment rates. For example, on the fee-for service side, BBA increases payments to primary care physicians and decreases payments to surgeons. Payments to hospitals are being reduced: fiscal year 1998 hospital reimbursements are frozen at 1998 levels and capital and disproportionate share reimbursements are being reduced. Capitation payments to Medicare+Choice plans will exclude graduate medical education costs in the future; however, the direct medical education expense component of these costs are included in the claims data used to develop the risk assessment model. As previously discussed, any distortion that results from including DME may be small.

C. Examination of the Re-Scaling Factor

It is the Work Group's understanding that one of the key factors in the implementation of the PIP-DCG model is the re-scaling factor. This factor is necessary to assure that average payments for the FFS population in a county remain the same, whether the current AAPCC methods or the new PIP-DCG model is used. In other words, risk scores of health plans must be synchronized with a new measure of the average risk of Fee For Service Medicare beneficiaries in a county.

Although the re-scaling factor is not a direct part of the PIP-DCG method (re-scaling would be required with any new method), the Academy cannot determine the effects of the PIP-DCG method on health plans without understanding its implications - which means understanding how the re-scaling factor works. It is likely that health plans will feel it important that the details be disclosed prior to final implementation of the new risk adjustment methodology, so that they can fully assess the impact of the new risk adjustment system.

The credibility of the factors used in small enrollment counties and in counties with relatively few remaining fee-for-service enrollees should be considered. Since this factor is critical in the operation of any formula, it is important to understand exactly what the effect is in small counties, how sensitive the factor is to changes in base years, and what techniques HCFA will use to increase credibility or to minimize the effects of statistical fluctuations. For example, the number of Medicaid recipients in a small county may vary dramatically from year to year, especially with changes in the economy.

HCFA provided the Work Group with a general outline of how the rescaling process will work. We understand that HCFA plans to compute average risk scores for each county using three years of FFS data (1994, 1995 and 1996). However, a single year (1996) will be used in all but the smallest counties to create a restandardized rate book for determining the risk adjusted payments beginning in 2000.

The Work Group cannot adequately estimate the impact of the health status risk adjustment formula on health plans without more detailed information about the method, the calculation and the data underlying the calculations. The Work Group was unable to fully analyze the

mathematical formula used to produce the rate book or how the formula would operate with the data collected from health plans. The Work Group suggests that HCFA continue to review the rescaling process and determine its overall impact on the risk adjustment payment system.

D. Phase-In of the Risk Adjustment Methodology

HCFA has indicated it will phase-in the new risk adjustment payment system rather than make the changes all at once. It is anticipated the new payment amounts based on the health status risk scores will be blended in some fashion with payments calculated under the current system. The shift to the new methodology will be made in incremental steps over the next few years.

The Work Group believes this decision to phase-in the risk adjustment methodology will help avoid significant market disruption that might otherwise occur. The potential impact on health plans from the risk adjustment formula could be significant and both HCFA and the plans need time to fully understand the changes. This phase-in approach will also provide HCFA with the opportunity to adequately test the accuracy of data collected from plans and to verify the underlying assumptions used to develop the risk adjustment formula.

VII. Recommendations

A. Sensitivity Testing

Health Economics Research performed a number of tests on the PIP-DCG risk adjuster methodology to determine how well it predicts total expected medical costs. The recommendations made by HER regarding several key components of the model such as the use of inpatient data only, exclusion of one-day stays and the number of PIP-DCG groups to be used, appear to be reasonable based on the FFS data which was reviewed. While Health Economics Research has discussed potential bias against managed care organizations that deliver care more efficiently than fee for service providers, HER did not have managed care data to determine what, if any, bias exists.

HCFA has completed some preliminary testing of the potential impact of the new risk adjustment methodology on Medicare+Choice plans, including managed care organizations. The Work Group believes HCFA should update these tests as additional data is available from plans and the agency and health plans gain more experience with the operation of the risk adjustment mechanism.

The Work Group recommends that HCFA consider the following testing protocols to allow a more thorough analysis of its health status risk adjustor methodology:

<u>Tests Using Managed Care Data</u>

1. Continue to test the impact of the risk adjustment methodology on managed care organization revenue. Managed care organizations could experience significant decreases in revenue due to the implementation of the risk adjustment methodology. The results of the tests should be compared among managed care plans with several different characteristics:

a. Medicare enrollment (large, medium or small).

b. Total enrollment (large, medium or small).

c. Urban versus rural service areas.

d. Level of county payment rates (low, medium or high).

2. Test the impact of coding practices.

a. Start with a given set of inpatient charts.

b. Code once, within legally allowable parameters, with the goal of producing the most revenue under the risk adjustment methodology.

c. Code again, within legally allowable parameters, with the goal of producing least revenue under the risk adjustment methodology.

d. Compare the results from (b) and (c). Significant differences indicate that the method is subject to gaming.

3. HCFA's data testing needs to address the scenario in which a member enrolls in a new Medicare+Choice plan and the data from the enrollee's previous Medicare+Choice plan was inadequate.

4. Test the sensitivity of re-scaling to choice of underlying data.

a. For a sample of counties, perform the re-scaling process calculations with one and more than one year of data.

b. Test the sensitivity of re-scaling factors to the quality of Medicaid and institutionalized status fields.

5. Test the sensitivity of working aged adjustments. As the actual development and methodology for the working aged adjustments have not yet been finalized by HCFA, suggestions here are premature.

<u>Tests on FFS Data</u>

1. Test the calculation of PIP scores for small counties using one year and multiple years of FFS data.

2. Review the calculation of PIP scores across similar small counties.

3. Test PIP scores on counties with very large HMO market penetration.

Tests on Risk Assessment Formulas

1. Test variations due to differing combinations of statuses, when combining together the aged and disabled factors.

2. Test variations in the FFS demographic scores using multiple years of institutionalized or Medicaid status information by county.

B. Cost-Benefit Analysis

Consideration should be given to producing a cost-benefit analysis of the PIP-DCG methodology and any subsequent modifications. The proposed system is relatively new and it is likely that there will be difficulties in implementation. It would be very helpful to establish more accurate estimates of the cost of implementing the PIP-DCG methodology and any modifications (such as using ambulatory data) and to determine the benefits to be derived from these systems before final decisions as to implementation are made. The analysis should specifically include the costs incurred by health plans due to changes to the system. To the extent there are significant additional expenses placed on plans, they may choose to drop out of the Medicare system rather than enter into new contracts or continue existing contracts using the new risk adjustment methodology.

C. Actuarial Oversight

HCFA apparently plans to conduct additional analysis of the impact of the PIP-DCG methodology on managed care plans. It is unclear what form that impact analysis will take. In addition, there is a need for continuing monitoring and testing of the system and future modifications. The Academy suggests that additional actuarial review be included as the system and subsequent changes are implemented.

D. Improving The Current Structure

Even if the use of ambulatory diagnoses is not considered as a future enhancement, the current system is cumbersome and could be improved in terms of simplicity, accuracy and predictability.

Simplicity - The current system uses rate books which are based on calculations that have been subject to various minimums, both in absolute amounts and in relative growth, and the rate book system is both cumbersome and inaccurate. One possible change to simplify the system would be to calculate directly the rate book amount for each geographic location to a value which represents the local cost for an individual with a demographic factor of 1 and a health risk assessment factor of 1.

Accuracy - The current system intends to use the 1997 local rate book values as the basis for future years' amounts. Increases to the factors are based on national increases and predetermined legislative minimums. One possible change to improve the accuracy would be to recalculate updates on an annual basis, using local data regarding cost, demographics and risk.

Predictability - The current system recalculates values annually using one full year's worth of data. One possible change to improve predictability would be to calculate amounts quarterly on either a year-to-date or a twelve-months-rolling basis, which would allow health plans to predict what future reimbursement levels will eventually result.

E. Timing of HCFA's Initial Testing and Analysis

The Work Group believes that testing should be done to assess the potential impact of risk adjustment in the market place. These concerns relate both to the potential changes created in the marketplace and to the delivery of care to beneficiaries participating in the Medicare+Choice program.

The introduction of the proposed PIP-DCG risk adjustment mechanism into the Medicare program creates increased uncertainty and risk to Medicare+Choice plans. Testing and analysis is the only way to reduce this risk and uncertainty.

HCFA has completed significant testing based on fee-for-service data. The Work Group believes that further testing must be done if the uncertainty to health plans is to be alleviated. The testing should include two components. First is the sensitivity of the formula to changes in key assumptions. Second is determining the impact of the proposal on managed care organizations. Changes in revenue should be examined to ascertain if large dislocations will occur, which may result in unfortunate consequences to beneficiaries. A reasonable test of the proposal is whether or not a majority of health plans receive sufficient income to cover costs on an ongoing basis, without reducing benefits dramatically

Another concern is the quality and timeliness of hospital encounter data submitted by health plans, which might also be part of the testing process. The Work Group urges HCFA to complete its planned data testing, and to include a comparison of the number of submitted claims with an external standard for the expected number. Risk adjustment will be inaccurate if claims are either held back or denied due to edits, or get caught in a data processing bottleneck.

In order to understand the value of detailed testing when a risk adjustment mechanism is significantly changed, it would be instructive to select recent historical examples of significant change to the risk adjustment mechanisms in current use. The following examples come to mind:

- The use of risk adjustment in the Health Insurance Plan of California (HIPC); and

- The use of risk adjustment in managed Medicaid programs (e.g., Maryland and Oregon).

In planning for a significant change to a current risk adjustment mechanism, actuarial soundness could be analysed in a prospective manner through sufficient testing or in a retrospective manner through an actuarial oversight function (such as through a committee or an independent evaluation).

In the example of the Health Insurance Plan of California (HIPC), a one year simulation was conducted prior to the actual start of the program. The HIPC is an inpatient only model, similar to the initial proposed HCFA (PIP-DCG) approach.

In the Oregon Medicaid program, two methods were tested; the Disability Payment System and the HIPC inpatient only model. Score variations were reviewed and a method of credibility was introduced through the use of corridors on final risk adjustment scores.

In the Maryland Medicaid program, carve-outs for behavioral health, maternity, AIDS and "rare and expensive conditions" are used. Stop-loss insurance is also provided. In addition, an article describing the process in the October 1998 *Journal of Ambulatory Care Management* noted that, "it will also be important to support adequate quantitative and qualitative evaluation including assessments of the actual accuracy of the payment system and its impact on patients and providers."

In each of the examples mentioned above, there was either extensive detailed initial testing or a follow-on actuarial oversight function (actual or implied). In several other states, implementation of a risk adjustment system without testing led to problems which required changes in the methodology.

VIII. Conclusion

A. Comparison of the PIP-DCG Method With Other Models

It is the Academy's understanding of the BBA provision that HCFA is mandated to use some form of health status based payment method starting in the year 2000. HCFA's current intention to implement the PIP-DCG model leads naturally to questions regarding the choice of the PIP-DCG model versus other available payment models.

The following are the Work Group's comments about various other types of risk adjustment

models in the context of HCFA's choice of the PIP-DCG method. We recognize that there is a substantial body of research regarding various risk adjuster models and that we are providing only an actuarial perspective.

What types of models satisfy the BBA requirements?

Based on our experience, there appear to be at least three main classes of models that meet the BBA requirement for health status-based payment. These are:

- Diagnostic-based models with data gained from administrative sources (typically from claim records or encounter data files);

- Diagnostic information from clinical records, such as medical charts; and

- Survey-based health status information.

A great deal of research has been completed recently in the diagnostic/administrative class of models. PIP-DCGs fall into this category, as well as the other versions of DCGs, Ambulatory Care Groups, the Disability Payment System, the Global Risk Adjustment Method and several others. It is our understanding that any of these methods would likely satisfy the BBA requirement and most of the models have approximately the same degree of predictive power on a prospective basis.

All of these data-intensive methods assume that reliable data from health plans will be available at little marginal cost. While the diagnostic data needed to produce risk assessment results will have many uses, gathering and transmitting that data in a fully operational implementation will likely prove to be challenging. Part of the challenge is just the management, transmission and audit of the data. A related issue is the need for some health plans to revise significantly their contracts, particularly if they are globally capitating providers for both professional and institutional services, without obtaining the requisite data under current contract terms.

A few researchers have explored the use of more detailed clinical risk adjustment models in a limited setting. While most clinicians would be likely to offer more support for this class of methods, the data-gathering cost is prohibitive. At this point in time, it is our experience as actuaries that clinical information from sources such as medical records and patient charts is nearly impossible to gather, except in the most manpower-intensive manner through actual chart audits, etc. As a result, models in this category are of great theoretical interest but of little practical help for the Medicare+Choice program.

Survey-based models have also been extensively researched. Many researchers report that surveys with sufficient response rates provide predictive power that is in the same range as the diagnostic/administrative class. The experience with self-reported surveys (e.g., the RAND SF-36) indicate that the health status information gathered by surveys is dependable and provides a good base for predicting future costs. There are, however, major drawbacks to use of surveys,

particularly with the elderly. These drawbacks include:

- High cost of administering the survey;
- Possible low response rates;
- Likely difficulties that the elderly would have completing the survey by themselves;
- Possibility of gaming of survey results when payment is dependent on survey answers; and
- Concerns about privacy of health information.

Because it appears that HCFA wants to obtain health-related data on every Medicare-eligible person, surveys don't appear practical, since response rates will never be 100% (except in special cases, like the SHMO and PACE demonstrations). In addition, the cost of obtaining the data is known to be significant.

As a result, the only practical class of health-based risk adjusters, at this time, would appear to be the diagnostic/administrative models.

What models are appropriate alternatives to the PIP-DCG model?

Many observers recognize that using only inpatient data in the PIP-DCG risk adjustment model may result in a bias toward the FFS system. The potential problems with using inpatient data and limiting DCGs to hospital stays over one day have been discussed earlier in this report. As discussed, HCFA's PIP-DCG model uses only inpatient data. A number of the other risk adjustment systems that the Work Group is familiar with use either both ambulatory and inpatient data or ambulatory data only. We also note that new methods are constantly emerging, such as research into the use of prescription drug data.

From various research (including that funded by HCFA), "comprehensive data models" (i.e., those that use inpatient and ambulatory data) have superior predictive power to inpatient-only methods. The Hierarchical Condition Categories model as well as various ambulatory cost group models perform at an R-squared value of around 0.09 versus the approximately 0.06 performance of the PIP-DCG model.[11]

If feasible, a model which uses ambulatory and inpatient data is preferable. However, there are very significant current barriers to implementation of a full data model within Medicare, including:

- HCFA initially collected inpatient data and does not plan to collect ambulatory data from plans until at least October 1, 1999;

[11] See: "Risk Adjustment For The Medicare Program," Greenwald, et al., pp. 10-12.

- As discussed earlier, many Medicare health plans do not capture ambulatory data for a variety of reasons. Most plans delegate data collection to their medical providers and ambulatory data is either not collected, is available only in very crude form, or is significantly under reported; and

- While there is standardization for the collection of inpatient data by hospitals, there is a great deal of variability in the coding of ambulatory claims. Even if data could be readily gathered from physician organizations, there may be considerable "noise" in the data due to the different coding practices used by medical providers.

As a result of HCFA's actions and the current state of ambulatory data available from physician organizations which contract with health plans, we believe that the PIP-DCG method is the only choice (by default) from the diagnosis/administrative class of models.

Are there other alternative payment models which may not satisfy the BBA provisions but would be an improvement over the current Medicare reimbursement system?

Although the BBA requires a health status based method, other methods which require less intensive methods may provide improvements over the current system. To the extent that HCFA has any experience or research involving other possible alternatives, we suggest HCFA discuss its findings with the appropriate legislative representatives and staff.

For example, it may be possible to create an "enhanced demographic model" which uses more readily available information, rather than health-status-based data. Enhancements could include a payment adjustment for new entrants into health plans (although it should be based on actual health plan experience, not on prior Medicare FFS claims) or modifying payment for health plan members who die in a payment year. Again, there are other issues that arise with these suggestions (such as possible perverse incentives which may appear to pay more to health plans which have more deaths). However, we believe that HCFA should consider all alternatives to improve the current AAPCC system, rather than be limited to only a narrow class of payment models.

Is there sufficient improvement in payment through the PIP-DCG model (or other health status based model) to justify the change from the current system?

Based on the preliminary report from the Health Economics Research firm, we understand the rationale which is the foundation of the PIP-DCG model. We have requested information from HCFA which would illustrate the changes in payment which would occur in a variety of circumstances. We were told any testing could not be completed until data from health plans was collected and processed, which has not yet occurred.

While this response is understandable, it fails to recognize that the health plans themselves will be required to submit Adjusted Community Rate "bids" to HCFA by May 1, 1999, shortly after the projected March 1, 1999 date when they receive information. Two months is not sufficient time reasonably to analyze such data and prepare corresponding bids. We therefore suggest that

HCFA consider speeding up the testing and disclosure process, or that it delay implementation until the results are known to HCFA itself and other stakeholders (health plans, policy makers, beneficiaries and taxpayers).

At this point, the Academy cannot comment fully on the effectiveness of the PIP-DCG payment mechanism. There appears to be potential for the perception of significant problems by health plans, which can only be remedied by further information.

B. How Well Does the PIP-DCG Method Satisfy Risk Adjustment Goals?

The goals of risk adjustment are identified by the American Academy of Actuaries' Health Risk Adjustment Monograph and were previously outlined in Section III of this report. In addition, risk adjustment systems ought to be easily administered and should not provide perverse incentives to health plans or providers. The PIP-DCG methodology developed by HCFA can be analyzed in relation to those criteria as follows:

Reducing the Effects of Risk Selection

Risk selection is a process used to enroll larger numbers of relatively low cost individuals and fewer relatively high cost individuals. High cost individuals are avoided because their cost may exceed premiums, while healthy individuals are sought because their costs will be less than the average.

Without risk adjusters, health plan revenue is based on average costs. To reduce the effects of risk selection, the health risk assessment mechanism must estimate the cost of the individual and adjust the health plan's compensation to at least partially reflect the expected difference in cost due to health status. If the additional compensation for high cost individuals is not high enough, a plan will lose money on the high cost individual and will still have an incentive to avoid covering him or her. If additional compensation is too high, the health plan will be overpaid for the risk and may seek out such individuals for whom payment exceeds future costs.

The PIP-DCG mechanism uses hospitalizations in a year to predict an individual's cost in the following year. Thus, incentives are created for health plans to identify individuals for which they will receive PIP-DCG based payments lower than expected costs and, in turn, avoid enrolling them. In particular, the proposed PIP-DCG risk adjustment mechanism will create incentives for health plans to avoid enrolling Medicare beneficiaries with:

- Chronic medical conditions, but no inpatient admissions in the previous year that would result in increased payments under the PIP-DCG mechanism;

- High cost admissions in the previous year that do not trigger increased payments under the PIP-DCG mechanism, but where significant follow-up health care costs are expected; or

- Medical conditions which are more likely than average to cause high costs from end-of-life hospitalizations, but no increase in payments under the PIP-DCG mechanism should the Medicare beneficiary die before the following year (this is an unavoidable consequence of a prospective system based on inpatient data).

Despite these concerns, we note that incentives exist in the present reimbursement mechanism to avoid enrolling any unhealthy Medicare beneficiaries (whose expected costs are greater than payments). While the PIP-DCG methodology will include some undesirable incentives for health plans, it does reduce some of the incentives that exist under the current payment mechanism.

The use of demographic-only factors for adjustment of new participants may allow for some incentive to select. However, since these individuals are typically healthier and since health plans know that a member may stay with them for life, the one year lack of adjustment is likely a relatively minor incentive.

Adjustments based on last year's hospital diagnoses may result in some mismatch between health care cost and health plan compensation. The additional premium received through a prospective risk adjustment system is not for the high cost of the initial hospitalization, but rather is for the anticipated higher cost in the following year. In non-chronic cases or in cases where most of the cost is associated with the initial high-cost hospitalization, the future premium will not be adjusted for most of the high cost. In end-of-life hospitalizations, no insurer will receive the increased premium in the subsequent year. These mismatches may result in some continued incentive for selection.

Compensating Health Plans Fairly and Equitably for the Risks That They Assume

Fair and equitable compensation implies that the actual average health care cost is within a predictable range of the anticipated average health care cost. If a plan experiences average health care costs within this range for each age-gender, institutional, ever-disabled, Medicaid risk category, it is fairly compensated. If a higher or lower than average number of high-cost individuals are enrolled (due to risk maldistribution), the plan will not be fairly compensated unless the PIP-DCG adjustment accurately adjusts for the discrepancy. The Work Group has specific concerns about the treatment of short hospital stays, the accuracy of the Medicaid status indicator and the accuracy of the working aged status information.

If a health plan enrolls an unusually high number of individuals in a category with high cost hospitalizations, it may not be fairly compensated for the additional risk, since there is no additional premium adjustment in the year of hospitalization. However, the plan will be compensated in the year after hospitalization for projected ongoing costs.

For an individual the PIP-DCG accounts for 6% of the following year cost variability on an individual basis, which may not be sufficient to compensate for the additional cost due to potential selection abnormalities. For non-random groups of health plan enrollees, however, the PIP-DCG

method represents a significant improvement to age and gender based payments. Predictive ratios for non-random groups are much closer to 1.0, meaning that the amount of over or underpayment is much reduced.[12]

Maintaining Consumer Choice

One of the goals of the Medicare+Choice program is to increase the choice given to seniors for their Medicare coverage. The health status risk adjusters may encourage health plans to offer more choices, without the fear of being selected against if they offer programs that may attract less healthy individuals. To the extent health status risk adjusters compensated plans fairly, this may result in more choices for seniors.

It should be noted, however, that carrier participation is very sensitive to the level of funding. For example, last year over 40 Medicare+Choice health plans either withdrew entirely from the Medicare market or reduced service. These withdrawals impacted 440,000 beneficiaries who were forced to switch to another Medicare+Choice plan or a traditional FFS Medicare program.[13] A reduction in Medicare+Choice funding might also harm consumer choices by forcing a reduction in benefits. Without HCFA's risk adjustment method being more fully defined and tested, neither HCFA nor the Work Group can predict the impact of risk adjustment on consumer choice.

Protecting the Financial Soundness of the Health Care System

Financial soundness is maintained when costs do not consistently exceed income. Health plan costs primarily consist of health care expenses and administrative expenses. The proposed PIP-DCG methodology does not have a direct impact on administrative efficiency or costs, which in a poorly run health plan can threaten the financial soundness of the organization. Health status risk adjusters will help match the financial risk of a plan's health care expenses with income and thus could improve the financial soundness of the health care system.

Without additional testing, it is difficult to determine the effect of implementing the risk adjustment method on the current Medicare program. Since payment adjustments are not made in the year of initial hospitalization, there will be a lag in increasing plan premiums that could impact a small or financially challenged health plan.

In addition, many observers feel there is a positive impact or "spillover effect" on the health costs incurred by fee-for-service plans in those markets with a high degree of penetration by managed care plans. This effect is believed to result from the improved operational efficiencies of medical

---

[12] See: "Risk Adjustment For The Medicare Program," Greenwald, et al., pp. 10-12.

[13] "Quick-Fix Push May Derail Long-Term Medicare Plan," Wall Street Journal, January 5, 1999.

providers in their managed care contracts which are then transferred to their FFS patients. It is possible that if fewer Medicare+Choice plans are in the market, then there will be less of this "spillover" on Medicare fee-for-service plans.

Administrative Feasibility

Many other programs such as state small group reform programs have had barriers to using a PIP-DCG mechanism because of the lack of data and administrative difficulties. This may be less of a barrier to Medicare due to the central nature of the program that allows for data on individuals to be consolidated, even if beneficiaries change plans. The program's success will depend on the accuracy and completeness of the data provided across all plans. Medicare performs extensive auditing of information on payers that increases the quality of payer data. Medicare risk programs have not had as detailed reporting requirements and will have to provide data to a centralized organization so that each individual can be classified using their hospitalization diagnosis data. The advantage of the PIP-DCG mechanism is that the information needed is part of the standardized hospital record and, therefore, generally available.

Resistance to Gaming Behavior by Insurers and Minimizing Perverse Health Plan or Provider Incentives

When reimbursement is based on claim coding, there is always a chance of "gaming" behavior through "upcoding." When there is discretion concerning the diagnosis used on an admission, the code with the higher reimbursement may be likely to be selected in order to increase the reimbursement. The PIP-DCG methodology would be less susceptible to upcoding because hospital reimbursement is not directly impacted by the risk assessment score derived from the coding and the insurer is usually removed from the actual coding. This statement is less true when the provider essentially "is" the plan, such as in the case of a provider-sponsored organization or a staff-model HMO.

Since future health plan capitation is increased when there is a qualifying hospitalization, there will be an incentive to hospitalize patients rather than use outpatient settings. Once a patient has been hospitalized during a year there is no further incentive to hospitalize or to over-utilize services, since there is no further increase in future premiums.

One-day admissions were not included in the original study. If this component is chosen as a permanent part of the method and plans understand that if they routinely keep patients more than one day for a given diagnosis, this admission will be scored as part of the PIP-DCG payment methodology, there will be an incentive to increase some stays.

C. Meeting the Needs of the Medicare System

As noted in this report, the proposed risk adjustment methodology may tend to penalize health plans which efficiently manage the delivery of health care because of the use of inpatient data and

the design of the DxGroups which make up the formula. As a result, HCFA probably wants to take steps to minimize this bias wherever practical, assuming the bias does not have a public policy basis. The Work Group recommends the implementation of a risk adjustment system based on more comprehensive data as soon as administratively feasible. The Work Group recognizes the critical limits on the current collection of Medicare data and strongly recommends that HCFA test both any new risk adjustment methods and the ability to collect comprehensive data prior to implementation. This testing should include an independent verification of mathematical calculations which make up the risk adjustment formula.

The Work Group also suggests that HCFA periodically audit the quality of data it is using to develop risk scores, and conduct tests to verify the accuracy of the assumptions which underlay the risk adjustment methodology, as the program operates over the next several years. It is important for HCFA to develop a mechanism which allows the agency to continually monitor how well the proposed methodology is meeting the provisions of BBA as well as the goals of a risk adjustment system as discussed in this report.

The Work Group acknowledges the time and effort that HCFA and its contractors have expended in preparing to implement health status-based risk adjustment for the Medicare+Choice program. The PIP-DCG risk adjustment method is a practical interim step towards better payment to Medicare+Choice contractors and is an improvement over the current demographic factor-based payment method (i.e., the AAPCC methodology), if implemented cautiously.

In addition, the PIP-DCG method appears to represent the only practical health status risk adjustment alternative available for implementation on January 1, 2000 as required by the Balanced Budget Act. The Work Group believes that more sophisticated payments methods are a step in the right direction to successful implementation of Medicare+Choice and its goal of better managing Medicare expenditures and providing choice for Medicare beneficiaries.