



Lessons from a Legislative Disaster

PRACTICE AREAS

Tax

Rosina B. Barker

Tax Notes

May 14, 1990

LENGTH: 10627 words

DEPARTMENT: Special Reports (SPR)

CITE: 47 Tax Notes 843

HEADLINE: 47 Tax Notes 843 - LESSONS FROM A LEGISLATIVE DISASTER.

AUTHOR: Barker, Rosina B.

Tax Analysts

TEXT:

[*843]

Rosina Barker has been on the Tax Staff of the Committee on Ways and Means, U.S. House of Representatives, since 1985, where she worked, among other issues, on section 89. Her views do not necessarily represent those of any member of the Committee. Ms. Barker is grateful to Daniel I. Halperin and Harry Conaway for extensive comments they made on an earlier draft of this article.

In this article, Ms. Barker examines what led to the repeal of section 89 of the Internal Revenue Code. Section 89 provided coverage rules for health care and other employer-provided fringe benefits. While observing that any curtailment of a large tax subsidy will be unpopular, Ms. Barker agrees with the critics of section 89 that it was excessively complex and generally unworkable. She argues that four basic features of its design underlie the problems of section 89: (1) testing of coverage received, rather than made available; (2) exclusive reliance



| 2 | Lessons from a Legislative Disaster

on mechanical rules; (3) too many goals that could be better met in other ways; and (4) an inappropriate sanction for noncompliance. Because of these fundamental flaws in design, the problems of section 89 could not be solved either by technical corrections, such as those enacted in 1988, or by regulatory efforts. When Congress returns to the issue of fringe benefit nondiscrimination rules, any test should take into account what was learned from the experience of section 89.

Copyright 1990, Rosina B. Barker

Table of Contents

- I. Introduction 843
- II. Structure, Purpose, and History 844
 - A. Statutory Structure 844
 - B. Statutory Purpose 845
 - C. Initial Business Response, TAMRA Modifications, and Repeal 845
- III. Fundamental Flaws 846
 - A. Test of Actual Coverage Received 846
 - B. Mechanical Rules 848
 - C. Too Many Goals 850
 - D. Inappropriate Sanctions 851
- IV. Conclusions 851

I. INTRODUCTION

Late on the night of Tuesday, November 7, 1989, the House and Senate passed a bill to increase the borrowing authority of the Federal government and forestall impending default on Federal obligations. Attached to this must-pass legislation was another urgent piece of legislative business: the repeal of nondiscrimination rules for employer-provided health care plans (and other fringe benefits). Enacted in 1986, these rules expressed Congress' judgment that the tax subsidy for employer-provided fringe benefits was warranted only for plans providing significant coverage to rank and file employees. By the end of 1989, these rules, contained in section 89 of the Internal Revenue Code,^{1/} enjoyed wide-spread notoriety in the pages of the business press, and massive organized opposition by employers. Its few supporters confined their defense to a stripped- down provision that, in the heated political climate surrounding section



| 3 | Lessons from a Legislative Disaster

89, proved an irrelevant and inadequate substitute for repeal.

Neither the goals of section 89, nor its development, explain the controversy it aroused. The tax expenditure involved was large, by 1989 totaling \$ 32 billion for employer-provided health care alone.^{/2/} Ensuring that a subsidy of this magnitude be directed at low-paid employees seems potentially unobjectionable. Indeed, this goal has long been embodied in the coverage and nondiscrimination rules governing employer-provided pension plans. The rules themselves were developed on a bipartisan basis over a period of a year and a half, evolving from their initial form [*844] in then-President Reagan's tax reform proposals to the Congress,^{/3/} through successive versions in each of the House-and Senate-passed tax reform bills,^{/4/} to a final, overhauled version in the Tax Reform Act of 1986. At each step in the process, there were hearings and extensive staff discussions with employers and their representatives. Modifications enacted in 1988 further reflected recommendations by employers. What then, led to the total repudiation of these rules

Arguably, the most fundamental cause of the opposition to section 89 -- a cause this paper will not explore -- is that it restricted access to a very large tax expenditure. In the case of health care, it provided that a \$ 32 billion tax expenditure was available only if coverage was provided to employees not typically covered, and who many employees firmly believed as a matter of policy should not be covered: part-time workers, 'leased' employees, and low wage, high turn-over employees. The provision applied in equal measure to large employers and to small employers, with their typically less extensive health coverage. It is possible that no structure, no matter how well designed, could have withstood the political pressure inevitably arising from a measure with such dramatic economic effects.

Nonetheless, the purpose of this paper is not to explore the politics of the curtailing of large government subsidy. Rather, this paper takes as a given that section 89 was indeed excessively burdensome. Specifically, its structure was complex, placing the statute beyond the comprehension of all but the most sophisticated consultants and lawyers. It required data gathering that was costly, time consuming and intrusive. It required excessively complicated testing to be applied to this data. And its complexity so obscured its purpose that it could command little political support by those who might otherwise defend its objectives.

This paper argues that specific basic features of section 89 led to these results. Because of certain fundamental elements in the design of the statute, the problems of section 89 could not be solved, and in some ways were made worse, by tinkering efforts such as those enacted in 1988. In particular, this paper argues that four basic features of section 89 made it ultimately unworkable and politically indefensible:



| 4 | Lessons from a Legislative Disaster

(1) The rules required testing of actual health coverage received, rather than merely coverage available. While the tax policy objectives underlying nondiscrimination rules have traditionally been satisfied with coverage-received tests, there is strong reason to conclude that, at least in the area of health care, appropriate rules must necessarily be confined to availability tests only.

(2) The rules relied almost exclusively on mechanical, or numerical ratio, tests. When applied to the unique features of health benefits, these led to an inevitable, and eventually unacceptable, trade-off between rigidity and complexity;

(3) The rules tried to accomplish too many goals -- not merely an excessively large number of goals, but goals that were better accomplished in other ways. Specifically, limitations on tax-free health benefits, or caps, and minimum benefit specifications would have more simply fulfilled much of the purpose of section 89. Instead of recognizing and deferring to political constraints on caps and minimum benefits, the rules tried to accomplish them indirectly through a set of tests that only became overburdened with the task, and ultimately collapsed under their weight.

(4) The rules employed an inappropriate sanction. If plans were discriminatory, only the coverage of highly compensated employees was taxable. Of this coverage, only the 'discriminatory excess' was subject to tax -- that is, the coverage in excess of what would be received under a nondiscriminatory set of plans. While adopted to make the rules politically more palatable, these refinements increased the operational complexity of the rules, and solidified the perception that their function was to limit benefits for high-



| 5 | Lessons from a Legislative Disaster

paid employees, rather than extend them to the rank and file. Their net effect thus may have been to further weaken support for the rules.

This paper concludes with a discussion of the implications of its analysis for future efforts to design health care nondiscrimination rules. Because health care benefits are growing in importance, this issue will certainly arise again, so this discussion is not moot. This paper confines its discussion of section 89 and its conclusions to employer provided health care benefits. Other benefits were involved, but health care was by far the most significant in dollar size and complexity.

II. STRUCTURE, PURPOSE, AND HISTORY

A. Statutory Structure

Section 89 was enacted as part of the Tax Reform Act of 1986. It provided a uniform nondiscrimination test for two fringe benefits: health care and group term life insurance. In addition, it could be elected for all other tax-subsidized fringe benefits: dependent care, group term legal assistance, and educational assistance. Under section 89, health cover-age received by highly compensated employees was excludable from taxable income only to the extent it did not exceed what would be received under a nondiscriminatory employer- provided health program.^{5/} [*845]

Under section 89, benefits were nondiscriminatory if they passed either a four-part test, or an alternative two-part test. These tests were (with one exception) 'mechanical tests' -- that is, they set numerical ratios which were required to be met when the value of plans provided to nonhighly compensated employees was compared with that provided to highly compensated employees.^{6/} This contrasted with an alternative approach available in nondiscrimination statutory schemes before 1986: nonmechanical tests providing general goals. For example, under section 410(b)(1)(B) (as in effect before the Tax Reform Act of 1986), a pension plan could be qualified if it provided benefits to a nondiscriminatory classification of employees. Treasury regulations provided alternative mechanical tests fleshing out the meaning of this more general statutory prescription.

As originally enacted, the four-part and alternative two-part tests found in section 89 were as follows:

1. Four-part test. The four-part test consisted of a three-part eligibility test and a one-part benefits test.
 - a. Three-part eligibility test.
 - i. 90/50 test: Under this test, 90 percent of an employer's employees must be eligible to participate in a health plan at least 50 percent as valuable as the best health plan available to any highly compensated employee.^{7/}
 - ii. Fifty percent test (executive-only plan test):



| 6 | Lessons from a Legislative Disaster

Under this test, each plan must be available to a group comprised of at least 50 percent nonhighly compensated employees.^{8/}

An alternative way of satisfying this test was available if the plan was available to a group in which the ratio of highly compensated employees to nonhighly compensated employees did not exceed the ratio of all highly compensated to all nonhighly compensated employees of the employer.^{9/}

iii. Nondiscriminatory provision test: Each plan must

be nondiscriminatory as to terms and conditions. This is the only non-numerical test, and was designed to capture non-quantifiable discrimination that could not be detected in a numerical test.^{10/}

b. Benefits test.

Any health plan would meet the benefits test if on average, coverage provided to nonhighly compensated employees under all health plans of the employer equals 75 percent of the average value of coverage provided under all health plans to the employer's highly compensated employees.^{11/}

2. Alternative two-part test. A plan would meet the alternative two-part test if:

(a) Eighty percent of the employer's employees actually receive benefits under the plan; and

(b) the plan is nondiscriminatory as to terms and conditions with respect to eligibility to participate (as provided under section 89(d)(1)(C)).^{12/}

B. Statutory Purpose

In its essential structure, section 89 was logical and its purpose explicable. Under the basic four-part test, coverage had to be widely available; 90 percent of the employer's employees had to be eligible to receive at least some significant coverage (no less valuable than 50 percent of the best coverage available to any highly compensated employee).

Executive-only plans such as executive physical examinations were denied tax-exempt status under the 50 percent test.

Coverage had to be not merely available; it had to be attractive and affordable enough so that enough nonhighly compensated employees would elect to buy it to allow the program to pass the 75 percent benefits test.

The tests were designed to offer significant flexibility to accommodate the many different kinds of health programs available. Two prongs of the basic four-part test -- the 75 percent test and the 90/50 test -- were applied to the employer's entire health program, rather than to each separate plan. This allowed many different kinds of plans to be offered to different segments of the employer's workforce, bound only by the constraint that together they meet the specified numerical tests. The 50 percent test was applied on a plan-by-plan basis, as was the alternative 80 percent benefits test. However, the statute provided aggregation rules so that a plan that did not meet these tests when considered alone could be aggregated with a plan worth no less than 95 percent of the first plan and the two thus tested



| 7 | Lessons from a Legislative Disaster

together./13/

C. Initial Business Response, TAMRA Modification, and Repeal

During the initial development of section 89, rules were added to make the tests substantively less difficult to pass.

These included rules allowing separate testing by different lines of business;/14/ provisions allowing for separate testing of employees with families and employees [*846] without;/15/ and a rule permitting the disregard of employees with employer-provided coverage elsewhere (e.g., from a spouse or parent)/.16/

After its enactment, the mechanical tests of section 89 immediately led to criticisms from those employers who began to prepare for compliance with the rules. (Because the rules did not become effective until 1989, employers had two years to prepare compliance.) Employers charged that the rules were complex, required excessive data gathering, and were excessively rigid given the multiplicity of existing health plan structures. These complaints led to the enactment of lengthy design modification in the Technical and Miscellaneous Revenue Act of 1988 (TAMRA)./17/

Changes enacted under TAMRA further eased the testing burden, most significantly by reducing the frequency of required testing. In addition, TAMRA increased the flexibility of the statute by providing additional aggregation rules and alternative tests. However, the cost of these improvements was a significant increase in the length, difficulty, and complexity of the statute.

Business opposition, especially by smaller businesses and their representatives in 1989, led to pressure for total re-peal of section 89. Rep. Dan Rostenkowski, D-Ill., Chairman of the Committee on Ways and Means, responded with several simplified stripped-down versions of nondiscrimination rules./18/ However, these efforts failed to assuage what had become full-scale demands for repeal. On November 7, 1989, President Bush signed the debt ceiling legislation sent by Congress that also provided for full repeal of section 89, and reinstatement of prior law./19/

III. FUNDAMENTAL FLAWS

The inadequacy of the TAMRA modifications lay not in the bill, but in that it made marginal changes to a structure that was intrinsically ill-suited to the task of providing a framework for testing health benefit plans. These structural flaws made the statute, no matter how refined, unable to respond to the tremendous variety and complexity already existing among employer-provided health benefits. These flaws were fourfold. First, the rules tested for actual health coverage received by employees, rather than that made available. Second, it relied almost exclusively on mechanical rules. Third, it attempted to meet too many goals that would have been more suitably met with other statutory mechanisms. Fourth, it applied an inappropriate sanction.

This paper examines each of these problems in turn.

A. Test of Actual Coverage Received



| 8 | Lessons from a Legislative Disaster

The basic four-part test included a requirement that, on average, the value of health coverage received by nonhighly compensated employees be no less than 75 percent of the value of the coverage received on average by highly compensated employees.

Actual coverage tests have historically been a part of fringe benefit nondiscrimination rules. They reflect two basic tenets of conventional tax policy: equity and efficiency. Under the principle of vertical equity, it is unjust that high-income individuals receive significantly more generous tax subsidies than low-income individuals. The principle of horizontal equity implies that a tax subsidy is not warranted unless it is widespread among similarly situated individuals. The principle of efficiency implies that the best use of a tax subsidy to promote health care, or the consumption of any other socially valued good, is to direct the subsidy to those who are least able to afford it, rather than those who are likely to buy it without the aid of the subsidy.

A slightly different statement of these goals, based on the fact that under nondiscrimination rules the fairness of a tax benefit is judged by its allocation among employees of one employer, rather than across society, is this: the purpose of providing a tax subsidy for the fringe benefits of highly paid employees is to encourage employers to establish and support benefit arrangements which benefit low-paid employees as well. If this incentive to disseminate benefits does not work, the tax subsidy is not justified.

Despite historical reliance on coverage tests, certain features of health benefits make them exceedingly difficult to subject to actual coverage tests, and the structure of section 89 reflected this difficulty in its earliest enacted version. At the outset data collection to test benefits under the 75 percent test did not threaten insoluble problems. Enough data were required to determine two denominators (total number of highly compensated employees, and total number of nonhighly compensated employees) and two numerators (the value of coverage received by highly compensated employees, and by nonhighly compensated employees).

As with pension benefits, to calculate the numerators was merely a matter of identifying the number of covered employees, the compensation of each, and the value of coverage received by each. Generally, this information [*847] was obtainable from employer records. Problems that arose with valuation, and frequency and timing of testing, were addressed in TAMRA.

The most significant difficulty of collecting data for actual coverage testing arose from cultural perceptions concerning what constitutes an appropriate health benefit. These generally shared beliefs have been discussed little, and analyzed even less, but proved to have a powerful effect on the content of the rules.



| 9 | Lessons from a Legislative Disaster

If an employee has a family, there is a widely shared consensus that his appropriate health benefit includes coverage for family members. This is significantly more expensive (between two and four times) than single coverage. If the employee has no family, the consensus holds that his appropriate health benefit is less. If the employee (or family) is covered under the plan of another employer, for example, through a spouse or parent, the appropriate benefit is deemed zero.

This perception distinguishes health benefits from pension benefits. An employee's pension benefit is generally considered to provide income replacement. Its appropriateness is judged in relation to salary, rather than, as in the case of health benefits, a host of factors specific to the employee's family. If he has a working spouse with a large retirement benefit, there is no policy consensus that his pension should therefore be reduced. If his spouse has no pension, his own retirement benefit is not considered to be an appropriate candidate for enlargement.

The shared attitude towards health benefit needs is grafted onto two circumstances frequently asserted by employers: Highly compensated employees have families to a greater extent than do nonhighly compensated employees. In addition, they are less likely to have coverage under the plan of another employer through a spouse or parent than are nonhighly compensated employees./20/

If true, these two circumstances would mean that even if all employees with identical health needs, as the term is understood, receive identical coverage, the employer's plan may fail any unadjusted test of actual coverage received. This is because, to a greater or lesser extent, highly compensated employees will be more likely to receive costly family coverage than nonhighly compensated employees, and will be less likely to waive coverage because of coverage under another plan.

A strict construction of the purpose of the actual coverage test would imply that if low-paid employees tend not to have families and thus do not require family coverage, the tax benefit for family coverage for high-paid employees is not warranted. Similarly, if low-paid employees tend more than high paid to have coverage from another employer, a less generous tax subsidy might be appropriate.

However, section 89 did not incorporate such a strict philosophical position. Instead, it adopted as a given the notion that appropriate health benefits differ among individuals in different situations. Accordingly, section 89 included a rule that permitted an employer to test employees with families separately from those without. Thus, the coverage received by highly compensated employees with families could be tested against that received by nonhighly compensated employees with families. In addition, the rules permitted employers to disregard altogether those employees covered by the plan of another employer through a spouse or parent./21/



| 10 | Lessons from a Legislative Disaster

While representing a politically necessary and philosophically defensible outcome, separate testing led to data collection problems that were severe and ultimately politically indefensible. To test separately employees with families, and to disregard employees with other coverage, employers first had to identify them. Some kind of questionnaire procedure had to be developed for every employee, whether or not covered by a health plan. To prevent abuse in the questionnaire procedure, section 89 initially required that employers collect the data annually, in the form of a sworn statement from every employee, on a form prepared by the Internal Revenue Service.^{22/}

These procedures caused an early firestorm of protest. To comply with these rules, employers were required to prepare a new data collection procedure. Employers immediately complained that efforts to collect responses from every employee were costly, especially because of the statutorily prescribed collection methods. Employers in addition charged that the questions were intrusive, requiring questions about sensitive family information that in many cases employees would resist answering. And finally, they pointed out that the information collected was inherently biased toward overstating discrimination in favor of highly paid employees. In answering questions about other coverage, employees suspicious about their employer's possible intent to curtail coverage may have an incentive to deny or under-state the amount of other coverage provided. This could lead to overstatement of the health coverage needs of nonhighly compensated employees.

In TAMRA, the data collection requirement was modified. Instead of collecting annual data on an IRS form from every employee, employers were permitted to collect sworn statements every three years, on forms that met IRS guide-lines, and to use random sampling.^{23/} [*848]

It is likely that the random sampling and other simplifications would have reduced the cost of data collection for very large employers. But it is unclear that the random sampling ever would have responded to the problems of employers too small to employ random sampling, and too large to keep track of the data by hand. In addition, the measures in large part failed to address the problems of intrusiveness and inherent bias in these proceedings.

The sworn statement rules underscore another, more fundamental problem. The specification of random sampling in the statute goes well beyond the specificity normally employed in statutory schemes. The rules suggested future vulnerability. If they did not work adequately, their minute detail precluded much regulatory flexibility to develop alternative testing methods. Future modifications were almost necessarily left up to legislation. This process is inherently slow, and its ability to respond to multiple problems compounded by small staff size. Yet the need for data collection, and concern for its accuracy, drove the statute to this result.



| 11 | Lessons from a Legislative Disaster

The tangled statutory scheme, and its minimal effect in addressing the fundamental problem raised by the data collection requirement, strongly points to a conclusion: Actual benefits testing may be inappropriate in the case of employer-provided health benefits. It is necessarily the case that many employees will turn down family coverage, or in-deed any coverage at all, for the simple reason that they do not need it, either because they have no family, or they are covered by another plan. To separate out these employees, data collection is necessary. And this data collection will run into the problems encountered under section 89.

This conclusion is a sharp departure from the tradition of subjecting pensions and other fringe benefits to actual coverage tests. But pensions are distinguished from health benefits in the important way already discussed: the appropriate size of an employee's pension benefit is determined almost exclusively by compensation. This information is by and large readily available to the employer. To determine an employee's appropriate health benefit requires information that is intrinsically difficult to collect.

If actual benefits testing is not possible for employer-provided health coverage, testing must be confined to the availability of health coverage. Under a set of rules governing availability, employers could test for compliance looking only at the design of their health plans. This raises problems of its own. What should be available, and to whom? What constitutes truly available health coverage? Nonetheless, in light of the history of section 89, it seems necessary to conclude that, with the possible exception of small plans provided to a small concentrated group of employees, benefits testing of large health programs will impose costs that strain the political tolerability of the rules.

B. Mechanical Rules

The second structural weakness of section 89 was its almost total reliance on mechanical rules. Mechanical rules employing numerical ratios were not new in the employee benefits field. Section 410(b)(1)(A) (as in effect before the Tax Reform Act of 1986) contained an elective mechanical test. The alternative reasonable classification test of section 410(b)(1)(B) was fleshed out with examples in the regulation essentially providing mechanical tests. What was new in section 89 was a complete and exclusive statutory prescription of numerical standards for both availability and benefits tests.

Not only was the approach more radical than earlier numerical schemes, it was imposed on a world with a tremendously varied and complex array of health programs. The inherent rigidity of the rules, combined with the enormous variety of health care programs, led to immediate difficulties.



| 12 | Lessons from a Legislative Disaster

1. Cliff effects and resulting complexities. The significant features of statutory mechanical rules is their 'cliff' effect. That is, a health program that is very near compliance nevertheless faces total failure if it falls below the prescribed numerical boundary. This cliff feature places enormous strains on the political acceptability of such rules, as employers present policymakers with plans that by some standard of reasonableness appear fair, but just barely fail, or plans that pass the tests under one set of circumstances but may fail if small changes in the work force occur. Thus, Congress was faced with an alternative: require employers with reasonable plans to undergo costly massive restructuring to meet the rules, or make the rules flexible enough to accommodate a greater variety of plans. By necessity, Congress chose the latter course. But to make mechanical rules more flexible, without abandoning the scheme of mechanical rules, requires more rules. These took several forms. First, for purposes of the 75 percent benefits test, employers were allowed to treat only the discriminatory excess of highly compensated employees -- that is, that portion of the benefit that exceeded what could be received under a nondiscriminatory plan -- as includable in income. This excess inclusion feature minimized the 'cliff' effect of failing the 75 percent benefits test; the more massively the average benefits of highly compensated employees exceeded the 133 percent boundary, the more income inclusion resulted. A limited version of this approach was adopted for failures of the 90/50 test.

A second solution to the inflexibility problem was the adoption of mechanical aggregation rules, whereby different nonidentical plans could be tested as one. This approach was adopted immediately; for purposes of the 50 percent availability test and the 80 percent alternative test, a plan could be aggregated with any plan worth no less than 95 percent of its value.²⁴ This rule was liberalized [²⁵849] in TAMRA, and three comparability or aggregation rules were added.²⁵

A third solution was to provide alternative ways of passing the rules. The 80 percent test was provided by the original statute as an alternative, simpler way of passing section 89. The aggregation rules, by expanding the number of plan configurations that would pass the tests, could also be viewed as a development of alternative standards.²⁶

While adding some flexibility to the rules, these mechanical refinements had two overwhelming effects. First, they added immensely to the complexity and difficulty of the rules. Second, the number of alternative ways of passing the tests, and the availability of alternative aggregation rules to do so, resulted in employers expending significant resources trying to figure out plan design changes that would minimize income inclusion for highly compensated employees, or minimize the expansion of benefits for nonhighly compensated employees. Thus, even though added for the convenience of employers, the greater flexibility had the perceived (and real) effect of increasing the cost of benefit plan design and testing.



| 13 | Lessons from a Legislative Disaster

The proliferation of mechanical tests, as refined by successive legislative efforts, increased the complexity of the rules, and the cost of plan design and testing. But they may not have adequately responded to the inflexibility inherent in mechanical tests. This conclusion becomes more apparent upon examination of the examples in the legislative history to TAMRA. Some of these examples are of such specificity that it is doubtful that they could be useful to more than a handful of taxpayers.²⁷ That is, given the tremendous number of health plans, there may be no politically logical end to the number of mechanical refinements and alternative rules that are necessary in the statute to accommodate those that seem intuitively reasonable. This raises the question of whether exclusive mechanical rules, at least those applicable to an entire health program, are appropriate in a statute, rather than in regulations. The administrative process may be intrinsically better suited to deal with the specific features of individual health programs than is the legislative process. Given the complexity of health plan designs, and given the reluctance of Congress to remove the tax benefits of pro-grams of long standing, it may be that sweeping mechanical rules are better left to the regulatory process.

2. Work force count problems. A second, related problem flows from the use of mechanical tests. The need to calculate a ratio requires determination of a numerator and a denominator. Generally, the denominator is the employer's entire work force, less statutorily excluded categories of employees: part-time and seasonal workers, collectively bar-gained employees, and individuals who perform services but are not 'employees' for purposes of the rules.

Clearly, the more noncovered employees that may be excluded from the denominator, the more easily a mechanical test can be met. This simple arithmetic fact posed political and technical problems. Who should be included in the de-nominator reflects a policy judgment about who deserves coverage. Employers objected most immediately and most vehemently to the inclusion of part-time employees -- defined as individuals who worked less than 17-1/2 hours per week,²⁸ and 'leased' employees. Leased employees are individuals who may or may not be common law employees, but who provide services not uncommonly performed by employees for a minimum number of hours per year.²⁹ Both categories of workers are typically excluded from employer- provided health care plans.

The requirement that an exact denominator be determined implied that a failure to count leased and part-time employees accurately might cause inadvertent failure and disqualification of the plan. The consequent need to identify such workers with precision added significantly to the data collection problems raised by section 89. Employers expressed concern about their ability to count hours worked by certain categories of casual workers so that they could identify all leased employees and all employees exceeding the 17-1/2 hour threshold. TAMRA included changes to make counting hours easier. For example, employers were permitted to look at scheduled hours, rather than actual hours worked, for an employee's first year. But complaints still remained in the case of [*850] counting hours worked by contract workers, counting hours by employees compensated on piece-rate systems, and keeping track of overtime.



| 14 | Lessons from a Legislative Disaster

The technical difficulties posed by the need to count leased and part-time employees only increased the severe political strain imposed by requiring coverage of them in the first place. Substantive coverage rules for both categories were significantly curtailed in the section 89 alternative reported by the Senate Finance Committee (but ultimately not enacted).³⁰

C. Too Many Goals

The third vulnerable feature of section 89 was the sheer number of tests it required, at least in its basic, four-part form. The number of tests had real costs. First, it contributed to the cost of the data gathering, since employers had to collect data in several different forms to satisfy all the tests. Second, the number of tests added to the complexity of the statute. The complexity, in turn, contributed to the political unacceptability of the statute by making its basic purpose harder to fathom. Even to many of its potential supporters, it looked more like an arbitrary morass than the articulation of a sensible legislative goal.

The number of tests flowed directly from the number of objectives legislators sought to meet through the statute. The simplest objective was ensured by the 75 percent test: Under this test, on average, the tax-favored coverage received by high-paid employees had to be roughly commensurate with the tax-favored coverage received by low-paid employees. But the concerns of section 89 reached beyond the comparability of the aggregate health benefits of high- and low-paid employees. Under a simple averaging test (such as the 75 percent test) it would be possible for a very small number of highly compensated employees to receive a grossly disproportionate benefit (provided the number was small enough not to cause the average benefits of high-paid employees to flunk the test). Conversely, under a straight averaging test it is possible to exclude a significant portion of the work force from coverage. (This can be accomplished by excluding some low-paid employees, and giving the rest a sufficiently generous benefit to meet the averaging test. Or it can be accomplished by excluding comparable fractions of high- and low-paid employees from coverage.) section 89 foreclosed this possibility by additional rules: the 50 percent test and the 90/50 test. With these tests, section 89 went beyond comparing average health benefits, and towards setting maximum and minimum permitted health benefit availability standards.

The 90/50 test, for example, was designed to satisfy two alternative objectives. On the one hand, it could be seen as ensuring that low-paid employees have access to health care that met a certain minimum standard. On the other hand, it could be seen as ensuring that if any high-paid employees had available a health benefit significantly better than those available to low-paid employees, the high-paid benefit for the highly paid employee was not tax subsidized.



| 15 | Lessons from a Legislative Disaster

The 50 percent test may be viewed as supplementing the objective underlying the 90/50 test by ensuring that plans covering primarily high paid employees ('executive only plans'), which are presumably more generous than plans covering the entire work force, are not tax subsidized.

When viewed in conjunction with these other tests, the 75 percent benefits test may be seen as another component of an availability test, rather than a separate kind of test. This test ensured that plans were not merely nominally available. That is, in order to pass the 75 percent test, benefits had to be attractive enough that highly compensated employees actually elected them. For example, to be sufficiently attractive, a plan had to have an affordable employee premium and an appropriate benefits structure.

When all these objectives are laid out together, it is apparent that at least a few of them could be met more simply in other ways. The goal of preventing disproportionate tax subsidies for a small number of highly paid employees could be more directly met by capping the permitted amount of excludable benefit. In conjunction with a cap, an average benefits test alone might have seemed more satisfactory. The goal of insuring widespread availability of health care that is adequate and affordable could be more easily met by plan design specifications. For example, these specifications might require the availability of a plan with a specified maximum employee premium, and a specified minimum employer premium.

However, both these mechanisms are controversial, at the very least. Health care caps are now considered almost certainly politically unacceptable. Included in President Reagan's tax reform proposals to Congress, they were vigorously opposed by organized labor, and were not included in the version of tax reform reported by the Committee on Ways and Means. Minimum plan design specifications have hardly reached even the level of congressional debate. [*851]

Notwithstanding the controversial nature of caps and minimum benefit standards, the experience of section 89 suggests that trying to approximate their results through nondiscrimination rules leads to excessive complexity and obscurity.

D. Inappropriate Sanctions

The fourth structural flaw of section 89 was the nature of the sanction imposed on plans that failed to meet the non-discrimination tests. Historically, the sanction for discriminatory employee benefits plans has been disqualification. When a plan is disqualified, the total benefit of every participant becomes includable in income.

Section 89 departed from this tradition in two respects. First, only highly compensated employees were subject to income inclusion. Second, only the 'discriminatory excess' coverage was includable. That is, only the value of the coverage received in excess of the maximum coverage that could be received under a nondiscriminatory arrangement was included in income. The number of highly compensated employees affected by income inclusion was further limited



| 16 | Lessons from a Legislative Disaster

by a 'ratcheting' formula that provided for successive inclusion of the coverage of each highly paid employee, starting with the employee with the most valuable coverage. This process was reiterated until the remaining nonincludable benefits of all employees taken together met the tests.

The sanction was designed to be philosophically sensible, as well as to make the rules more politically palatable. By limiting income inclusion to the discriminatory excess, the rules related the size of the sanction directly to the degree of discrimination, and thus, like the Lord High Executioner, made the punishment fit the crime. It was hoped that this limitation would minimize the 'cliff' problem arising in the case of plans which only narrowly failed the tests. It was further hoped that the rules would be more acceptable if the sanction were imposed only on highly compensated employees -- for most employers, a relatively small fraction of the work force.

While the sanction was designed to make the rules more palatable, in some respects it exaggerated those very features that made them most unacceptable. First, limiting the sanction to highly paid employees meant that the penalties for discrimination would, for most employers, affect only a small fraction of their work force. Hence, many employers pointed out it was far cheaper to respond to the rules not by making coverage less discriminatory, but rather by maintaining discriminatory coverage, and then 'grossing up' the salary of those highly paid employees affected by income inclusion. This approach was made even cheaper by the discriminatory excess formula, which in many cases would have resulted only in partial inclusion.

It is not clear how many employers planned to follow a 'gross up' response rather than expand coverage, as the rules were repealed before they ever went fully into effect. However, the perception certainly grew that the principal result of section 89 would not be to expand coverage, but rather to tax the health coverage of high- paid employees. In short, section 89 began to look like a complicated version of what Congress had already rejected: a limitation on the amount of excludable health benefits, or cap.

This perception further weakened the political credibility of the rules. Members of Congress might have been willing to defend a set of rules that demonstrably increased health coverage among rank and file employees. They were much less supportive of a set of rules that, according to what they were being told, merely put a tax on the health benefits of upper-level management.

IV. CONCLUSIONS



| 17 | Lessons from a Legislative Disaster

Employer-provided health care continues to grow in cost and importance. In particular, employer-provided retiree health care is attracting increased attention by employers and legislators for a number of reasons. Its costs are skyrocketing; recent changes in accounting procedures require that the obligations be reflected on corporate balance sheets, and an important Medicare program for the coverage of catastrophic medical expenses was repealed last year.

Thus, despite the debacle of section 89, the issue of nondiscrimination rules for health care plans will not go away. If Congress enacts new tax-favored vehicles for health coverage such as prefunded retiree medical care, the issue will reemerge in full force. What are the lessons that can be learned from the section 89 experience in designing health care nondiscrimination rules in the future?

First, it is likely that any comprehensive set of rules must be based on the availability of health coverage, rather than actual coverage received. This conclusion follows from the necessity under a coverage test to identify employees with families, and employees covered by plans of other employers. This task appears to be so difficult that coverage rules are not a practical possibility.

If a decision has been made to test the availability of health coverage, rather than coverage received, a second issue emerges. What does it mean for coverage to be available? Is an HMO located in a wealthy suburb truly 'available' to residents of faraway localities, even if nominally available? Is expensive coverage with high employee premiums truly 'available' to low-paid individuals who might willingly pay for coverage with a lower employee cost?

Of course, all these issues could be addressed with mechanical rules, specifying, for example, the minimum percentage of employees to whom coverage must be available and maximum employee premiums. This possibility raises a third issue: What is the appropriate role of mechanical rules in an availability test?

The experience of section 89 suggests that exclusive reliance on any set of mechanical rules will cause problems. For starters, mechanical design specifications [*852] would certainly lead to the disqualification of plans that appear intuitively reasonable. For example, a mechanical availability test would almost certainly specify a maximum employee premium, so as to ensure that the plan was truly available to rank and file employees who might be unwilling to pay high premiums. Yet if an employer whose plans did not meet this standard could show that in fact all employees participated, it is unlikely the legislative process could firmly abide by a rule that caused the disqualification of the arrangement. Thus, mechanical availability rules would almost certainly lead to alternative, mechanical coverage rules.

Given the complexity of the health care world, it is likely that alternative tests, aggregation rules, and other refinements would inevitably occur, and would encounter the same twin problems of complexity and inadequacy. For example, if an HMO located in the CEO's hometown suburb is considered not 'available' to nonresidents, it would nevertheless be



| 18 | Lessons from a Legislative Disaster

unreasonable to conclude its benefits were inherently discriminatory if a comparable health arrangement or HMO were available elsewhere to other employees. Thus, comparability rules would be required.

In addition, a rule that did not rely exclusively on mechanical rules would instead provide general standards. More detailed guidance would be provided by Treasury. This alternative raises problems of its own. Mechanical rules at least have the advantage of being fair. That is, all plans are judged by the same standard. Even if the standard becomes more byzantine as it evolves through the legislative process, it applies to all alike. If more general standards are used, the administrative process is more flexible, but correspondingly less fair. At its worst, the flexibility of the administrative process allows the boundaries of the rules to be set by the most aggressive private practitioners.

Given the problems raised by abandoning mechanical standards altogether, a more promising approach might be to use mechanical rules in a more limited fashion than used in section 89. One example of this kind of approach is shown in H.R. 1864, the first alternative introduced by Chairman Rostenkowski. This bill required an employer to offer to 90 percent of the employer's nonhighly compensated employees a plan with an employee premium not exceeding a \$ 520 per year (\$ 1300 for family coverage) Under the bill, the employer could offer other plans, provided that at least one plan met the 90 percent availability and \$ 520/\$ 1300 employee premium standards. In order to ensure that the plan was of more than de minimis value, the bill restricted the maximum exclusion for any highly compensated employee to a multiple of the most valuable plan meeting the 90 percent availability and maximum employee premium requirements.

Had this bill been enacted in 1986, would it have avoided the problems encountered by section 89? The answer is not clear. On the one hand, by using an availability test, the bill avoided the testing and data collection problems of section 89. On the other hand, by using a mechanical rule, it raised at least some of the difficulties of mechanical rules; specifically, the problems of work force count and rigidity. In addition, like section 89, the bill would have taxed only high-paid employees on the discriminatory excess of any coverage. Many employers would undoubtedly have responded to this trade-off just as they did with section 89 -- by permitting income inclusion for high-paid employees. What is unclear is whether Congress would have tolerated this result had it been arrived at with simple rules, instead of complicated ones.

Another vulnerability of this approach was its reliance on plan design specifications. While this approach is intrinsic to a mechanical availability standard, to many observers it looked like the first step towards required minimum benefits.

Other approaches using a modified form of mechanical rules are possible, and have not yet been tried. For example, it is possible that mechanical rules might work as safe harbors to more general rules. The advantage of this approach is that employers with complicated health arrangements would not be forced to redesign them, but could make a showing to



| 19 | Lessons from a Legislative Disaster

the Treasury Department that the plans are nondiscriminatory. And the mechanical rules may serve as a benchmark guiding against excessive administrative discretion. The hazard of this approach is that the safe harbor may be viewed as the substantive rule, and subject to the same pressures as was section 89. Another, more limited possible use of mechanical tests would be to disqualify highly discriminatory 'executive only' plans, rather than more broad based arrangements.

In conclusion, when Congress is willing to return to the issue of health care nondiscrimination rules, the experience of section 89 can offer some useful lessons on what form the rules should take. This paper argues that four features of section 89, when combined, together made the statute unworkable: coverage tests; exclusive reliance on mechanical rules; too many goals that could be better met in other ways; and an inappropriate sanction. This it not to say, however, that some features of section 89 might not be appropriate if used on a more limited fashion. For example, mechanical rules might play an important role in an availability test.

FOOTNOTES

/1/ I.R.C. sec. 89 (originally enacted in the Tax Reform Act of 1986, P.L. 99-514, sec. 1151; amended by the Technical and Miscellaneous Revenue Act of 1988, P.L. 100-647, sec. 1011(B); repealed by P.L. 101-140, sec. 202 (1989)).

/2/ Staff of the Joint Committee on Taxation, Estimates of Federal Tax Expenditures for Fiscal Years 1990-1994, JCS-4-89, February 28, 1989, at 12.

/3/ The President's Tax Proposal to the Congress for Fairness, Growth and Simplicity, May 1985, at 33-35.

/4/ H.R. 3838, 99th Congress, 1st session (1985), and Senate Amendment to H.R. 3838, 99th Congress, 2d Session (1986).

/5/ I.R.C. sec. 89(a)(1) (as in effect before repeal).

/6/ Under I.R.C. sec. 414(q) (as added by P.L. 99-514, sec. 1114(a) and amended by P.L. 100-647, sec. 1011) a 'highly compensated employee' is defined generally as any employee who is a five percent owner of the employer; received compensation in excess of \$ 75,000; compensation in excess of \$ 50,000 if the employer is among the top 20 percent of the employers' most highly compensated employees; or was at any time an officer of the employer and received compensation in excess of 50 percent of the defined benefit plan dollar limit under section 415(b)(1)(A) for the year.

/7/ I.R.C. sec. 89(d)(1)(A) (as in effect before repeal).

/8/ I.R.C. sec. 89(d)(1)(B) (as in effect before repeal).

/10/ I.R.C. sec. 89(d)(1)(C) (as in effect before repeal).



| 20 | Lessons from a Legislative Disaster

/11/ I.R.C. sec.89(e) (as in effect before repeal).

/12/ I.R.C. sec. 89(f) (as in effect before repeal).

/13/ I.R.C. sec. 89(g)(1)(A) and (B) as enacted by P.L. 99-514, sec. 1151(a). The application of section 89(g) was modified by P.L. 100-647, sec. 3021(a)(6) to provide that, for purposes of the 80 percent test: (1) '90 percent' would be substituted for '95 percent'; and (2) an employer could elect to aggregate a plan with any other plan valued at 80 percent of that plan, provided that total coverage under the aggregated plans equaled 90 percent of the employer's nonhighly compensated employees.

/14/ I.R.C. sec. 89(g)(4) (as in effect before repeal).

/15/ I.R.C. sec. 89(g)(2)(A)(ii) (as in effect before repeal).

/16/ Under I.R.C. sec. 89(g)(2)(A)(ii) as enacted by P.L. 99- 514, sec. 1151(a), the rule permitting disregard of any employees covered by the health plan of another employer was applied only for purposes of the 75 percent benefits test of section 89(e). P.L. 100- 547, sec. 3021(a)(7)(A)(i) and (ii) extended the disregard rule to the 80 percent test of section 89(f), but only in the case of an employer who made coverage available to 80 percent of the employer's nonhighly compensated employees, without regard to the provision permitting disregard of employees with other coverage. Also as a result of P.L. 100-647, the employer could elect to use a 90 percent availability rule instead of an 80 percent availability rule if the employer also elected the 80 percent aggregation rule of I.R.C. section 89(g)(1)(D)(ii)(III). (See footnote 12, above).

/17/ Technical and Miscellaneous Revenue Act of 1988, P.L. 100- 647, sec. 1011B.

/18/ Rep. Rostenkowski's first effort to reform section 89 was contained in H.R. 1864, 101st Congress, 1st Session (1989). A second version was included in H.R. 3150, section 11331, as introduced by Rep. Rostenkowski, August 4, 1989. A third and final version was included in H.R. 3299, sec. 11331, as reported by the Committee on the Budget, September 20, 1989.

/19/ P.L. 101-140.

/20/ There is no data supporting or refuting either of these contentions.



| 21 | Lessons from a Legislative Disaster

/21/ I.R.C. sec. 89(g)(2)(A)(i) and (ii). In early draft versions of the nondiscrimination rules, the 75 percent benefits test was designed specifically to respond to the claim that high-paid employees have families, and need family cover-age, more than do low- paid employees. That is, the 75 percent test was not initially conceived as a legislative endorsement of providing low-paid employees with 25 percent less valuable coverage than high-paid employees, but as a rough response to the family coverage problem. As the rules evolved, however, the 75 percent test became perceived as embodying the allowable margin of discrimination between the coverage of high- and low-paid employees. Employers requested, and received, a separate testing rule.

/22/ I.R.C. sec. 89(g)(2)(B) before amendment by P.L. 100-647 sec. 1011B(a)(4).

/23/ P.L. 100-647, sec. 1011B(a)(4). Sampling under this provision could be conducted only by an independent third party, and was required to produce a 95 percent level of confidence that the margin of error not be greater than three percent.

/24/ I.R.C. sec. 89(g)(1)(A) and (B) before amendment by P.L. 100-647, sec. 1011B(a).

/25/ P.L. 100-647, sec. 1011B(a)(2), and Conference Report to Accompany H.R. 4333, H. Rep. No. 1104, 100th Congress, 2d Session (1988) at 80.

/26/ For example, the 80 percent test could initially be passed by aggregating plans within 95 percent of one an-other's value. Under TAMRA modifications, employers could elect to pass an 80 percent coverage test by aggregating plans within 90 percent of one another's value; or an alternative 90 percent coverage test by aggregating plans within 80 percent of one another's value. In addition, a fourth test was created by the addition of the rule allowing employers to disregard employees with other coverage for purposes of the 80 percent test; to make use of this rule, employers had to make coverage available to 80 percent of their employees (not excluding those with other coverage), essentially resulting in a three-pronged test (80 percent availability/nondiscriminatory availability/80 percent coverage). For employers electing to make use of the 80 percent aggregation rule, another three-pronged test was available (90 percent availability/nondiscriminatory availability/90 percent coverage).

/27/ For example, see safe harbor 3(c) under the comparability rules enumerated at Technical and Miscellaneous Revenue Act of 1988, Conference Report to Accompany H.R. 4333, Report 100-1104, 100th Congress, Second Session, October 21, 1988, Volume II, pages 40-41.

/28/ I.R.C. sec. 89(g)(1)(B) (as in effect before repeal).



| 22 | Lessons from a Legislative Disaster

/29/ I.R.C. sec. 414(a)(3)(C) (as in effect before modification by P.L. 100-140).

/30/ The problems of obtaining an accurate workforce count, of course, arise for any fringe benefit rule subject to mechanical tests. But it is probably the case that the workforce count problem is more significant in the case of health benefits than pension benefits, for two reasons. First, in the specific context of section 89, the problem was superimposed on a rule requiring availability to almost all of the employer's workforce -- 90 percent. Failure to offer coverage to 90 percent resulted in total disqualification and income inclusion for all highly compensated employees in the plan, so the severity of the test was combined with a steep 'cliff.' The combination only increased employers' anxiety about the difficulty of obtaining an accurate workforce count. Second, health benefits are more expensive than pension benefits for high turnover, part-time marginal workers. In the case of pension benefits, five-year vesting and typical pension accrual formulas keep the cost of early year accruals fairly low. By contrast, a health benefit may be a significant percentage of the compensation of a part-time or low-wage worker. Therefore, the cost of extending coverage to ensure passage of coverage rules may be more significant in the case of health than pension benefits.

***** End of Document *****

10741D

***** Print Completed *****

Time of Request: October 17, 2005 04:46 PM EDT

Print Number: 1822:66189535

Number of Lines: 552

Number of Pages: