Mark A. Hall

The Structure and Enforcement of Health Insurance Rating Reforms

Requiring health insurers to cover everyone who applies regardless of health status—an approach called "guaranteed issue"—is severely hampered without accompanying rating restrictions that keep insurance affordable for higher-risk people. The degree of rating flexibility also determines how much insurers can continue to compete based on their skills at risk selection, and how well they can counter adverse selection. Therefore, the structure and enforcement of rating reforms are essential to how insurance market reforms function. Based on an in-depth qualitative study in seven states with insurers, agents, and regulators, this article explains the factors that determine the stringency of rating reforms, and details how various aspects of rating restrictions can be used strategically to engage in greater risk segmentation than first appears possible. The article concludes by reflecting on the appropriate degree of complexity in rating rules, and it offers recommendations for crafting rating reforms that avoid unintended consequences.

The Health Insurance Portability and Accountability Act of 1996 (HIPAA) has been called the "most significant federal health care reform in a generation" (Atchinson and Fox 1997), one that President Clinton said would "seal the cracks that swallow as many as 25 million Americans who cannot get insurance when they change or lose jobs" (Kuttner 1997). Others, such as Rep. Fortney Stark, have called HIPAA placebo legislation that does little or nothing, and so amounts to a "cruel hoax" (Bureau of National Affairs 1997). This disjuncture is explained in part by the fact that, while HIPAA has sweeping protections against being denied insurance coverage, it says nothing about what prices insurers may charge when coverage is mandated. HIPAA leaves rate regulation entirely to the states, and so the degree of effective protection varies widely across the country. The General Accounting Office (GAO) reports that some states allow insurers to charge as much as six times their standard rates for "federally eligible" subscribers who move from group to individual coverage (U.S. GAO 1998). This vividly documents that guarantees of insurance coverage are severely hampered without provisions that limit how much insurers can increase rates for higher-risk subscribers.

Most states in fact have set such limits in the small-group market. Because both small-group rating reforms and HIPAA's guaranteed issue requirement were in place in many states several years prior to HIPAA, and because the rating reforms vary among the states, the pre-HIPAA experience offers a rich source to draw from in anticipating the likely impact of extending similar protections to the indi-

Mark A. Hall, J.D., *is a professor of law and public health in the Department of Public Health Sciences at Wake Forest University School of Medicine.* *The Robert Wood Johnson Foundation supported this study under its Changes in Health Care Financing and Organization initiative.* *Address correspondence to Prof. Hall at Department of Public Health Sciences, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157-1063.*

Inquiry 37:376–388 (Winter 2000/2001)

 0046-9580/00/3704–0376$1.25

vidual market (Nichols and Blumberg 1998). Such extensions are being debated across the country, both by states seeking to craft workable reforms for the individual market and in Congress, as a possible expansion of HIPAA.

To better inform this debate, this article reports findings from an extensive, three-year qualitative study of insurance market reforms in seven states.[1] The primary focus is on rating restrictions in the small-group market, but the study also includes some examples from states with individual market reforms. A qualitative approach is especially useful here because it allows us to focus separately on different versions of the rating component of market reforms. In contrast, quantitative evaluations typically study rating restrictions in combination with other reform measures and without distinguishing different types of restrictions (Jensen and Morrisey 1999).

The article begins with an overview of the structure of rating reforms and a summary of the study methodology. It then describes how these reforms have affected a variety of rating practices, and concludes with recommendations for shaping effective regulatory policy.

The Structure of Rating Reforms

The aim of rating reforms is to prevent insurers from varying their prices among purchasers more than a defined amount for policies with similar benefits and subscribers with similar demographics (or "case characteristics"). Despite this common aim, states vary considerably in their rating rules. There are three basic approaches: rating bands, adjusted or modified community rating, and pure community rating—each of which requires successively greater degrees of rate compression.[2] The key distinguishing factor is the extent to which insurers may reflect various risk factors in their rates.

Rating bands allow health status to affect rates, but only within a defined range. In the small-group market, states originally allowed ranges of ±25% to 35%, but many since have tightened the range to ±10% to 20%. Allowing no rate variation based on individual health status is called modified community rating. This is "modified" or "adjusted," rather than pure, community rating because full or substantial adjustment still is allowed for age and sometimes for gender. States using this approach also typically allow adjustments for other factors, such as the employer's industry, or whether employees smoke.

Pure community rating eliminates most of these factors (including age/gender factors) and retains only location, benefits, and family size as rating factors. Some states use a form of community rating that allows some additional variation in rates, but if the rating rules greatly restrict the degree of age/gender rating, the rating method still qualifies as nearly pure community rating. Unconstrained age/gender rating is the critical feature because, according to interviews with actuaries, these factors produce variations of fivefold or more, even if individual health status is entirely removed. If other rating factors are allowed for individual health status or other characteristics, these create even wider variation than demographics alone, but these other factors usually are limited to much smaller ranges than are permitted for demographics.

Proposals that allow separate blocks of business add another layer of complexity. Traditionally, many private insurers have treated all small-group business as a single block (or "book" or "class") of business for purposes of rating, product design, and marketing. However, some maintain distinct blocks when their products are sold through separate sales forces, when they are acquired from another insurer, or when they have fundamentally different designs, such as health maintenance organizations (HMOs) vs. indemnity products. In keeping with this tradition, many states apply rating limits separately to a limited number of blocks defined in this manner. To prevent circumvention of the rating limits by block gerrymandering (segregating high and low risks into different blocks), a number of states also limit the pricing variation *among* blocks. For this purpose, a 20% limit is typical.

Regardless of the rating rules adopted, every state allows rates to reflect certain fundamental factors that drive health care costs. These include the actuarial value of differences among benefits in different policies, the number of people covered in the family unit, and geographic location. Thus, the layers of allowable factors outlined above permit rate variation in addition to what is justified by benefits, family size, and location.

A final dimension in which rating reforms restrict price variation is in the amount that insurers may increase a purchaser's rates over time. Pure community rating requires that an insurer increase prices in lock step for all purchasers within a region and with the same coverage, and adjusted community rating allows for increases only due to age. Rating bands, however, create the possibility that purchasers with

initially low rates could receive substantial renewal increases, a practice known as "low-balling" or "churning." Most states with rating bands limit year-to-year premium increases for any given subscriber to 10% to 15% above the insurer's "trend." Trend is defined as the increase in the insurer's rates for new business. The concept is to allow marketwide cost increases that are driven by technology advances, inflation in the medical sector, and the like, but to limit increases that reflect group-specific health risk. (Trend is keyed to new business rates because this is where insurers are the most competitive.)

To summarize, consider the treatment under these rating reforms of two different applicants, one a very good risk and the other a very bad risk, who seek coverage from an insurer whose average yearly rate is $3,000. Pure community rating requires the insurer to charge each of these subscribers exactly the same price for the same set of benefits, if they live in the same location. Also, their renewal rates must be the same as those that the insurer charges new subscribers. Adjusted community rating allows rates to vary only due to age and gender, but these factors alone might cause rates to range from $1,000 to $6,000. Rating bands allow additional rate variation to reflect each purchaser's particular health characteristics, but only within the defined range. In a state with a ±25% range, rates for a purchaser in a demographic bracket with a median price of $3,000 could vary from $2,250 to $3,750, depending on health status. A low-risk purchaser whose health status worsens could be moved toward the top of the band, but only in increments of 10% to 15% a year (on top of changes in age and overall average rate increases).

Study Methodology

To evaluate the functioning of these complex rating rules, seven study states were selected in 1996 to reflect a range of intensity in individual and small-group market reforms, as indicated in Table 1. These states also reflect different demographic, economic, and market characteristics. In each study state, in-depth interviews were conducted with the two to four regulators who have the most knowledge of these laws, with the five to six independent agents who specialize in health insurance (except in North Carolina, where only three agents were interviewed), and with actuaries, underwriters, marketers, product designers, or lawyers at up to four of the top insurers, including Blue Cross, leading HMOs, and commercial indemnity insurers. We also conducted interviews at 11 national insurance companies with business in some or all of these study states. In all, more than 100 subjects were interviewed in 1997, and more than 60% of these (or their substitutes) were interviewed again in 1998. Interviews lasted approximately one to two hours each and were based on an interview guide, but the discussions were free ranging and the coverage of topics varied somewhat.[3] Most interviews were in person and one-on-one, but a number were done over the phone and in groups of two to five people. Finally, documentary information was collected and analyzed using primarily qualitative methods. The information included insurers' operating data, data reported to state regulators,

Table 1. Rating reforms in each study state

State and effective date of reform law	Small-group rating	Guaranteed issue pre-HIPAA	Individual-market rating
New York, 1993	Pure community rating	All products	Pure community rating
Vermont, 1992	Nearly pure community rating	All products	Nearly pure community rating
Colorado, 1992	Adjusted community rating	Designated products	
Florida, 1992	Adjusted community rating	All products	
North Carolina, 1992[a]	±20% rating bands	Designated products	
Iowa, 1992	±25% within a block, 20% among blocks	Designated products	30% among blocks
Ohio, 1993	±35% rating bands	Limited open enrollment	

[a] North Carolina was not studied quite as extensively as the other six states since it was used primarily to field test and refine interview techniques.

prior public policy and industry studies, and news articles in local and national publications.

The Functioning of Rating Reforms

This paper explores, in several dimensions and with different levels of complexity, the impact rating reforms have on insurers' rating practices. The simplest question is whether it makes sense to allow any rating flexibility at all. Then, I examine how effective these reforms are in constraining rate variation within the stated ranges. Next, I look at how these reforms affect rate increases and rating by blocks of business. Fourth, I explain how the ability to adjust rates for differences in benefits introduces important complications. Finally, I address issues relating to adjustments for age, location, and family size.

The Case For and Against Rating Flexibility

One might be tempted, for simplicity's sake, to adopt the purest form of community rating. However, it is noteworthy that even this requires adjustments for location, benefits, and family size. Adding adjustments for age and gender is relatively straightforward, but in combination with location, this allows rates to vary by as much as 1,000% at the furthest extremes. Whether this step is warranted is primarily an issue of social justice, which this article does not explore.[4] But once age or gender adjustments are allowed, adding limited flexibility for health risk factors increases rate variation only a modest amount. Rating bands are controversial not so much because of their breadth (although, as subsequently explained, they are broader than they first appear), but because they allow overt medical underwriting and therefore keep risk selection as an explicit competitive variable. Accordingly, rating bands deserve further critique beyond their magnitude and complexity.

One benefit of allowing some health risk adjustment is simply that it helps to keep more insurers in the market, since insurers, especially commercial indemnity insurers, strongly oppose community rating. Indeed, community rating has caused large numbers of indemnity insurers to pull out of certain states, particularly in the individual market. However, we found in this study that indemnity insurers with significant market shares often were willing to remain in states even with pure community rating, as long as the insurers were assured of being able to increase average rates across the board as needed to stem mounting losses.[5] Nevertheless, some ability to vary rates among purchasers is critical to many indemnity insurers, especially smaller ones. We observed a number of examples where smaller indemnity insurers were leaving states in which they had market shares that to them were significant, primarily because of the states' move to adjusted or pure community rating (Hall 2000). These insurers explained that it is especially critical to be able to offer lower rates to new business than to renewing subscribers in order to avoid being locked out of the market with a bad pool of existing business. These insurers also were sensitive to restrictions that prevent higher rates for very small groups.

Additionally, indemnity insurers argue forcefully that pure or adjusted community rating does not allow them to compete fairly with HMOs. They contend, with some justification, that sicker people, on average, prefer indemnity and preferred provider organization (PPO) products over HMOs because these impose fewer restrictions on choice of physician and covered prescriptions. If indemnity insurers are allowed to reflect this increased health risk in only their *average* rates, then they cannot offer a price-competitive product to healthier subscribers. This has the potential, they claim, to make indemnity-based products affordable only for higher-risk people, and perhaps not affordable for anyone, even though some lower-risk people are willing to pay extra money for increased choice and coverage.

The extent to which these claims are true and, if so, whether this is due to adverse selection or to less use of cost controls by indemnity insurers, can be resolved only through more quantitative and direct empirical measures.[6] Also debatable is how serious a public policy concern it is to lose some smaller indemnity insurers, as long as some remain in the market. Based on our observations, PPO products are still widely available even in tightly regulated small-group markets. However, their prices are increasing significantly faster than HMO products, and pure indemnity products have all but disappeared from the more tightly regulated small-group and individual markets.

Some reformers seek to address these concerns through explicit risk adjustment mechanisms. However, it is plausible to view rating flexibility as an alternative to risk adjustment. Allowing insurers to reflect risk differences upfront in their rates avoids the need for risk adjustments done through regulatory controls behind the scenes. Although this has the obvious impact of requiring unhealthy purchasers to pay more, rating bands limit the extent of this burden at the same time that they allow some effi-

ciency-enhancing price variation to occur. While this approach is plagued with complexities of design and enforcement, so is regulatory risk adjustment, whose flaws and complexities are only dimly understood because its methods have not yet been fully developed and implemented.

Flexibility in Rating Bands

Reform laws that create rating bands typically restrict rating flexibility to a prescribed range above or below an "index rate," which is the midpoint between the highest and lowest rates. This appears to tightly compress rate variation, so that the highest risks pay only moderately more than average risks. Even the broadest range of ±35%, used in Ohio, on first glance appears to mean that high-risk groups never may be charged more than 35% above average, standard risks. This appearance is deceiving. As one Iowa agent noted, "When you do actuarial calculations, it just doesn't come out the same as when a human typically looks at something." To start, it is important to understand that this range is not fixed on any certain "standard" rate. Insurers are not required to set their standard rates in the middle, or to have their midpoint rates match any particular spot in the distribution of health risks. Instead, this range is defined solely by a comparison of the top and the bottom rates. It is completely up to the insurer where and how to distribute risks within this range.

To make maximum use of the allowable range, many actuaries told us that they rate standard groups near the bottom end of the premium range, not the middle. Actuaries said this gives them more room to increase rates for higher risks, either at the outset or through subsequent rate increases for groups whose risk status worsens over time. A ±35% band allows insurers to charge high risks as much as 108% more than low risks (1.35 vs. .65), and a ±25% band allows a 67% increase (1.25 vs. .75) (Curtis et al. 1999). If an insurer sets its standard rate at the lowest allowable level rather than in the middle of the allowable range, then it effectively can double the range that it has available to increase rates for higher risks. An Ohio regulator commented that this degree of rating flexibility is a "field day" for insurers. However, this practice is fully consistent with how the law is written.

The rating strategy of setting the "standard" rate at the low end of the rating band has an important drawback, however, that deters insurers from using this strategy to the fullest extent. Rating standard risks at the lowest tier leaves no room for insurers to give discounts to the best risks, so it is difficult for them to compete for this business. These insurers would be using all their rating flexibility to deter high risks and none to attract good risks. Thus, a number of insurers leave some room at the bottom of their bands to issue discounts, but typically this is only about 10 percentage points. This still leaves substantial room to increase rates for higher risks. Other insurers issue no discounts. Few insurers that we interviewed issue standard rates at the middle of the bands.

An Ohio agent explained how this degree of flexibility tends to undercut guaranteed issue: "In my opinion, all that [the rating reform] has created is a way for this not to be guaranteed issue. I mean, who's going to pay that kind of premium? . . . There's still a mechanism in place where, if they don't want a group, all they have to do is double the premium and nobody's going to take it." The same point was made in states where increases of only 50% or 67% were allowed. However, other agents explained that, for people with the worst health problems, price is no barrier; their need for insurance is so great, they will buy insurance at almost any price, if they have the money.

Even if rating flexibility does not deter the purchase of insurance, it still has real impact when some insurers use this flexibility aggressively and others do not. To avoid a high risk, all an insurer needs to do is to offer a slightly higher price than someone else. If all insurers used the same rating methods, this would mean, however, that the higher-priced insurer would lose lower-risk business as well. Still, if this insurer used the full width of an allowable rating band and other insurers did not, it could offer both lower prices for better risks and higher prices for worse risks, relative to its competitors. For instance, in Vermont, which has nearly pure community rating, commercial indemnity insurers are allowed to vary rates ±20% for demographic factors. Blue Cross and HMOs, however, are required to use pure community rating. As a result, some Vermont agents and insurers believe that commercial indemnity insurers are receiving a more favorable selection of risks.

One agent noted that some Vermont insurers are "very sophisticated and disciplined in their exploitation" of this rating flexibility. An actuary said these insurers are "speed boats" that are very good at "nipping off" the best risks, and there is "no limit to how clean they can pick" the market. When some

insurers can vary rates and others cannot, he said, making money through risk selection is "as easy as falling off a log." An industry analyst commented that Blue Cross is being "pillaged by the competition," and a Blue Cross subject commented that price differences of 20% to 40% are more than enough to discourage higher risks from purchasing health coverage, which means that insurers who use this rating flexibility have nearly as much underwriting impact as they did when they could decline applicants outright.

In other states, Blue Cross and HMO plans have avoided this adverse selection by intensifying their rating and underwriting practices. Several insurers said that, as the result of HIPAA's requirement that all small-group products be guaranteed issue, they moved away from pure or adjusted community rating and adopted underwriting practices that use the allowable rating flexibility much more aggressively. Others, however, noted these changes were not entirely due to HIPAA, but also resulted from an intensely competitive environment in which insurers were eager to find any possible price advantage. In either event, subjects in several states noted that underwriting and rating practices intensified following HIPAA. For instance, two Ohio insurers, one an HMO and the other an indemnity insurer, shifted in 1997 from adjusted community rating to a complex system of multiple rate tiers, designed to make full use of the ±35% rate band through detailed medical underwriting. The HMO was very concerned about preparing for the first time to conduct medical underwriting with this level of precision since it did not think it had the "savvy" its competitors had developed. The indemnity insurer, which specializes in life insurance, formerly kept a simplified underwriting and rating process since health insurance was not its main line of business. Its health underwriter complained that the company had to undertake a major investment at great cost and disruption in order to gain the rating sophistication required to continue selling health insurance in a guaranteed-issue environment.

Likewise, both Kaiser Permanente and Medical Mutual of Ohio, which sell through the Cleveland purchasing cooperative known as the Council of Smaller Enterprises (COSE), now use rating tiers spread across the full allowable range, rather than their past practice of adjusted or pure community rating. One subject explained this is a "very un-Kaiser like" practice, but it was necessary for Kaiser to avoid becoming the carrier "of last resort." Several agents noted the irony that when insurers adopted these new rating strategies in response to HIPAA, which was meant to promote portability, it became more difficult for higher-risk groups to switch insurers than previously. New insurers now can rate these groups at their highest tier, whereas existing insurers are constrained by a 15% limit on increases.

So, while rating bands impose significant constraints on how much insurers can raise rates for the highest-risk people, insurers still have considerable room to compete based on risk selection and rating strategies. Whether this is beneficial for consumers and for the market is explored subsequently. Here, it suffices simply to document that insurers and agents view rate variations as small as 10% or 20% to be critical to employers' decisions about whether or where to purchase insurance.

Durational Rating and Rating by Blocks

Rating bands and community rating are limits that operate only at a point in time. They restrict how much an insurer can vary its rates in a given period, but they set no limits on how much rates can rise from one period to the next. However, most states also limit how much an insurer can raise a particular purchaser's rates, relative to the insurer's average increases.[7] These limits are intended to prevent a practice known as "churning," which results from insurers giving steep discounts initially, but then increasing rates steeply at renewal, forcing subscribers to look for new coverage, even if claims did not exceed first-year estimates. Insurers have justified these increases by observing that claims costs tend to increase rapidly after the first year or two of coverage. This phenomenon, known as the "durational effect," occurs for several reasons. When medical underwriting is allowed, the advantage from careful underwriting wears off after a year or two as initially healthy groups or individuals regress to the mean. Also, claims costs increase simply because pre-existing condition exclusion periods expire. Limits on durational rating require insurers to anticipate these effects in their initial pricing, so that purchasers do not receive rate shocks one or two years after starting coverage.

States with rating bands typically limit annual rate increases to 15% over trend. Like the rating bands themselves, this limit can be deceptive. This 15% is allowed on top of any increases in the insurer's overall rates (the trend). It also is in addition to any increase resulting from the purchaser moving into a

higher age bracket. Therefore, agents report that it is still common for purchasers to receive increases of 30% or more.

Another way that the durational limit can be deceptive is if it is applied separately to different blocks of business. In the individual market, many states set a strict 0% limit on durational increases, meaning all renewals must receive the same rate increase, no matter how high their claims costs have been. However, states commonly apply this limit only within separate blocks of business, which are liberally defined. For this purpose, many states allow insurers to maintain different rating blocks *for each approved version* of their coverage contracts. This allows insurers to free themselves from durational limits simply by making minor changes in their benefits or contract language and filing new documents for approval.

This leads to a convoluted manner of doing business, which absent reform has become increasingly widespread but which is poorly understood outside the health insurance industry. An insurer will aggressively underwrite and favorably price a new policy to a select pool of purchasers. After a few years, as the durational effect takes its toll, the rates for this policy will become uncompetitive for new subscribers, so the insurer will close off that block of business (stop selling the policy), and will begin to market a slightly revised policy at lower rates to a new pool of freshly underwritten subscribers. Subscribers to the old contract are protected by guaranteed renewal and limits on renewal increases, but only relative to other holders of the same policy. An iron law of actuarial science, however, states that a closed pool will deteriorate rapidly, meaning that average claims will increase quickly, because the pool is not being "refreshed" with new, low-risk people. Closing a pool accelerates the durational effect because adverse selection starts to occur. Healthier subscribers have every reason to switch to a less expensive and freshly underwritten policy, either with the same insurer or with a new one, leaving behind older or sicker members. Thus, simply by making minor changes in their contract forms, insurers are able to maintain distinct blocks of business based on health risk, despite even the most stringent limits on price increases for renewals.

A number of states use one of two strategies to prevent this manipulation of rating rules. Some reforms require that all policies sold within a market segment (individual or small group) be treated as a single block. Others still allow separate blocks of business, but limit how many an insurer may have or how these blocks are defined; these reforms also limit the extent to which rates can vary among blocks. Such strategies appear to effectively prevent the type of block gerrymandering just described. However, these reforms have not eliminated entirely the durational effect that gave rise to these market dynamics.

Insurers expected HIPAA's guaranteed issue requirement to greatly reduce the durational effect, which was thought to arise mainly from careful medical underwriting. However, the durational effect appears to remain strong. One experienced actuary said this is the most surprising result of the small-group reforms. In this insurer's large block of small groups, new subscribers have 15% to 20% fewer claims than existing subscribers, an effect that largely disappears in a year despite the lack of any initial underwriting. Actuaries at other insurers report that their loss ratios for small groups that renew are substantially higher than for those whose policies lapse, an indication that the durational effect continues despite guaranteed issue.

This suggests that much of the "select" risk quality of new, small-group purchasers was due to generic pre-existing condition exclusions and to natural risk selection behavior, rather than to insurers' underwriting prowess in accurately identifying and excluding bad risks.[8] Pre-existing condition exclusion periods typically expire after one year, resulting in an increase in claims. The natural selection effect stems from the fact that people with health problems are averse to making any changes in their health insurance, a phenomenon that Altman, Cutler, and Zeckhauser (1998) describe as "adverse retention."[9] Since most people who buy insurance already have coverage elsewhere, people shopping for new coverage tend to be healthier than those who are not looking for new coverage.

The continuing strength of the durational effect is significant for insurers' pricing strategies because it means insurers improve their risk pools by attracting newer subscribers. This can intensify price competition, but it also can lead to market volatility through strategic "low balling or "buying market share." A number of those interviewed observed that the durational effect allows new market entrants to offer lower premiums initially; however, this advantage wears off rather rapidly, and so new insurers often find that they have to raise their rates steeply after just a year or two, especially if they underpriced initially relative to the risks they received. If so, their

enrollment will deteriorate rapidly, since existing healthy subscribers will leave and the insurer will not attract new enrollees because its new business rates must keep within 15% of the pace set by its renewal rate increases. As a result, one actuary explained, "one year's genius can be next year's bozo," since a great rate that attracts lots of new business will end up locking the insurer into a rate structure that cannot sustain increased claims as the durational effect takes its toll. We observed several instances in different states (Florida, New York, North Carolina, and Vermont) of small or new insurers aggressively gaining market share following insurance reforms but then quickly losing this business as their loss ratios soared and their attractive pricing disappeared.

There are differing views on whether the reform laws are to blame for this low balling and market volatility. Volatility is increased by portability provisions that allow groups to easily leave insurers when they receive steep rate increases. A number of insurers reported very high "lapse rates" of 30% to 40% among groups dropping coverage each year. In other respects, however, the small-group laws promote market and pricing stability. The fact that rating bands allow some rate flexibility means that existing insurers can respond to new entrants to some degree by offering preferred rates, so long as their normal pricing is in the middle of the bands (which may not be due to other strategic concerns mentioned earlier). New entrants, for their part, usually realize they have only a limited leeway to impose rate increases greater than the trend in new business rates, so they are reluctant to engage in "low-balling" strategies to gain an artificial advantage from new business.

Adjustment for Benefit Differences

The discussion so far has revealed how rating rules allow greater variation in rates than first appears permissible, without any violation of the rules. Next, we examine possible ways that rating factors can be misused or manipulated to allow even greater variation, but in ways that are not consistent with these rules. The first of these techniques arises from adjustments to reflect the value of different benefits or plan designs. Rating reforms allow insurers to vary their rates however much is actuarially justified by differences in benefits among their various plans. Valuing benefit differences is a matter that entails actuarial judgment, however, and for which there are different techniques. One technique is to declare that benefits are worth the claims costs they generate, so that different benefit packages are rated according to the claims experience for the entire pool of subscribers to each package. The difficulty with this approach is that it confounds benefit differences and health status factors. If some plans are more attractive to healthier or sicker populations, then the claims experience will reflect underlying health risk as well as benefit differences.

Using only claims experience can result in anomalies, such as placing a higher actuarial value on a benefit package that actually has less coverage. For instance, if Plan A and Plan B are identical except that Plan B offers free membership in a health club, Plan B should be more expensive; however, measured by claims experience, this difference likely will be muted, or reversed, since health-conscious subscribers likely will gravitate toward the free membership. In states that, prior to HIPAA, required only certain standardized benefit plans to be guaranteed issue, this actuarial technique was employed in a way that led to pricing these guaranteed-issue plans much higher than similar, medically underwritten plans. This occurred even though the guaranteed-issue plans had somewhat leaner benefits, since the guaranteed-issue plans had risk pools that generated considerably more claims than the plans with richer benefits.

Sometimes, the rate differences were dramatic. For instance, in North Carolina, a 1993 compliance audit found that several insurers were charging two or three times more for the leaner statutory plan than for their richer plans that had more favorable risks. Allowing insurers to price their different products according to the risks they attract, rather than according to an objective measure of the value of the benefits, obviously undermines the goal of rating reforms to spread risk across the entire market segment. This makes it much more difficult for higher risks to afford guaranteed-issue coverage. In North Carolina, guaranteed-issue plans accounted for only about 3% of the new small-group plans sold in the years prior to HIPAA.

A different sorting of risks occurs even when all plans are guaranteed issue, because higher-risk people then tend to seek out richer benefits. For instance, plans with lower deductibles or better drug benefits are likely to attract sicker patients who expect to use these benefits. If a risk-neutral actuarial valuation is used, this selection effect is ignored

in adjusting rates for differences in benefits. But if actual or projected claims experience is used to value these increased benefits, then insurers can use the benefits adjustment to reflect the full impact of the increased health risk of those who purchase these richer policies.

Viewed more broadly, the actuarial difference between two plans can reflect one or more of three distinct effects: 1) the benefits effect, which is simply the difference in how much of treatment costs are covered by insurance; 2) the "moral hazard" effect, which reflects behavioral changes in subscribers' decisions to seek treatment due to benefit differences; and 3) the "risk-selection" effect, which is due to subscribers with different health status selecting different plans because of what they cover. Thus, if two plans differ only in the amount of their deductibles, say $100 vs. $500, one naturally would expect the lower deductible plan to cost more. But how much more? The simple value of the benefit is worth somewhat less than $400, since not everyone will need to incur the full deductible each year. However, not having to pay as large a deductible encourages some people to seek more care, which increases overall utilization and treatment costs somewhat. Finally, and this is our present concern, more generous benefits attract subscribers who have a greater expectation of using those benefits. If the latter factor is reflected in premiums, then rates are being allowed to reflect health status, at least to some degree.

Through interviews and observation, we inquired into whether the latter is occurring under rating reforms. Some actuaries said they do not attempt to use this leeway in the rating rules, but set benefit factors according to national actuarial standards and data, which they apply equally to low- and high-risk pools. However, actuaries at other insurers candidly admitted that, where allowed to do so by regulators, they use benefit factors to include the full impact of risk-selection effects in rates.

We observed a number of examples where plans that were identical except for their deductibles had rates that differed considerably more than the deductible differential. In Vermont, we observed several instances in which insurers had two plans whose only difference was a $400 greater deductible, yet the cost of the lower deductible plan was at least $600 higher. In North Carolina, we observed plans with leaner benefits costing considerably *more* than those with richer benefits, an effect that is clearly due to risk selection since the leaner plans were sold only on a guaranteed-issue basis while the richer plans were subject to full medical underwriting.[10]

A few state regulators whom we interviewed appeared oblivious to this issue. Others were aware of the issue, but considered these rating tactics to be legitimate since claims experience is a relevant measure of the value of benefits. Some regulators require that benefit factors be based primarily on claims experience and do not permit departure from actual experience without special justification. For instance, in New York, rates for point-of-service (POS) products must be based on a claims experience pool separate from that for HMO or PPO products, even though this tends to result in rates that reflect the characteristics of the risk pools that are attracted to each product rather than the inherent value of these different product designs. Nevertheless, these regulators refused to allow the extremes observed in New Jersey and North Carolina. Benefit factors at least have to be pointed in a logical direction (higher rates associated with richer benefits) and they must remain within reasonable bounds (e.g., rates for lower deductibles cannot exceed the difference in the deductibles). Within these limits, however, different actuaries might use benefit adjustments of different magnitudes, in part because of the associated risk-selection effects.

Still other regulators attempt to avoid letting health risk factors influence benefit factors to any extent by focusing on this aspect of rating in the actuarial certifications that they require insurers to file. However, there is no established way to disentangle these combined factors, so, necessarily, these judgments are made more by intuition than by hard data. Typical actuarial certifications assert that "rate differences due to differences in plan design only reflect benefit differences," and "neither rates nor rating factors associated with the statutory standard and basic plans give recognition to the guaranteed-issue feature of those plans." Similarly, some state statutes require that "premiums for identical groups differ only by amounts attributable to plan design and [do] not reflect differences due to the nature of the groups assumed to select particular health benefit plans." Language like this prevents actuaries from explicitly including health risk factors in their benefit adjustments, but it still does not prevent them from using benefit adjustments as an indirect surrogate for health risk. This could be done only by scrutinizing the data and detailed assumptions that actuaries use to develop their rating structures, but regulators rarely do this. State regulators usually

accept on face value the general certifications of compliance just quoted and do not ask even for the size of the relevant factors. One regulator conceded the insurance department is "willing to take almost anything" based on the broad certification of general compliance. An actuary with one insurer explained that certification requires only that the insurer have policies and procedures designed to be in compliance, not that the actuary has conducted a systematic audit or is certain everything is in compliance.

The state in this study that came closest to separating health risk from benefit factors was Colorado. There we found (prior to HIPAA's requirement that all plans be guaranteed issue) that guaranteed-issue statutory plans were priced very competitively with insurers' more favored, medically underwritten products. This was accomplished by insisting that the statutory plans be the starting point for an insurer's entire small-group rating structure. Deviations from rates for the statutory plan are allowed based on the actuarial value of benefit differences in other plans. But because rates for statutory plans determine rates for all other plans, there is an inherent check against setting rates for statutory plans so high that only high-risk groups are willing to purchase them.

Age, Location, and Family Size Factors

Other rating factors have produced some controversy, though to a lesser degree, by even more subtle ways in which some insurers use them to their advantage. Agents and actuaries pointed to various techniques for loading the allowable rating factors in ways that produce higher rates for less attractive purchasers. For example, insurers view very small groups with five or fewer members as inherently higher risk than larger groups, because adverse selection is more severe for the smallest groups and because these groups lack economies of scale for administrative costs (Curtis et al. 1999). However, some states do not allow insurers to use a rating factor for group size. Therefore, we were told that some insurers raise rates for these groups by using different age and gender factors than actuarial data justify, based on calculations about the demographic composition of these less desirable groups.

Similarly, in community-rated states, insurers have some leeway to use geographic factors as a surrogate for health risk characteristics in different locations. To some extent, this is perfectly legitimate, but it also can lead to some degree of manipulation. Some states mandate uniform geographic boundaries for rating purposes, but others leave this for insurers to decide, prescribing only the size of the geographic units (county level, metropolitan area, or three-digit zip-code areas). Depending on the precise rules, insurers may be able to define geographic rating in a way that separates higher-risk from lower-risk populations. Even when areas are standardized, some insurers can be more aggressive than others in setting this rating factor to favor one area over another, based more on health status than on delivery system characteristics. One Florida agent explained:

> [Insurers] look at a county and they go, okay, that county is mostly blue collar or fishing industry . . . For example, [Outlying] is the county next to us. Their rates are 15% higher than [Center City] rates. And I've argued with carriers that it doesn't make sense because the groups in [Outlying] County are going to come to [Center City] for their care . . . [But the insurers are] looking at the county as the type of industries that are mainly in there and using those factors . . . [Center City] is considered a white-collar county . . . very little industry. Then you go look at [Outlying County], which is mainly a fishing industry, a lot of blue-collar workers there, and they're getting impacted [unfavorably].

Naturally, the same can be done more overtly through selective marketing. Insurers are not required to maintain equal presence throughout a state. And, although these practices can be criticized for their risk-selection purposes, they are legal and may reflect accurately the risk characteristics of different geographic locations.

Finally, several agents spoke about difficulties relating to family size. Some states mandate uniform family-size categories, while others leave this to insurers' discretion. Regardless, insurers have some discretion over the shape and magnitude of the family-size factor. As with geography and age, this allows insurers to devise the rating structures they believe are the best surrogates for health risk factors. Some insurers use several tiers for family size (single, spouse with no children, one child, more than one child, etc.), which makes the price relatively more attractive for smaller family units. Other insurers go in the opposite direction, sometimes using only two categories (single and dependents), with the result that coverage is much more expensive for single parents, for instance. Several agents noted that changes in family-size rating have caused employees to drop coverage for their dependents or to

Table 2. Summary of drawbacks and remedies for various rating reform components

Reform element	Intent	Drawback	Remedy
Community rating	Simplicity and fairness	Leads to adverse selection, reduced or distorted competition	Rating flexibility
Modified community rating	Allow rates to vary by age/gender	Demographic adjustments are large	Pure community rating
	Prohibit health status adjustments	Indemnity insurers say this prevents them from countering adverse selection from HMOs	Rate bands
Rate bands	Limit higher rates for those in worse health	Standard rates can be set at bottom of range	Decrease the allowable range; self-correcting since this limits room to compete for better risks
	Limit medical underwriting as a competitive variable	Small rate differences create large market advantages	Self-correcting: more insurers will make aggressive use of rate bands
Limits on renewal increases	Prevent "churning" based on "durational rating"	These limits apply only within blocks of business, which can be manipulated to segregate low vs. high risks	Impose limits on variation among blocks
		Favors new entrants, makes it more difficult for insurers to compete for new business	Self-correcting, since new entrants will feel durational effects quickly and be trapped in an unsustainable rate structure.
Adjustments for differences in benefits	Allow rates to reflect legitimate actuarial differences in benefits	Difficult to separate benefit from selection effects	Require more detailed actuarial justification
Age, location, and family-size adjustments	Permit legitimate factors that affect claims costs	Additional opportunities to circumvent by using allowable factors as proxies for limited or prohibited factors	Define allowable regions and demographic groupings; increase actuarial scrutiny

purchase dependent coverage in the individual market.

Implications for Crafting Reforms

Rating restrictions give rise to a plethora of potential techniques for manipulation, circumvention, and gaming. This may cause one to question the entire enterprise. However, some degree of complexity is unavoidable in a competitive insurance market. Most of these rating strategies are equally or even more prominent in an unregulated market. Therefore, this account is not intended to build a case against rating restrictions, or for any particular set of restrictions. Instead, it is meant to reveal the mechanics of these restrictions to help predict the likely results of particular regulatory strategies. Drawing the proper balance between rating flexibility and rating uniformity is one of the many dilemmas policymakers confront in crafting a workable set of reforms for the complex machinery of the private health insurance market.[11]

Where best to settle along the range from pure community rating to broad rating bands is a question that involves issues of social justice, and practicalities of market mechanics and regulatory oversight. This article focuses on the latter issues. Regardless of how much or little rating flexibility insurers are allowed to retain, careful attention must be paid to how various allowable rating factors might be used strategically to achieve risk selection objectives not intended by reforms. As summarized in Table 2, even factors as simple as location or age are subject to subtle forms of manipulation. Rating bands for health risk factors create more leeway than first appears. And adjustments for differences in benefits, which are necessary in any rating system, are difficult

to disentangle from the natural selection patterns that inevitably follow benefit differences. Careful regulators have learned to minimize these potential forms of circumvention, but doing so requires regulatory vigilance and expertise.

In conclusion, the complexity of rating reforms cannot be avoided. Some type of rating reform is essential if guaranteed issue is to have any meaning. Competitive insurance markets are inherently complex, so any effective rating reform also will be somewhat complex.[12] However, complex reforms require careful monitoring to eliminate possible avenues for circumvention. Some of this slippage is sufficiently minor or self-correcting as to be not worth the worry. But in other respects, rating rules can be used to undermine themselves or other reform components. Avoiding these larger-scale problems requires careful construction of the rules and diligent monitoring of their implementation. So far, not all states have lived up to this ideal.

Notes

Participating in this research were Elliot Wicks, Ph.D., Janice Lawlor, M.P.H., Allen Feezor, M.A., Mark Smith, Ph.D., and Robert Goodman, Ph.D. Although the analysis and conclusions are solely the author's and do not necessarily reflect the views of the Robert Wood Johnson Foundation or these colleagues, the author is deeply indebted to their support and assistance, which made this work possible.

1 A full description of this study and its findings on numerous topics can be found at http://www.phs.wfubmc.edu/insure.

2 For a more extensive explanation of the content of these reforms, see the study web site mentioned above, Hall (1992, 1994a), and Curtis et al. (1999).

3 In general, interview topics included: 1) the purposes and political history of the reform law; 2) successes and failures overall and for each reform component; 3) the impact on premiums, products, and competition; 4) circumvention by insurers, agents, or employers; and 5) enforcement and compliance problems. Interviews focused on small-group reforms, but also included individual-market reforms for relevant states.

4 See Hall (1994b), Stone (1993), Daniels (1990), Light (1992), and Epstein (1997).

5 For more elaboration and detailed examples, see Hall (2000) and the research reports posted at www.phs.wfubmc.edu/insure.

6 For additional discussion, see Pauly and Nicholson (1999).

7 Under pure community rating, all rates must be in lock step and renewal rates must equal rates for new business. Under adjusted community rating, renewal rates can increase only if the purchaser unit moves into a new age bracket. Under rating bands, however, the only inherent limit is whether the purchaser was previously near the top of the band. If not, then rating reforms prevent insurers from moving the purchaser up any more than 10% to 15% a year.

8 Another study found no difference in one large insurer's claims over six years between small groups that were underwritten and those that were guaranteed issue. See Glazner et al. (1995). Another reason the durational effect might continue to be strong following underwriting reforms is that the effect arises in part from preexisting condition exclusions, and, although insurance reforms purport to limit these exclusions, the one-year limit in HIPAA for small groups is no shorter than what was already standard in the industry.

9 Based on their analysis of a large block of government employees, Altman, Cutler, and Zeckhauser (1998) found that adverse retention accounts for approximately two-thirds of the total adverse selection effect—that is, selection effects due to losing existing low risks are twice as large as selection effects due to attracting new high risks.

10 Even more dramatic, in New Jersey (which was not included in this study), leading insurers in 1999 charged, on average, $3,312 more for single coverage, whose only difference from the lower-cost option was a $500 lower deductible. We were informed that this disparity was largely attributable to insurers' deliberate effort to move subscribers into higher deductible plans, but the disparity also reflects, to some extent, the higher risk status of those who continued to purchase the lower deductible plan.

11 For more elaboration, see Curtis et al. (1999).

12 This theme is developed in more depth in Hall (1998).

References

Altman, D., D. M. Cutler, and R. J. Zeckhauser. 1998. Adverse Selection and Adverse Retention. *AEA Papers and Proceedings* 88(2):122–126.

Atchinson, B. K., and D. M. Fox. 1997. The Politics of the Health Insurance Portability and Accountability Act. *Health Affairs* 16(3):146–150.

Bureau of National Affairs. 1997. *Health Law Reporter* (October 2):1529–1530.

Curtis, R., S. Lewis, K. Haugh, and R. Forland. 1999. Health Insurance Reform in the Small-Group Market. *Health Affairs* 18(3):151–160.

Daniels, N. 1990. Insurability and the HIV Epidemic: Ethical Issues in Underwriting. *Milbank Quarterly* 68(4):497–525.

Epstein, R. A. 1997. *Mortal Peril: Our Inalienable Right to Health Care?* Reading, Mass.: Addison-Wesley.

Glazner, J., W. R. Braithwaite, S. Hull, and D. C. Lezotte. 1995. The Questionable Value of Medical Screening in

the Small-Group Health Insurance Market. *Health Affairs* 14(2):224–234.

Hall, M. A. 1992. Reforming the Health Insurance Market for Small Businesses. *New England Journal of Medicine* 326(8):565–570.

———. 1994a. *Reforming Private Health Insurance.* Washington, D.C.: AEI Press.

———. 1994b. *Is Community Rating Essential to Managed Competition?* Washington, D.C.: AEI Press.

———. 1998. Public Choice and Private Insurance: The Case of Small Group Market Reforms. *University of Illinois Law Review* 1998(3):757–775.

———. 2000. The Impact of Health Insurance Market Reforms on Market Competition. *American Journal of Managed Care* 6:57–68.

Jensen, G. A., and M. A. Morrisey. 1999. Small Group Reform and Insurance Provision by Small Firms, 1989–1995. *Inquiry* 36:176–187.

Kuttner, R. 1997. The Kassebaum-Kennedy Bill: The Limits of Incrementalism. *New England Journal of Medicine* 337(1):64–67.

Light, D. W. 1992. The Practice and Ethics of Risk-Rated Health Insurance. *Journal of the American Medical Association* 267(18):2503–2508.

Nichols, L. M., and L. J. Blumberg. 1998. A Different Kind of "New Federalism"? The Health Insurance Portability and Accountability Act of 1996. *Health Affairs* 17(3):25–42.

Pauly, M., and S. Nicholson. 1999. Adverse Consequences of Adverse Selection. *Journal of Health Politics, Policy and Law* 24(5):921–930.

Stone, D. 1993. The Struggle for the Soul of Health Insurance. *Journal of Health Politics, Policy and Law* 18(2): 286–317.

U.S. General Accounting Office (GAO). 1998. *Health Insurance Standards: New Federal Law Creates Challenges for Consumers, Insurers, Regulators.* Washington, D.C.: GAO.