

National Centre for Classification in
Health



PROFESSIONAL RELATIVITIES STUDY

RESOURCE MATERIAL C

Literature Review

*Review of the literature on relative values carried out by
NCCH in 1998 to inform the PRS*

prepared for

Medicare Schedule Review Board
December 2000

Literature Review

Context

Escalating health care costs have prompted governments, insurers and medical professionals¹ worldwide to search for ways to reform the traditional fee-for-service payment (Harris, Richardson, Farish and Saltman, 1996; Hsiao, Braun, Kelly and Becker, 1988).

In Australia, there has been mounting pressure over the last decade for a fundamental reassessment of the Medicare Benefits Schedule (MBS) from the medical profession and patients, and from Government initiated enquiries. To achieve this reassessment, the former Government and the Australian Medical Association² (AMA) announced a joint review of the MBS. The Medicare Schedule Review Board (MSRB) endorsed the four main objectives of the review: 1) to introduce equity into the MBS; 2) to provide a common structure across professional groups as far as possible; 3) to promote and appropriately reward good clinical practice; and 4) to remove perverse incentives which discourage good clinical practice. (Department of Health and Family Services, 1998).

Medical Fee Setting

It appears that the problems experienced with traditional fee structures both here and overseas are remarkably similar. The "customary, prevailing and reasonable" system for setting fees prevalent in the USA has been criticized as inflationary, complex and unpredictable and is seen as responsible not only for distorting relative charges but influencing issues such as access to medical service, practice location, specialty choice and quality of care. Moreover, such systems are difficult to administer and understand, and provide financial incentives for overutilisation (Blumenthal and Epstein, 1992; Hsiao and Becker, 1989; Lee and Ginsberg, 1988; Revicki, Orkin, Luce, McMenamin and Weschler, 1990; Roe, 1985; Stason, 1987).

Alternative fee setting mechanisms have been implemented in an effort to constrain the burgeoning cost of health provision. Capitation, voucher systems, DRG based funding, full salaried remuneration, negotiated fee and negotiated access mechanisms have been implemented (as primary mechanisms or in combination with fee for service) in Canada, France, West Germany, Switzerland and the USA (Egdahl and Hertenstein, 1987; Penington, Cashman and

¹ A medical professional is an inclusive term describing doctors, surgeons and other clinicians. USA studies use the term "physician" to describe all medical professional.

² AMA refers to both the Australian and American Medical Associations. Within this document AMA will be used when referring to the Australian Medical Association and the American Medical Association will be referred to in full.

Kearney, 1984; Rakich and Becker, 1992; Sax, 1984; Scotton, 1982). In Ontario for instance, where physician fee schedules are negotiated within a budget constraint, additional strict measures to control expenditure were implemented during 1991-2. These mechanisms included expenditure caps (capitation) and reductions in payment to high volume physicians (Chan, Anderson and Theriault, 1988).

It is apparent that these initiatives have had limited impact on the increasing cost of providing medical services. Despite numerous attempts and permutations on existing themes, few fee-setting models can claim to be effective in both containing costs and providing a tenable long-term and up-dateable solution to physician payment. Some of these approaches have been able to achieve short-term budgetary restraint, but many have had unintended consequences (for example: code creep - Chan et al, 1998; fee rigidity - Hsiao, Braun, Becker and Thomas, 1987). None are regarded as being economically tenable fee setting mechanisms for the longer term.

The most compelling criticism of these mechanisms is that they lack a rational, equitable and objective basis to determining medical professional fees. Most fee systems have evolved over time, with remuneration levels being "ratcheted" in line with economic growth or inflation indices. Until recently there had been no attempt to establish the true **value** of medical services **independent** of their **price** in the market place.

Parallel experiences in other occupations

Value, worth and price are concepts that have been considered by other industry and corporate sectors in considering remuneration issues. Job evaluation techniques vary, but one of the most commonly used evaluation methods assigns "points" to component tasks and adds these points to give a numerical total of "job worth". This rating allows the relative ranking of jobs, usually within a corporation. However, this process is heavily dependent on which job attributes are rated and how weights are differentially applied. Historically, factors and factor weights have been selectively chosen and used to maximize remuneration rates (Hartmann, Roos and Treiman, 1983). More recently competency- based pay techniques have gained favor as a tool for assessing the relative worth of particular employees. Under this paradigm, employees are assessed in terms of their ability, capacity or demonstrated experience in making a contribution to the performance of an organization and remunerated accordingly. However, these systems also encounter difficulties when measuring personal traits and behavioural competencies, particularly when the way competencies are usually introduced "is to hang them on the back of an existing performance related pay system" (Armstrong and Brown, 1998). Remunerating an employee on the basis of

characteristics the jobholder possesses, rather than what the job notionally requires from any employee in the same discipline puts the spotlight on **results** rather than on **how** those results are achieved (Armstrong and Brown, 1998). Because of its reliance on an assessment of an individual's performance it does not lend itself to an industry wide study of relative worth or value. Despite the fact that competency based pay systems appear to offer a means of establishing the value of an employee's contribution, Armstrong and Brown (1998) conclude that there is a "danger ...of implying a precision in competency measurement that does not exist".

Neither job evaluation nor competency based assessments are directly applicable to medical fee setting because they are designed to provide a relative assessment of work value within an organization, not across an entire industry. However, the basic principles employed by these methods and the results obtained from them are informative. While they do achieve some measure of objectivity, both these approaches are widely criticized because they do not assess the value of work or workers a priori³. The existing remuneration scales are either exploited by, or influence the determination of value. It is also apparent that the traditional medical fee schedules and fee setting mechanisms are afflicted with similar defects.

Clearly the major impediment to achieving a viable medical fee schedule has been the inability to objectively measure the true a priori value of work without reference to existing remuneration levels. In order to achieve a valid and equitable fee schedule a great deal must be known about the nature of work, its component factors and how these factors interact. And more importantly, how these factors can be accurately and reliably measured needs to be determined. Establishing the relative value of different tasks objectively is the key to consequently determining appropriate monetary compensation for these tasks - and not vice versa.

The literature detailing the US experience reveals that there was broad support for reforming the existing fee setting mechanisms (Cullen 1990; Hadley and Berenson, 1987; Stason 1987). Moreover, there was a general consensus that developing this system de novo offered the best opportunity to achieve a rational, equitable and sustainable fee setting mechanism (Berenson, 1987; Dunn and Becker, 1995; Orkin, 1995).

³ a priori: adj & adv. of concepts, knowledge; logically independent of experience; not derived from experience. The Concise Oxford Dictionary, 9th ed, 1995.

The Harvard Study

The US Congress, the Health Care Financing Administration (HCFA) and the Physician Payment Review Commission (PPRC) shared this view and funded a study conducted by Harvard School of Public Health and lead by William Hsiao. The terms of reference for the Harvard study were initially to develop a fee schedule and stipulated that the "inherent reasonableness" of such a schedule would be the primary criterion on which it would be judged. In a subsequent report, the PPRC recommended that a relative value scale should be based on resource costs, without endorsing any specific approach (American Medical Association, 1997; Hsiao, Braun, Dunn and Becker 1988; Hsiao, Braun, Dunn, Becker, Yntema, Verrilli, Stamenovic and Chen, 1992; Iglehart, 1988).

It is important to note that the Harvard study's terms of reference were to develop a fee schedule. However, Hsiao and his team were responsible only for determining the **mechanism** for fee setting. To do this they proposed that a resource based relative value scale was a function of three factors: total work of the physician, practice costs including professional liability insurance and opportunity cost of postgraduate specialist training (Hsiao, Braun, Dunn and Becker, 1988).

Determination of the monetary values of the cost conversion factors and the opportunity costs was not undertaken by the Harvard study and these factors were investigated independently of the study to develop the relative values of the total work component (American Medical Association, 1997).

Understanding the a priori value of work

Hsiao's work for the Harvard study is informative because it is the singular and most comprehensive overhaul of a medical fee schedule that includes most sectors of the profession. The literature reveals that this is rare undertaking. Most other attempts at reform have not been undertaken on this scale and these might be regarded as stopgap or band-aid measures (Katz, Charles, Lomas and Welch, 1997).

The Harvard study is also worthy of consideration because it alone attempts to define relative work value largely independently of market price determination mechanisms and existing remuneration scales. It should be noted that anesthesiologists in developing a relative value scale for anesthesiology services have undertaken similar work. However in this case the scope was greatly circumscribed, considering only one medical discipline. While the methodology for establishing relativities within this specialty is informative, this research does not provide benchmarks or guidance for

establishing cross-specialty relativities (American Society of Anesthesiologists, 1989; Revicki et al, 1990). Hsiao's work is also unique in that he and his team have developed an original methodology for determining total work value (Hadley and Berenson, 1987). This is a new and innovative achievement without parallel in other industries or organizations or in the psychological or human factors research.

This is not to say that Hsiao's work and methodology should be accepted as the only possible approach. Nor does it suggest that the work is without its own difficulties or criticisms (Roper, 1988). Some of the more balanced commentary in the literature regarding the Harvard study has cautioned that it should be regarded as research and therefore subject to further analysis and refinement. They also warn that an RBRVS should not be regarded as a panacea (Lee and Ginsberg, 1988; Roper, 1988; Todd, 1988).

Other applications of the Harvard model

In 1983, prior to the Harvard study, the Massachusetts Rate Setting Committee introduced a new fee schedule based on Hsiao's model. Relative fees provided by the model were not regarded as reasonable, understandable or acceptable by the medical profession. "Public outcry by the physician population led to a rescinding of this new fee schedule and a restoration of the old fees in the summer of 1984" (Egdahl and Manuel, 1985).

Several provincial medical associations also adopted the Hsiao model across Canada. Although many associations devoted several years to developing these fee schedules in line with Hsiao's methodology, none have been formally approved or implemented (Katz et al, 1997). Katz et al (1997) claim that this is the result of a failure to obtain physician consensus on the relative resource estimates, and that the initiative was abandoned and shelved indefinitely when threatened by a "mutiny of specialists". This criticism of Hsiao's method is echoed in the recommendation made by Egdahl and Manuel (1985) in their review of the Massachusetts experience, that "what is needed is a dialogue among physicians that attempts to arrive at an appropriate ranking of professional services." However, it could be argued that the failure of Hsiao's model in these two scenarios is more realistically a function of **how** the Harvard methodology was conducted rather than evidence of a basic methodological flaw. Hsiao et al gives prominence and emphasis to the importance of the consultative process as an essential process in developing relative work values. (Hsiao, Braun, Dunn and Becker, 1988).

Review of the Harvard Study Methods⁴

Hsiao's conception and measurement of the "total work" component is of particular interest because this component defines and encapsulates all the dimensions of work undertaken by medical professionals. They began with a systematic and consultative examination of the concept of work and its dimensions. Physicians described the important dimensions of their work as including time, mental effort, knowledge, judgment, diagnostic skill, physical effort and stress as well as the complexity of the patients medical problems, the seriousness of their condition and the possibility of patient harm. Ultimately multidimensional scaling analysis provided four salient dimensions of work: 1) time, 2) mental effort and judgment, 3) technical skill and physical effort and 4) stress. (Hsiao, Braun, Yntema and Becker, 1988). Using the last three dimensions in lieu of "intensity" allowed Hsiao and colleagues to simplify, measure and validate the definition of work (Hsiao, Yntema, Braun, Dunn and Spencer, 1988).

Estimating Total Work

Hsiao et al partitioned work into three periods: intraservice work, pre- and post- service work. Intraservice work is defined as that period of work where a medical professional sees a patient or performs a procedure. Pre-service work refers to the tasks undertaken in preparation for consulting that patient or performing the procedure, and post- service work includes all follow-up work. Because there is so much variation in the time and intensity that these components involve, it was not viable to measure the total work in a composite manner (Hsiao, Yntema, Braun, Dunn and Spencer, 1988).

Intraservice work

Because work dimensions such as mental effort, skill and stress are almost impossible to measure objectively, the Harvard study relies on subjective measures of these dimensions. Magnitude estimation (Stevens, 1975) was selected as the most reliable method of obtaining reproducible and valid measures of subjective judgments. Hsiao and colleagues asked respondents to rate services in relation to a reference service, which was assigned a value of 100 units. A surgeon, who judged another service to be 4 times more work than this reference service, assigned it a relative work value of 400. Ratings were unconstrained, allowing respondents to rate a service as high or low as they thought necessary to reflect reality.

The results showed that each of the four postulated dimensions of work, (time, mental effort and judgment, skill and physical effort and stress) all contribute to the overall concept of work, and

⁴ A more comprehensive overview of the Harvard Study Methodology is given in Appendix XX .

none of these dimensions can be discarded without significantly impairing the description of work (Hsiao, Yntema, Braun, Dunn and Spencer, 1988). The method yielded results that were considered reliable, valid and reproducible (Hsiao, Yntema, Braun, Dunn and Spencer, 1988).

Estimating Pre- and Post Service Work

The magnitude estimation was not thought to be suitable method for obtaining time estimates of pre- and post-service work. The intraservice period tends to be a discrete and well defined period allowing physicians to fairly accurately report the usual time spent on these tasks, whereas pre- and post service work is "...often fragmented and intermingled with other activities" (Dunn, Hsiao, Ketcham and Braun, 1988).

Instead Hsiao and colleagues used a twofold approach to quantifying pre and post service. They began by surveying medical practitioners on a set of 373 distinct services. Regression analysis produced predictions of pre and post service times for services outside this subset. Typologies of pre and post services were established and these were categorized into distinct classes.

Dunn et al (1988) report that pre- and post-service work accounts for a significant portion of the total work performed by medical practitioners and that the proportion may range from 7% to 70% of the total work for various specialties. However, the majority (75% of all medical practitioners) spend between 25 and 50% of their total work in providing pre- and post-service. "Therefore, with some exceptions, the *relative* work relationships among services are not changed greatly but incorporating pre-service and post-service work into computations of total work" (Dunn et al 1988).

Critique of the Harvard Study's approach to defining and measuring work

The literature offers a range of objections and criticisms of the methodology employed by the Harvard study, and these range from philosophical difficulties with the overall approach to denouncements of specific procedures.

A quote from Hadley and Berenson (1987) is illustrative of the more philosophical criticisms:

"Developing a formula for calculating relative costs from time inputs gives the appearance of replacing judgment by objective science. But it does not remove judgment from either the construction of the formula or the measurement of what goes into it. As soon as one permits the value of actual time to be modified by subjective notions of skill,

effort, and expertise, then the process becomes fundamentally a judgment-laden task for which there is no obviously correct solution...."

And further,

"...when the judgments take the form of assigning fixed numerical weights and restrictive and arbitrary mathematical formulas to vaguely defined and unmeasurable concepts such as skill, intensity of effort, complexity..... the results are castles built from sand".

Hadley and Berenson raise an important issue with regard to the subjective assessment of skill, effort, complexity and stress. While these are indeed nebulous concepts, the psychological literature does provide some support for the fact that they can be reliably measured. Tsang and Velazquez (1996) report that there is strong support for the notion that mental workload is multidimensional and that subjects are capable of accurately reporting the demands on separate workload dimension. Hendy, Hamilton and Landry (1993) also report that an "additive model provides an adequate method for combining factor ratings into an estimate of overall workload". The subjective rating scale technique for assessing the perceived amount of mental effort expenditure was found to be a reliable measure in two studies conducted by Paas and Van Merriënboer (1994). These researchers also found that the subjective ratings were far more sensitive and reliable with regard to changes in task complexity conditions compared with more objective psychophysical measures (cardiovascular measures) (Paas and Van Merriënboer, 1994).

Stason lends support to the Harvard study by observing that though rating concepts like complexity and stress are admittedly subjective and difficult, results have proved to be remarkably consistent and that this approach should be considered a methodologic challenge rather than an inherent weakness (Stason, 1987).

A more specific criticism of the Harvard study methodology relates to the use of vignettes describing procedures. Survey respondents used these specially developed vignettes to assign time and intensity ratings to each procedure. Lee (1988) reports that there were perceived inconsistencies between these vignettes and the CPT codes usually used by medical practitioners, particularly with reference to evaluation and management services. He argues that the use of vignettes may have hindered the assignment of accurate relative values in these instances.

Another criticism of the relative work value estimation method is that is based on averages. Roper (1988) makes the point that such averaging is likely to conceal geographic differences and will not

account for the extra effort and skill required from physicians who typically treat severely ill patients. Similarly, Todd (1988) argues that the methodology

"..uses estimates for average time, average total work and average costs. It assigns values for the average patient. Have any of us ever met an average physician or an average patient?"

However, he does acknowledge that this shortcoming is not an indictment of the relative value scale, but that these issues are a problem generic to any third-party payment system (Todd, 1988).

It is notable that few of these critics acknowledge that the Harvard study extensively used consultative processes for assigning and validating the ratings of time, mental effort, skills and stress. It is true that this consultative process resulted in a convergence, but it is not clear that these convergent ratings could be regarded as necessarily average ratings. Kahan reports that when consultative panels had widely different preliminary ratings of services, the discussion centered on the source of the difference. "Achieving a common understanding of the typical case would result in an agreement regarding what was to be rated and therefore a greater likelihood of similarity in the magnitude of rating"⁵ (Kahan, Morton, Farris, Kominski and Donovan, 1994).

Establishing links between specialties and developing a common relative value scale.

To this point, the Harvard study has been concerned with rating the intraservice, pre- and post-service work in each specialty against a reference item peculiar to that specialty. The research team had decided that it would not be valid to have individual specialist's rate the value of services with which they were not familiar. The cross-specialty comparison task "loomed from the beginning of the study as a seriously vexing problem" (Hsiao, Braun, Dunn and Becker, 1988). The essence of the problem was to establish a way of comparing the relative value of disparate specialty services and ranking them on one common scale. The objective was to determine the "proper parity" between these different services and procedures.

Method for establishing cross-specialty parity in relative work values

Panels of physicians from different specialties were consulted and they considered potential specialty -to- specialty "links". These link services were of two types: 1) the *same* service provided by one or more specialties, or 2) *equivalent services* where the service provided was different but of equal work value (American Medical Association, 1997; Dunn and Becker, 1995; Hsiao, Braun, Dunn, Becker, Yntema, Verrilli, Stamenovic and Chen, 1992).

⁵ Kahan et al are making a distinction between a statistical average (or mean) and a typical (most common, or modal) rating

A weighted least squares regression method was then used to place all the established cross-specialty link services on a common relative value scale, while largely preserving the within specialty relationships of services. (American Medical Association, 1997; Dunn and Becker, 1995).

Approximately 250 link services were identified from among 33 specialties and aligned onto a common relative value scale. "The number of links per specialty ranged from 4 to 80 with a median of 12.5 links" (Dunn and Becker, 1995; Hsiao, Braun, Dunn, Becker, Yntema, Verrilli, Stamenovic and Chen, 1992). Validation of the results using independent panels of physicians, as well a statistical verification using the "jackknife" method, showed the cross-specialty alignment to be reproducible" (Braun, Yntema, Dunn, DeNicola, Ketcham, Verrilli and Hsiao, 1988; Dunn and Becker, 1995; Hsiao, Braun, Dunn and Becker, 1988).

Extrapolation from surveyed to unsurveyed services

Because only a limited number of services (800) were assessed by survey, and only a subset of these were identified as being the same or equivalent across specialties, it was necessary to extrapolate work values for the larger "universe" of services defined by the Physicians' Current Procedural Terminology edition four (CPT4).

Initially the researchers grouped services into "families" (for e.g.: all services coded in coronary bypass surgery became a family, as did all new patient visits). The Harvard team then attempted to establish relative work values for unsurveyed services by extrapolating based on the differences in average charges for services. They theorized that if an unsurveyed service in a family had a 20% higher average charge than a surveyed service in that same family, then the physician work involved in the former should be 20% higher than in the latter. Within each family a surveyed service was nominated as the "benchmark" service for computing the relative work values for other services within that family. Multiplying the work value of this "benchmark" by charge-based ratios that represent the relationship between surveyed and unsurveyed services provided work values that can be used to construct a relative work value scale for a "dramatically " larger number of services. (Kelly, Hsiao, Braun, Sobol, and DeNicola, 1988). Kelly et al claim that this "...extrapolation method ...makes use of available charge data without building in their inherent distortions" (1988). However, this process produced relative work values that often seemed "incongruous or paradoxical" (American Medical Association, 1997).

Subsequently, in Phases Three and Four of the study, a new extrapolation method using small groups of physicians in consultation to assess and verify the relationships between surveyed and unsurveyed services was used and a review process was also developed (American Medical Association, 1997).

Kelly et al (1988) concluded that the analyses of results obtained by this family extrapolation method indicate that charges are highly predictive of relative work values across and within families of related services. They also report that where there is sufficient charge data available, the extrapolation method allows confident prediction of values of work for services. In cases where procedures are rarely performed or where there is insufficient charge data, they propose that a regression analysis method be utilized to establish relative work values of unsurveyed services.

Overview of the impact of cross-specialty alignment and extrapolation

As Hsiao expected, establishing the equivalence and parity of work across specialties proved to be one of the most innovative and difficult aspects of the Harvard study. A number of questions and concerns have subsequently been raised about the methods employed to establish a single scale of relative work values. For instance, Roper (1988) questions the choice of link services, and contemplates whether a different set of consultant physicians would have identified different link procedures, yielding different values. He also questions whether cross-specialty parity and equivalency would have been more accurately estimated had links been established using all measures of work, rather than only the intraservice work value.

Evaluation and Management Services

Apart from methodological concerns, the literature reveals that there was considerable disquiet among the medical profession with regard to the relativities between specialties. This was particularly the case when considering the relationship between "cognitive" and procedural services (Hadley and Berenson, 1987). The "undervaluation" of cognitive services (where tasks involve consultation, evaluation and management of patients rather than imaging, invasive or surgical procedures), is inherent in most fee-for-service payment systems (Berenson, 1987; Deeble, 1996). Removing this distortion while developing the resource based relative work value scale has proved difficult. Research undertaken by Braun, Hsiao, Becker and DeNicola (1988) and Hsiao, Braun, Dunn and Becker (1988) identifies the ambiguity inherent in CPT codes for these services. When using the vignettes, physicians closely agreed on work values for evaluation and management services (E/M). However, they found that "there may be large differences in the way that different specialties use the (CPT-4) billing codes available for E/M services. Moreover,

the work entailed within some of the E/M services may vary considerably" (Hsiao, Braun, Dunn and Becker, 1988). The fundamental problem here appears to be that the vignettes more accurately describe evaluation and management services. The "equivalent" CPT-4 codes are too ambiguous, allowing greater variation and therefore less accurate assignment of evaluation and management services, resulting in "disturbing rank order anomalies" (Lee and Ginsberg, 1988). Subsequent refinements and adjustments to the work value scale have addressed some of these value ratings of E/M services (American Medical Association, 1997).

Anesthesiology Services

As previously mentioned, anesthesiologists had developed a system of relative values for measuring the work of anesthetic services (American Society of Anesthesiologists Relative Value Guide, 1977-1989 (ASA RVG); Revicki et al, 1990) prior to the commencement of the Harvard study. The Relative Value Guide for these services took into account three components of work: 1) a value unit reflecting the average complexity or risk of the anesthetic procedure; 2) time units in either 10 or 15min periods; and 3) modifier units reflecting the severity of the patient's physical condition. It should be noted that the time component in each surgical procedure is controlled not by the anesthesiologist but by the surgeon (Cullen, 1990).

Revicki et al comment that the resource based relative value model developed by Hsiao assumes a constant relationship between time and other factors used to define the amount of physician work involved in a procedure. However, their research reports that this assumption may not hold true for anesthesiologists, particularly for some selected high-volume procedures (Revicki et al 1990). The Relative Value Guide developed by the American Society of Anesthesiologists is advocated as an alternative approach to establishing the value of anesthetic services because it includes a method for including factors of patient's physical condition as well as variable time factors and complexity of procedure factors. It also includes a review process for updating basic relative value units in light of new technology or procedural advances. (Revicki et al, 1990). Cullen reports that anesthesiologists have been able to demonstrate the importance and variability of "actual time" in the provision of anesthetic services. The ASA RVG scale also demonstrates that it is possible to include a method for measuring variable time factors in a relative value scale (as compared with the average time measures included in the Hsiao model). These variable time components peculiar to this specialty were factored into the resource based relative value scale developed by the Harvard study (Cullen, 1990).

Critique of the cross-specialty alignment and extrapolation procedures

The purpose of a resource based relativities scale was to establish the relative value of each (and hopefully) every medical service provided, across the range of specialist practitioners. Apart from a few anomalies exemplified by evaluation and management services and anesthesiology, the Harvard study largely achieves this objective. However, Hsiao and his colleagues do identify a number of possible limitations of the Harvard study, and these are reiterated by other commentators.

The reliance on CPT-4

Specifically, there seems to be widespread concern that the CPT-4 coding system is inadequate because it was not "...designed for payment purposes, and it lacks a clear conceptual basis and a set of agreed-upon principles for classifying, defining and coding the large and growing repertoire of physician services" (Hsiao, Braun, Dunn, Becker, Yntema, Verrilli, Stamenovic and Chen, 1992).

The CPT-4 is in widespread use among USA physicians and forms the basis of their charges. This charge data was used by the Harvard study as a primary factor in the extrapolation methodology. The CPT-4 system will also remain the primary mechanism of billing for physicians under the RBRVS. Thus, this coding system is central to both the development and the real-world functionality of the RBRVS (Kelly et al, 1988).

It is well understood that any classification system should strive for an optimal balance between the level of detail and the total number of classification codes, and to some extent the level of detailed codes between and within specialties will influence cross-specialty definitions of work and consequently their relativities. Coding errors will be most damaging when they yield misleading estimates of work among services. The impact of coding errors was not assessed in the Harvard study and Hsiao reports having some reservations about the quality and use of such secondary data (Kelly et al 1988).

Concerns about what was *not* measured

It is difficult to determine the extent to which the available literature is critical of Harvard study purely in terms of its methodology and measurement of relative work values, or whether these criticisms are a reflection of a concern about the ultimate influence this scale will have on remuneration levels. It could be argued that pecuniary considerations might have motivated some of these objections. However, these arguments also have some relevance in terms of objective and unbiased assessments of work in the medical profession.

One argument alludes to the phenomenon of "swings and roundabouts". The Hsiao model does not account for the "severity" of patient problems. "Typical" physicians who provide services to a range of patients would be overpaid for that proportion of their practice that is straightforward and uncomplicated, and conversely underpaid for those patients who are severely ill or at risk. This system reflects traditional physician custom, where some overpayment compensates for some underpayment. The Harvard model relies on typical or average practices in establishing relative work values and for the typical medical practice this will prove appropriate. It is also possible that skilled or expert physicians will perform some tasks in less time and with less effort and stress. These physicians will be "over-compensated" under the Hsiao model. This would similarly be the case for physicians who routinely work with straightforward or non-complicated cases. By comparison, physicians who typically perform services for severely ill patients, or those who regularly do more difficult "revision work" would be undercompensated (Roe, 1985).

A similar argument centres on "quality of care". It is argued that physicians who provide qualitatively better patient care should receive higher values (Hsiao, Braun, Becker and Thomas, 1987; Stason, 1987). The Harvard researchers recognize that their methodology does not include a measurement of competency or quality outcomes among physicians, nor does it recognize or account for patient outcomes or the benefits of various service procedures (Hsiao, Braun, Dunn and Becker, 1988; Hsiao, Braun, Becker and Thomas, 1988). Other researchers agree arguing that a complete analysis of value includes an assessment of both cost and benefit (Hadley and Berenson, 1987; Stason, 1987). These variables should ideally be included in physician fee scales as well as resource cost; "Cost alone is an incomplete measure of value" (Stason, 1987).

Consultative procedures during RVU development

A feature common to all phases of the Harvard study is the reliance on consultation with physicians and the role of the technical consulting groups (TCG's). These procedures were integral to the development of a relative value scale (American Medical Association, 1997; Dunn and Becker, 1995). Technical consulting groups were small groups of physicians from a variety of specialties (nominated by the American Medical Association). These TCG members acted as expert advisors to the Harvard research team and assisted in developing the vignettes used to survey a larger sample of physicians about work values. These TCG's were also involved in selecting reference items (for within specialty relativities) and in determining linkage items (for cross specialty alignment). In addition, the TCG's reviewed the data and work values returned by the survey data and subsequent statistical analysis (Dunn and Becker, 1995).

Post-implementation revision procedures

After implementation of the Resource Based Relative Value Scale (RBRVS) the Health Care Financing Administration (HCFA) developed a RVS refinement process which involved 24 Review Panels. Physicians from 42 specialties were represented on these Panels. The American Medical Association took the position that updating and maintaining the RVS should remain a scientific and clinical activity, and these activities should remain the responsibility of the medical profession. The American Medical Association formed and continues to support the RVS Update Committee (RUC). This Committee meets annually to reconcile, recalibrate and refine coding and work values in the RVS in light of identified anomalies, changing clinical procedures or technological innovations. Other researchers acknowledged and recommended that the RVU should be regarded as a first approximation, not a magic solution (Lee and Ginsberg, 1988) and that there should be a continued and iterative refinement process (Blumenthal and Epstein, 1992), consequently there is wide support for and participation in the revision process.

The RUC is supported by other committees - an Advisory Committee, a Research Committee, Cross Specialty and Reference List Sub Committees and a Health Care Professional Advisory Committee. These committees are responsible for surveying physicians and making recommendations to the RUC on necessary adjustments to the RVS (American Medical Association, 1997).

It is apparent that revision processes incorporating comprehensive and ongoing consultation with physicians are crucial to refining the RBRVS. In the five years since implementation interim RVU's have been ratified, new procedures and codes have been evaluated and included and adjustment have been made to RVU's to more accurately reflect physician work practices. For example, in the 1992-3 review, the Panel reviewed 791 codes and values. Of these 360 were assigned a higher value, and 35 were assigned a lower value. In subsequent years, the number of procedural codes recommended for review and adjustment has significantly declined. This indicates the increasing accuracy, validity and acceptance of the RBRVS (American Medical Association, 1997).

Roper (1988) suggests that the assumptions, statistical estimation and extrapolation techniques necessary to construct the Harvard scale for use as a national fee scale for physicians raises methodological concerns. "Each of the individual assumption, estimations and extrapolated values

by themselves might be quite reasonable. Nonetheless, a system built on multiple layers of estimates may not necessarily yield valid results. The sheer magnitude of the building process may constitute cause for concern" (Roper, 1988).

If the Harvard study was unsupported by the comprehensive consultative revision process, Roper's argument might be considered a valid criticism. However, the RBRVS has been and continues to be, closely scrutinized and adjusted in line with the realities of the medical profession.

The Canadian experience stands in contrast. Katz reports that the attempt to implement a Harvard model RBRVS for provincial Canadian physicians was unsuccessful largely because of the lack of consultation with, and involvement of physicians in the development process. (Katz et al, 1997).

The Australian Initiative

One Australian study has examined whether a Harvard model RVU could be applied to General Practice physicians in Australia. The study by Harris, Richardson, Farish and Saltman (1996) looked at whether GP's assessment of the value of their work correlated with independently measures assessments and whether different methods of measuring work value produced comparable value ratings. They found that scaling methods were of little use to GP's in reliably assessing the relative value of their work. Harris et al also reports that duration of patient consultation may be a reasonable proxy for relative work value assessment in General Practice. (Harris et al, 1996) Deeble also emphasizes that the Harris found that the reliable application of the Harvard scale required extensive GP training. It should be noted that this study had a limited scope and did not use a consultative process in developing or refining the methodology or results.

The present study is a more comprehensive assessment of the extent to which the Harvard model might be applied in the Australian environment. It has been established to undertake a review of professional relativities and work component of individual service items in the Medical Benefits Scale (MBS). This portion of the larger review process is the most involved of all the review activities, and this is consistent with the effort and emphasis in the Harvard study. It was the Board's intention that the definitive work undertaken in the USA should form the common reference for this study. The National Centre for Classification in Health has undertaken an extensive information gathering exercise which has included a "mapping" of all MBS items to the US CPT-4 schedule. The Feasibility Study (National Coding Centre, 1996, National Centre for Classification in Health, 1997) found that it was possible to correlate similar items in the US and Australian schedules, although there was an incomplete and imperfect alignment of items.

Subsequent work on collecting, collating and modeling work values has been undertaken in a consultative fashion with the medical profession. The results of this work are presented later in this document.

Bibliography

- Altman, D.G., Gore, S.M., Gardner, M.J. and Pocock, S.J. 1997 Statistical guidelines for contributors to medical journals. In Gardner, M.J. and Altman, D.G. (eds) 1997. *Statistics with Confidence. Confidence intervals and statistical guidelines*. British Medical Journal., The Universities Press, Belfast.
- AMA (American Medical Association) 1997. *Medicare RBRVS: The physicians' guide*. AMA Health Policy Group, USA.
- Armstrong, M. and Brown, D. 1998. Relating competencies to pay. The UK experience. *Compensation and Benefits Review*. Vol 30 (3), 28-39.
- Berenson, R.A. 1987. Physician payment reform: Finally. *Annals of Internal Medicine*. Vol 107, (6), 929-931.
- Blumenthal, D. and Epstein, AP.M. 1992. Physician payment reform - Unfinished business. *The New England Journal of Medicine*. Vol 326, (20), 133-1334.
- Braun, P., Yntema, D.B., Dunn, D., DeNicola, M., Ketcham, T., Verrilli, D. and Hsiao, W.C. 1988. Cross-specialty linkage of resource based relative value scales. Linking specialties by services and procedures of equal work. *Journal of the American Medical Association*. Vol 260, (16), 2390-2396.
- Braun, P., Hsiao, W.C., Becker, E.R. and DeNicola, M. 1988. Evaluation and management services in the resource based relative value scale. *Journal of the American Medical Association*. Vol 260, (16), 2409-2417.
- Chan, B., Anderson, G.M. and Theriault, M.E. 1998. Fee code creep among general practitioners and family physicians in Ontario: Why does the ratio of intermediate to minor assessments keep climbing? *Canadian Medical Association Journal*. Vol 158, (6), 749-754.
- Cullen, D.J. 1990. Resource based relative value scale and reimbursement for anesthesiology services. Time is of the essence. *The Journal of Anesthesiology*. Vol 73, (4), 593-595.
- Deeble, J.S. 1996. Valuing medical work. No fee schedule can reflect all work dimensions. *Medical Journal of Australia*. Vol 165, 7.
- Department of Health and Family Services. 1998. The Medical Benefits Schedule Review. On the Departmental Web Site at www.msrtf.gov.au/~msrtf/backgrnd.htm.
- Dunn, D.L. and Becker, E.R. 1995. Validating the resource based relative value scale cross specialty alignment. A survey of double boarded physicians. *Medical Care*. Vol 33, (10), 975-987.
- Dunn, D., Hsiao, W.C., Ketcham, T.R. and Braun, P. 1988. A method for estimating the preservice and postservice work of physicians' services. *Journal of the American Medical Association*. Vol 260, (16), 2371-2378.
- Egdahl, R.H and Hertenstein, R.D. 1987. An access-oriented negotiated fee schedule. The Caterpillar experience. *Annals of Surgery*. Vol 206, (3), 349-357.

- Egdahl, R.H. and Manuel, B. 1985. A consensus process to determine the relative complexity-severity of frequently performed surgical services. *Surgery, Gynecology and Obstetrics*. Vol 160, 403-406.
- Hadley, J. and Berenson, R.A. 1987. Seeking the just price: Constructing relative value scales and fee schedules. *Annals of Internal Medicine*. Vol 106, 461-466.
- Harris, E.A., Richardson, J., Farish, S.J. and Saltman, D.C. 1996. The reliability and validity of work measurement in Australian general practice consultations. *Medical Journal of Australia*. Vol 165, 18-21.
- Hartmann, H.I., Roos, P.A. and Treiman, D.J. 1983. An agenda for basic research on comparable worth. In Hartman, H.I (Ed) *Comparable work. New directions for research*. 1983, p11.
- Hendy, K.C., Hamilton, K.M and Landry, L.N. 1993. Measuring subjective workload: When is one scale better than many? *Human Factors*. Vol 35, (4), 579-610.
- Hsiao, W.C. and Becker, E.R. 1989. Paying physicians according to their resource costs: the development of a resource based relative value scale. *Health Policy*. Vol 12, 257-261.
- Hsiao, W.C., Braun, P., Becker, E.R. and Thomas, S.R. 1987. The resource based relative value scale. Toward the development of an alternative physician payment system. *Journal of the American Medical Association*. Vol 258, (6), 799-802.
- Hsiao, W.C., Braun, P., Dunn, D. and Becker, E.R. 1988. Resource based relative values. An overview. *Journal of the American Medical Association*. Vol 260, (16), 2347-2353.
- Hsiao, W.C., Braun, P., Dunn, D.L., Becker, E.R., Yntema, D., Verrilli, D.K., Stamenovic, E. and Chen, S-P. 1992. An overview of the development and refinement of the resource based relative value scale. The foundation for reform of US physician payment. *Medical Care*. Vol 30, (11), Supplement 1-9.
- Hsiao, W.C., Braun, P., Kelly, N.L. and Becker, E.R. 1988. Results, potential effects and implementation issues of the resource based relative value scale. *Journal of the American Medical Association*. Vol 260, (16), 2429-2438.
- Hsiao, W.C., Braun, P., Yntema, D. and Becker, E.R. 1988. Estimating physicians' work for a resource based relative value scale. *The New England Journal of Medicine*. Vol 319, (13), 835-841.
- Hsiao, W.C. and Stason, W. 1979. Toward developing a relative value scale for medical and surgical societies. *Health Care Financing Review*. Vol 1, 23-28.
- Hsiao, W.C., Yntema, D., Braun, P., Dunn, D. and Spencer, C. 1988. Measurement and analysis of intraservice work. *Journal of the American Medical Association*. Vol 260, (16), 2361-2370.
- Iglehart, J.K. 1988. Health Policy Report. Payment of physicians under Medicare. *The New England Journal of Medicine*. Vol 318, (13), 863-868.
- Kahan, J.P., Morton, S.C., Farris, H.H., Kominski, G.F. and Donovan, A.J. 1994. Panel Processes for Revising Relative Values of Physician work. A pilot study. *Medical Care*. Vol 32, (11), 1069-1085.

BIBLIOGRAPHY

- Katz, S.J., Charles, C., Lomas, J. and Welch, H.G. 1997. Physician relations in Canada: Shooting inward as the circle closes. *Journal of Health Politics, Policy and Law*. Vol 22, (6), 1413-1431.
- Kelly, N.L., Hsiao, W.C., Braun, P., Sobol, A. and DeNicola, M. 1988. Extrapolation of measures of work for surveyed services to other services. *Journal of the American Medical Association*. Vol 260, (16), 2379-2384.
- Lee, P.R., and Ginsberg, P.B. 1988. Physician payment reform: an idea whose time has come. *Journal of the American Medical Association*. Vol 260, (16), 2441-2443.
- National Coding Centre, 1996. *Consultancy to examine the potential for mapping the professional components of services described in overseas schedules with the Medicare Benefits Schedule*. Stage 1: Report to Medicare Schedule Review Task Force - 26 November 1996 (unpublished).
- National Centre for Classification in Health, 1997. *Consultancy to examine the potential for mapping the professional components of services described in overseas schedules with the Medicare Benefits Schedule*. Stage 2: Report to Medicare Schedule Review Board - 2 March 1997 (unpublished).
- Orkin, F.K. 1993. Moving toward value based anesthesia care. *Journal of Clinical Anesthesiology*. Vol 5, 91-98.
- Paas, F.G.W.C and Van Merrienenboer, J.J.G. 1994. Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*. Vol 79, (1, Pt2), 419-430.
- Penington, D.G., Cashman, J.D. and Kearney, B.J. 1984. *Committee of Enquiry into Rights of Private Practice in Public Hospitals*. Final Report AGPS Canberra. p 98.
- Rakich, J.S. and Becker, E.R. 1992. United States physician reform: Background and comparison with the Canadian model. *Health Care Management Review*. Vol 17, (1), 9-19.
- Revicki, D.A., Orkin, F.K., Luce, B.R., McMenamain, P. and Weschler, J.M. 1990. Physician payment reform: Anesthesiology as a case study. *Anesthesiology*. Vol 73, 760-769.
- Roberts, R. 1988. Medical fee setting. A case study. MBA thesis (unpublished).
- Roe, B.B. 1985. Sounding board. Rational remuneration. *The New England Journal of Medicine*. Vol 313, (20), 1286-1289.
- Roper, W.L. 1988. The resource based relative value scale: A methodological and policy evaluation. *Journal of the American Medical Association*. Vol 260, (16), 2444-2446.
- Sax, S. 1984. *A strife of interests. Politics and policies in Australian health services*. George Allen and Unwin. p81.
- Scotton, D. 1982. Discussion following George Palmer. Commonwealth/State fiscal relationships and the financing and provision of health services. In Tatchell, P.M. (Ed) *Economics and Health*, 1981. Australian National University. Canberra.
- Stason, W.B. 1987. What cornerstone is best for physician payment reform? *Annals of Internal Medicine*. Vol 106, (3), 468-470.
- Stevens, S.S. 1975. *Psychophysics*. New York. John Wiley.

Todd, J.S. 1988. At last, a rational way to pay for physicians' services? *Journal of the American Medical Association*. Vol 260, (16), 2439-2441.

Tsang, P.S. and Velazquez, V.L. 1996. Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*. Vol 39, (3), 358-381.

REVIEW OF THE HARVARD STUDY

The following provides a more detailed, and somewhat technical review of the Harvard methodology. It is included for the reader interested in issues such as sampling techniques and statistical analysis and results.

Definition of Work

Hsiao et al propose the formula:

$$\text{RBRV} = (\text{TW}) \times (1 + \text{RPC}) \times (1 + \text{AST})$$

where RBRV = resource based relative value

TW = Total work

RPC = Relative specialty practice cost

AST = amortized value for the opportunity cost of specialist training

It is the "total work" component that is of particular interest because this component defines and encapsulates all the dimensions of work undertaken by medical professionals. The Harvard study draws on research previously conducted by Hsiao and Stason (1979). This early study posited that there were two components of total work: complexity and duration (time taken on each procedure). Examination of the data from this study revealed that there was a high degree of correlation between these two factors, which suggested that physicians were not adequately distinguishing between the two concepts. Hsiao and Stason hypothesized that the overlap between time and complexity led to an "exaggeration of their product" (Hsiao, Braun, Yntema and Becker 1988). In an attempt to overcome this problem a subsequent study sought a global estimation of work that incorporated both these dimensions. This approach however was also thought to be unrealistic as it reliability elicited an underestimation of work, particularly in lengthy procedures. (Hsiao, Braun, Yntema and Becker, 1988). For the Harvard study, they began with a systematic and consultative re- examination of the concept of work and its dimensions. Physicians described the important dimensions of their work as including time, mental effort, knowledge, judgment, diagnostic skill, physical effort and stress as well as the complexity of the patients medical problems, the seriousness of their condition and the possibility of patient harm. Ultimately multidimensional scaling analysis provided four salient dimensions of work: 1) time, 2) mental effort and judgment, 3) technical skill and physical effort and 4) stress. (Hsiao, Braun, Yntema and Becker, 1988). Using the last three dimensions in lieu of "intensity" allowed Hsiao and colleagues to simplify, measure and validate the definition of work (Hsiao, Yntema, Braun, Dunn and Spencer, 1988).

Estimating Total Work

Hsiao et al partitioned work into three periods: intraservice work, pre- and post- service work. Intraservice work is defined as that period of work where a medical professional sees a patient or performs a procedure. Pre-service work refers to the tasks undertaken in preparation for consulting that patient or performing the procedure, and post- service work includes all follow-up work. Because there is so much variation in the time and intensity that these components involve, it was not viable to measure the total work in a composite manner (Hsiao, Yntema, Braun, Dunn and Spencer, 1988).

Intraservice work

Because work dimensions such as mental effort, skill and stress are almost impossible to measure objectively, the Harvard study relies on subjective measures of these dimensions. Magnitude estimation (Stevens, 1975) was selected as the most reliable method of obtaining reproducible and valid measures of subjective judgments. Hsiao and colleagues asked respondents to rate services in relation to a reference service, which was assigned a value of 100 units. A surgeon, who judged another service to be 4 times more work than this reference service, assigned it a relative work value of 400. Ratings were unconstrained, allowing respondents to rate a service as high or low as they thought necessary to reflect reality.

Hsiao et al postulated a general functional relationship between the judgments of work and its dimensions:

$$W = A \times t^{\beta} \times M^{\tau} \times S^{\phi} \times R^{\delta}$$

where W = work

A = a constant

t = time

M = mental effort

S = technical skill

R = stress

β , δ , τ and ϕ give quantitative weights of each dimension and predict the percent change in the level of work predicted by a percent change on the ratings of dimensions. (Hsiao, Yntema, Braun, Dunn and Spencer, 1988).

Survey Methods

Hsiao et al used a stratified random sample to ensure that the sample was representative of each specialty and geographic region. Physicians who worked in patient care less than 20 hours per week were excluded, as were those in residency training or those over 65 years of age. After a

pilot survey of 90 physicians a national survey of 3164 physicians was undertaken. One thousand, nine hundred and seventy-seven (1977) respondents completed in recorded telephone interviews. The overall response rate to the survey was 62.5%. (Hsiao, Yntema, Braun, Dunn and Spencer, 1988).

Analysis and Results

The data was transformed using common logarithms, and this procedure provided reasonable symmetry and homoscedasticity which are desirable for statistical analysis such as regression (Hsiao, Yntema, Braun, Dunn and Spencer, 1988; Altman, Gore, Gardner and Pocock, 1997).

For more detail refer to *Table 1. – Mean Values of Intraservice Work and its Dimensions Before Cross-Specialty Linkage for Services in 18 Specialties*, Hsiao, Yntema, Braun, Dunn and Spencer, 1988, p.2364.

Reliability and validity of the data obtained

By comparing the ratings obtained from the pilot survey with those obtained from the national survey, Hsiao et al were able to verify that the ratings are reproducible and that the method for obtaining these ratings is reliable (Hsiao, Yntema, Braun, Dunn and Spencer, 1988).

The intraclass correlation method (Spearman-Brown formula) indicates the extent to which two physicians in the same specialty agree on ratings of various services. The results here produce correlation coefficients in the range of 0.991 to 0.998. This means that ratings made by different groups of experts are in close agreement. It is possible to get close agreement between experts and for both of them to be objectively wrong. To test whether this was the case, multiple linear regressions were also performed and the coefficients (r^2) were all greater than .964 indicating "surprisingly close unity..... This high degree of consistency constitutes statistical evidence that the ratings of work are probably valid" (Hsiao, Yntema, Braun, Dunn and Spencer, 1988). As another check on the validity of the ratings made by respondents, the time durations estimated by physicians in the survey based on their recall were compared to external, objective measures of the duration of surgery obtained from operating room log times. "The times reported by the two methods were very similar" (Hsiao, Yntema, Braun, Dunn and Spencer, 1988).

Overall it was also found that each of the four postulated dimensions of work, (time, mental effort and judgment, skill and physical effort and stress) all contribute to the overall concept of work, and

none of these dimensions can be discarded without significantly impairing the description of work (Hsiao, Yntema, Braun, Dunn and Spencer, 1988).

Estimating Pre- and Post Service Work

The magnitude estimation was not thought to be suitable method for obtaining time estimates of pre- and post-service work. The intraservice period tends to be a discrete and well defined period allowing physicians to fairly accurately report the usual time spent on these tasks, whereas pre- and post service work is "...often fragmented and intermingled with other activities" (Dunn, Hsiao, Ketcham and Braun, 1988).

Instead Hsiao and colleagues used a twofold approach to quantifying pre and post service. They began by surveying medical practitioners on a set of 373 distinct services. Regression analysis produced predictions of pre and post service times for services outside this subset. Typologies of pre and post services were established and these were categorized into distinct classes.

To predict pre and post service times from intraservice time data the following formula was used:

$$T_p = a + (b \times W_i) + (c \times T_i)$$

where T_p = pre or post service time

W_i = intraservice work

T_i = intraservice time

To statistically estimate the pre- and post service work from intraservice work data, the following formula was used:

$$W_p = (T_p)(W/T)$$

where W_p = indicates pre- or post- service work

T_p = pre-or post service time

W/T = rate of work per unit of time for pre-and post- services (W_p/T_p) equal to work divided by time (from intraservice time data)

Because there were different categories of services, and different typologies of pre- and post-service work, different regression models and separate parameters were used to estimate the pre- and post service times for different services (i.e.; in statistical terms, evaluation and management services, anesthesia, hospital and emergency department services were treated differentially). (Dunn, et al, 1988).

Survey methods

In this section of the Harvard study, Hsiao and colleagues again conducted a pilot survey and a subsequent national survey of physicians. Five hundred and sixty (560) were surveyed and 367 completed interviews (for a response rate of 66%). The two surveys generated pre- and post-service time data for a total of 154 different services. (Dunn et al, 1988)

Analysis and results

Data was log transformed and regression analysis was used to obtain estimates of pre- and post-service time for unsurveyed services.

For more detail refer to *Table 3. – Preservice and Postservice Time for Selected Services*, Dunn et al, 1988, p.2375.

Reliability and validity of the data

Dunn et al found that the data were reproducible and reliable (1988, p2376). They used two methods to assess the "reasonableness" or validity of the data. Physicians in technical consulting groups reviewed the data and they concluded that the estimates of pre- and post-service times "generally reflected their clinical experience", although in 4% of cases they detected systematic errors in the predictive method. These errors were subsequently modified (Dunn et al, 1988). Although information from other more objective sources were limited, the Harvard study compared their predicted estimates of pre- and post- service times with data from two surveys of physicians' practice, the American Medical Association Socioeconomic Survey of Physicians and the Mendenhall study (cited in Dunn et al 1988). They found that the results of their own estimations of pre- and post- service times were within 20% of those reported in the American Medical Association and Mendenhall studies. (Dunn et al 1988).

Once the pre- and post-service time data was analyzed it was used in calculating the estimates of pre- and post-service work. Again the reliability and validity of this data was assessed by a consultative group of physician investigators. Their assign of work value differed from the estimated work values in only 7% of cases. The predicted work values were adjusted in line with clinician opinions. (Dunn et al, 1988).

For more detail refer to *Table 7. – Preservice and Postservice Work for Selected Services*, Dunn et al, 1988, p.2377.

Overall, Dunn et al (1988) report that pre- and post-service work accounts for a significant portion of the total work performed by medical practitioners and that the proportion may range from 7% to 70% of the total work for various specialties. However, the majority (75% of all medical practitioners) spend between 25 and 50% of their total work in providing pre- and post-service. "Therefore, with some exceptions, the *relative* work relationships among services are not changed greatly but incorporating pre-service and post-service work into computations of total work" (Dunn et al 1988).

Establishing links between specialties and developing a common relative value scale.

To this point, the Harvard study had been rating the intraservice, pre- and post-service work in each specialty against a reference item peculiar to that specialty. The research team had decided that it would not be valid to have individual specialist's rate the value of services with which they were not familiar. The cross-specialty comparison task "loomed from the beginning of the study as a seriously vexing problem" (Hsiao, Braun, Dunn and Becker, 1988). The essence of the problem was to establish a way of comparing the relative value of disparate specialty services and ranking them on one common scale. The objective was to establish the "proper parity" between these different services and procedures.

Method for establishing cross-specialty parity in relative work values

Panels of physicians from different specialties were consulted and they considered potential specialty -to- specialty "links". These link services were of two types: 1) the *same* service provided by one or more specialties, or 2) *equivalent services* where the service provided was different but of equal work value (AMA, 1997; Dunn and Becker, 1995; Hsiao, Braun, Dunn, Becker, Yntema, Verrilli, Stamenovic and Chen, 1992).

A weighted least squares regression method was then used to place all the established cross-specialty link services on a common relative value scale, while largely preserving the within specialty relationships of services. (AMA, 1997; Dunn and Becker, 1995).

Approximately 250 link services were identified from among 33 specialties and aligned onto a common relative value scale. "The number of links per specialty ranged from 4 to 80 with a median of 12.5 links" (Dunn and Becker, 1995; Hsiao, Braun, Dunn, Becker, Yntema, Verrilli, Stamenovic and Chen, 1992). Validation of the results using independent panels of physicians, as well a statistical verification using the "jackknife" method, showed the cross-specialty alignment to

be reproducible" (Braun, Yntema, Dunn, DeNicola, Ketcham, Verrilli and Hsiao, 1988; Dunn and Becker, 1995; Hsiao, Braun, Dunn and Becker, 1988).

Extrapolation from surveyed to unsurveyed services

Because only a limited number of services (800) were assessed by survey, and only a subset of these were identified as being the same or equivalent across specialties, it was necessary to extrapolate work values for a larger "universe" of services defined by the Physicians' Current Procedural Terminology edition four (CPT4).

Initially the researchers grouped services into "families" (for e.g.: all services coded in coronary bypass surgery became a family, as did all new patient visits). The Harvard team then attempted to establish relative work values for unsurveyed services by extrapolating based on the differences in average charges for services. They theorized that if an unsurveyed service in a family had a 20% higher average charge than a surveyed service in that same family, then the physician work involved in the former should be 20% higher than in the latter. Within each family a surveyed service was nominated as the "benchmark" service for computing the relative work values for other services within that family. Multiplying the work value of this "benchmark" by charge-based ratios that represent the relationship between surveyed and unsurveyed services provided work values that can be used to construct a relative work value scale for a "dramatically " larger number of services. (Kelly, Hsiao, Braun, Sobol, and DeNicola, 1988). Kelly et al claim that this "...extrapolation method ...makes use of available charge data without building in their inherent distortions" (1988). However, this process produced relative work values that often seemed "incongruous or paradoxical" (AMA, 1997).

Subsequently, in Phases Three and Four of the study, a new extrapolation method using small groups of physicians in consultation to assess and verify the relationships between surveyed and unsurveyed services was used and a review process was also developed (AMA, 1997).

Kelly et al (1988) concluded that the analyses of results obtained by this family extrapolation method indicate that charges are highly predictive of relative work values across and within families of related services. They also report that where there is sufficient charge data available, the extrapolation method allows confident prediction of values of work for services. In cases where procedures are rarely performed or where there is insufficient charge data, they propose that a regression analysis method be utilized to establish relative work values of unsurveyed services.