A Study of Schedule Rating.

BY

ALBERT W. WHITNEY.

The National Council on Workmen's Compensation Insurance has undertaken a revision of the schedule-rating system. Although the present schedule is unsatisfactory, there is no such pressing need of a new schedule that the work can not be carried on deliberately and thoughtfully and thoroughly; unless a schedule can be evolved which shall show marked improvement over the present schedule, which shall bear indubitable signs of being structurally right, which upon that right structure shall carry a content that is based upon fact rather than mere judgment, there will be no use of making a change, for every change is disturbing and to that extent to be avoided.

The present schedule is not based upon a fundamental analysis: it therefore is lacking in adaptability and flexibility; it is not based upon statistical facts: quantitatively it is therefore presumably not correct. It is more valuable for its accident prevention effects than as a measure of the hazard. It should be possible, however, to create a schedule that will also be a reasonably correct measure of the hazard.

The making of a schedule involves three fairly distinct processes: first the creation, or more properly the discovery, of the proper structure, second the determination of the particular elements (accident causes) which should enter into the schedule, and third the determination of the weights of these different causes. The first process is primarily actuarial, the second primarily engineering, and the third primarily statistical.

The actual work of revision grew out of a joint meeting of the Actuarial and Engineering Committees; the first stage of the process was intrusted to a subcommittee consisting of Messrs. Mowbray (chairman), Newell, Paine, Perkins, Wheeler and Whitney. This committee has developed a schedule-structure which bears the earmarks of being correct. To what extent, however, it will prove to be amenable to actual application still remains to be seen, for only the first stage of the process has been completed; the engineers have scarcely had their hands upon the problem and the statisticians not at all.

There is, however, much reason to feel encouraged. For one thing the present structure has been the result of the convergence of several independent lines of thought, one of which was that of Mr. Downey, the chairman of the last Schedule Revision Committee, who at the end of the last revision foreshadowed the new developments to a considerable extent. Furthermore, the new structure has successfully stood all the tests that can be applied short of actual use.

At the request of Mr. Mowbray I am describing the result of the work of the committee up to date; this is not a report of the committee, but only a personal view of how the matter stands. It is hardly necessary to say that the comparatively direct way in which the results are here derived was not the way in which the results were originally secured, for the committee did much wandering before it got upon the right track. I may express the satisfaction, and surprise indeed, of the committee itself that the problem has seemed to yield so well to mathematical analysis.

The committee has made use of the following notation:

Let N be the number of employees in the standard (average) risk of the classification in question.

Let N' be the number of employees in the particular (in general non-standard) risk of the class.

In general the unprimed letters refer to the standard (average) risk of the class and the corresponding primed letters to the particular (in general non-standard) risk of the class. Briefly the unprimed letters may be said to refer to class and the primed letters to risk.

Let all accidents be separated into classes according to cause, the causes being numbered and indicated by subscript.

Let A_i equal the number of employees injured through cause i.

Let D_i equal the number of danger-points associated with cause i; as there are within the same cause different types of danger-points, this involves the conception of a standard danger-point for each cause and the reduction by a system of weights of other types of danger-points to this basis.

Let N_i equal the average number of employees exposed per danger-point associated with cause i.

Then N_iD_i will equal the number of employee danger-points—that is, the number of "exposures." An exposure is a double-ended entity, an employee at one end and a danger-point at the other; it is fundamental in this analysis, the quantities N_i and D_i not occurring separately, but only in this combination.

Let E_i equal the number of "careless exposures"—that is, the number of exposures in which the employees are careless. This presupposes a standard degree of carelessness, and that the employees are separable into two classes, the perfectly careful and standard careless. This is an actuarial fiction; it is, however, the commuted expression of actual facts—that is, for actuarial purposes the effect of a certain distribution of varying degrees of carelessness is the same as a certain separation into standard careless and perfectly careful. It may also be explained that the term "careless" is used only in a suggestive sense. We shall desire in practice to include in this category other "personnel" qualities having a bearing upon the frequency or seriousness of accidents. The "careless exposure" is conceived of as consisting of a standard danger-point on the one hand and a standard "careless" employee on the other, more correctly an employee who is standard in his susceptibility to accident. In this category we propose to include such elements affecting personnel susceptibility as light, sanitation, safety organization, first aid, hospital, use of goggles and approved clothing, personnel work, education, etc. The effect of such factors as first aid, hospital, etc., which in reality reduce the seriousness of accidents (and therefore K_i, later to be defined) rather than their actual number, is conceived of as commuted into an equivalent effect, expressed in ϵ_i (defined in the next paragraph) in reducing the number of accidents.

Let ϵ_i (the coefficient of susceptibility) equal E_i/N_iD_i.

Let p_i (a probability) equal A_i/E_i, or the proportion of careless exposures which result in injury.

Let π equal the pure premium for the class; then π' will be the pure premium for the risk.

Let π_i equal the part of π that is allocable to cause i, then

$$\pi = \sum_{i=1}^{n} \pi_i$$

and

$$\pi' = \sum_{i=1}^{n} \pi_i',$$

where n is the total number of causes.

Let W equal the average annual wages.

Let $K_i W$ equal the average cost of each accident.

Then:

$$\pi_i = \frac{\text{Losses due to cause } i}{\text{Total payroll}} = \frac{A_i K_i W}{NW} = \frac{A_i K_i}{N} = \frac{\epsilon_i N_i D_i p_i K_i}{N}. \quad (1)$$

Similarly

$$\pi_i' = \frac{\epsilon_i' N_i' D_i' p_i' K_i'}{N'};$$

and

$$\pi_i' = \frac{N}{N'} \cdot \frac{\epsilon_i' N_i' D_i' p_i' K_i'}{\epsilon_i N_i D_i p_i K_i} \pi_i. \quad (2)$$

$$\therefore \pi' = \frac{N}{N'} \sum_{i=1}^{n} \frac{\epsilon_i' N_i' D_i' p_i' K_i'}{\epsilon_i N_i D_i p_i K_i} \pi_i \quad (3)$$

Now K_i is the average percentage cost per accident caused by a standard danger-point to a standard careless employee (that is, an employee of a standard susceptibility to accident) in a risk in a given classification. It may therefore be assumed constant from risk to risk, therefore $K_i' = K_i$.

Similarly p_i' is the proportion of exposures of a standard careless employee to a standard danger-point in a risk of a given classification which result in injury. It also may be assumed constant from risk to risk, therefore $p_i' = p_i$.

Making these two simplifications, we have:

$$\pi_i' = \frac{N \epsilon_i' N_i' D_i'}{N' \epsilon_i N_i D_i} \pi_i = \frac{N}{N'} \frac{E_i'}{E_i} \pi_i ,$$

and

$$\pi' = \frac{N}{N'} \sum_{i=1}^{n} \frac{\epsilon_i' N_i' D_i'}{\epsilon_i N_i D_i} \pi_i = \frac{N}{N'} \sum_{i=1}^{n} \frac{E_i'}{E_1} \pi_i . \quad (4)$$

Now, let us suppose that the causes can be separated into two groups, first a group of what we may call major, or better schedule-ratable causes, and second a group of minor or non-schedule-ratable causes. The causes belonging to the first group we may number 1, 2, . . ., m; the causes belonging to the second group we may number $m+1$, $m+2$, . . ., n.

Making this separation, we shall have:

$$\pi' = \frac{N}{N'} \sum_{i=1}^{n} \frac{\epsilon_i' N_i' D_i'}{\epsilon_i N_i D_i} \pi_i + \frac{N}{N'} \sum_{i=m+1}^{n} \frac{\epsilon_i' N_i' D_i'}{\epsilon_i N_i D_i} \pi_i .$$

The non-schedule-ratable causes will in theory be causes the hazards of which will not vary from risk to risk—that is, causes for which $N_i'=N_i$ and $D_i'/D_i=N'/N$, the latter condition expressing the fact that the relative frequency of danger-points in risk and class will be the same as the corresponding relative number of employees.

We may now make the further assumption that ϵ_i'/ϵ_i is constant for all causes (with one exception, to be discussed later), say ϵ'/ϵ. This is not a violent assumption. The personnel conditions are peculiarly a matter of management, and the policy of the management may be generally assumed to extend throughout the plant and to affect all causes in practically the same way.

When these substitutions are made our expression for π' becomes

$$\pi' = \frac{\epsilon'}{\epsilon}\left(\frac{N}{N'}\sum_{i=1}^{n}\frac{N_i'D_i'}{N_iD_i}\,\pi_i + \sum_{i=m+1}^{n}\pi_i\right)$$

Letting the part of the pure premium that is allocable to minor causes be R, that is

$$R = \sum_{i=m+1}^{n}\pi_i\,,$$

we have

$$\pi' = \frac{\epsilon'}{\epsilon}\left(R + \frac{N}{N'}\sum_{i=1}^{m}\frac{N_i'D_i'}{N_iD_i}\,\pi_i\right). \tag{5}$$

I may now make some comments upon the schedule-structure which this formula represents. It will be noticed, in the first place, that it is not of the additive type (that is, in the manner of Taylor's series) by which the adjusted rate is expressed as the basic rate plus or minus certain increments (although it might readily be reduced to that form); in a general way it can be said that the rate is built up by adding together the parts of the pure premium for the average risk of the class that are allocable to the various causes, each affected by a factor expressing the relativity between the risk in question and the average risk as respects both number of exposures and degree of susceptibility to accident.

These factors, for the schedule-ratable causes, are of the form

$$\frac{N}{N'}\frac{\epsilon_{i_i}}{\epsilon_i}\frac{N_i'D_i'}{N_iD_i}\,.$$

N/N' takes care of the effect upon the pure premium of size of risk; whether this shall be measured by a comparison of the actual number of employees or by a comparison of payrolls is a practical consideration which need not be discussed here.

The remainder of the factor, which is in reality E_i'/E_i or the ratio of careless exposures in risk and class, breaks up into two factors, ϵ_i'/ϵ_i and $N_i'D_i'/N_iD_i$; the first takes account of all "personnel" elements of the exposure, the second takes account of the physical elements of the exposure.

It is significant that these two elements are related to each other multiplicatively. It is, in fact, clear intuitively that the effect of carelessness is not additive; a careless employee is a greater hazard in proportion to the greater physical hazard of the risk.

So much for the schedule-ratable causes. The residue R in theory should be made up of hazards that do not vary from risk to risk. There are probably none such, although there are doubtless hazards that vary little from risk to risk. In practice, however, R will doubtless have to be made up of those hazards which, while collectively substantial, are individually so small as to be impracticable to measure by actual inspection. The presence of R gives the formula a most valuable flexibility: if a very simple schedule is desired, only the most important causes will be thrown into the schedule-ratable class and the balance of the causes will contribute to R; in that case R will be relatively large; on the other hand, if an elaborate schedule is desired, a large number of causes will be thrown into the schedule-ratable group and will be the subject of inspection; in that case R will be small.

It should be noted that no limitations have been placed upon the nature of the hazard. The formula is therefore general and should include the several types of hazard that are to be found. The treatment of these types will be differentiated by assumptions with regard to the N_i's and D_i's. There are several of these types of hazard. There is, in the first place, the catastrophe hazard. This we may assume affects all employees alike; therefore $N_i' = N'$ and $N_i = N$. The personnel items with the exception of first aid and hospital will probably affect this hazard in a minor degree, and therefore for this hazard the personnel factor will have an individual value less than ϵ'/ϵ, which we may represent by ϵ_c'/ϵ_c. The charge for catastrophe, representing the catastrophe cause by the subscript c, reduces therefore to

$$\frac{\epsilon_c'}{\epsilon_c}\frac{D_c'}{D_c}\pi_c \tag{6}$$

D_c'/D_c is a factor representing the relative probability of catastrophe for risk and class.

The case of machines is too complicated to discuss in detail. We may assume, however, for illustrative purposes, an ideally simple case of an industry in which the machines are thoroughly standardized both as to type and relative number. In that case it will not be too violent an assumption to make $N_i'=N_i$; this will be strictly true if, in addition to the standardized conditions mentioned above, the machines are fully manned. In that case the inner multiplier reduces to D_i'/D_i and is obtained by forming the ratio of the actual number of danger-points revealed by inspection in the risk as compared with the corresponding number in the average risk of the class. It is assumed that in counting danger-points they will first be reduced to a common basis by a system of weights.

Another type of hazard is that of stairs and elevators. Under the present schedule two elevators, each of which had a certain defect, would receive twice the debit that one such elevator would receive, and still more absurdly two elevators that had a certain good characteristic upon which a credit was due would receive twice as great a credit as one such elevator. In theory, therefore, the rate would be reduced to zero by putting in a sufficient number of superior elevators.

Two elevators may produce twice the hazard of one, but that question can not be decided until we know how many people use the elevators. Suppose the following case: a factory has only one elevator. This is insufficient; it is overcrowded. A second is added; it relieves the congestion, and in that respect the hazard is actually reduced. However, in general the hazard to each employee remains the same, for he makes the same number of trips as before. Suppose the number of employees using the elevator to be 200. After the second elevator is put in the number using each elevator will be 100. The number of danger-points will be doubled, but the number of persons exposed per danger-point will be only half as great, so that the number of exposures, $N_i'D_i'$, will remain unchanged. The installation of the second elevator will not affect the hazard so far as riding employees are concerned. It has, of course, doubled the hazard, so far as the operatives are concerned, and probably doubled the danger of falling down shafts; all three of these hazards are, however, covered by the general formula when properly used.

These examples illustrate the flexibility of the schedule and the fact that each type of hazard must be given individual consideration.

Innumerable questions make their appearance when the schedule formula is applied to the various causes. I shall not undertake to discuss these with the exception of one very fundamental question. The question is briefly this: In counting danger-points shall we confine ourselves to bad conditions or shall we take account of the hazard that is inherent in even good conditions? The problem is practical rather than theoretical, for there can be no doubt that the theory of the schedule contemplates the latter procedure. We are dealing in our formula with absolute hazards, and if a hazard exists, even associated with a so-called good condition, we must take it into account in making an enumeration of the danger-points. This, however, is not done in the present schedule and there are certain practical difficulties in the way of carrying out such a procedure.

The especial province of the schedule is to carry the classification process beyond the manual. This it does by an analysis of the hazards that are actually to be found in industrial processes. A schedule that does not deal with absolute hazards can not only not produce right results, but it will obviously fail to penetrate into this territory.

The acceptance of the principle of absolute hazard means in practice that a schedule must express the hazard not only of sub-standard conditions, but of standard and super-standard conditions as well; the schedule must take account of the hazard not only of unguarded machines, but of guarded machines.

The carrying out in concrete practice of the principle of absolute hazard unquestionably involves very great difficulties, particularly of a statistical nature. How successfully they can be overcome it is too early to foretell.

A mathematical analysis can be made of the general statistical problem in the following way: Reverting to the formula (4), we have

$$\pi' = \frac{N}{N'} \sum_{i=1}^{n} \frac{E_i'}{E_i} \pi_i .$$

Let $\pi_i/\pi = h_i$, then $\pi_i = \pi h_i$ and

$$\pi' = \frac{N}{N'}\,\pi \sum_{i=1}^{n} \frac{E_i'}{E_i}\,h_i .$$

Multiply both sides of this equation by the payroll P' $(=N'W')$ and write similar equations for the risks 1, 2, . . ., s, indicating the number of the risk by the number of primes.

Then

$$\pi' N' W' = N\pi W' \left(\frac{E_1'}{E_1}\,h_1 + \frac{E_2'}{E_2}\,h_2 + \cdots \text{to } n \text{ terms}\right)$$

$$\pi'' N'' W'' = N\pi W'' \left(\frac{E_1''}{E_1}\,h_1 + \frac{E_2''}{E_2}\,h_2 + \cdots \text{to } n \text{ terms}\right)$$

$$\pi^{(s)} N^{(s)} W^{(s)} = N\pi W^{(s)} \left(\frac{E_1^{(s)}}{E_1}\,h_1 + \frac{E_2^{(s)}}{E_2}\,h_2 + \cdots \text{to } n \text{ terms}\right). \quad (7)$$

Adding, using for convenience the first term as type, we have:

$$\sum_1^s \pi' N' W'$$
$$= N\pi\left(\frac{h_1}{E_1} \sum_1^s E_1' W' + \frac{h_2}{E_2} \sum_1^s E_2' W' + \cdots \text{to } n \text{ terms}\right). \quad (8)$$

If we assume that the rates for these risks have been right in the aggregate, the term on the left will give the amount of the losses, but since

$$\pi = \frac{\text{losses}}{\text{payroll}},$$

we shall have

$$\sum_1^s \pi' N' W' = \pi \sum_1^s N' W' .$$

Making this substitution and cancelling π, we have:

$$\sum_1^s N'W' = N\left(\frac{h_1}{E_1} \sum_1^s E_1'W' + \frac{h_2}{E_2} \sum_1^s E_2'W' + \cdots \text{to } n \text{ terms}\right). \quad (9)$$

Now, observe that

$$\frac{E_1}{N} = \frac{\sum_1^s \frac{E_1'}{N'} N'W'}{\sum_1^s N'W'} = \sum_1^s \frac{E_1'W'}{\sum_1^s N'W'} .$$

This, in effect, says that the careless exposures per employee for the normal (average) risk is equal to the average value of the careless exposures per employee for all risks, 1, 2, . . ., s, where each value of the careless exposure per employee is weighted by the corresponding value of the payroll.

Substituting this value for E_1, and similar values for E_2, etc., in equation (9), the equation reduces to

$$1 = h_1 + h_2 + \ldots h_n \tag{10}$$

which is identically true.

Our schedule-formula is therefore satisfied identically by the following statistical equations compounded from the figures for s risks, which it is assumed are sufficiently numerous to provide for the working of the law of averages.

$$\pi = \frac{\sum_1^s \pi' N' W'}{\sum_1^s N' W'}, \tag{11}$$

$$E_i = \frac{N \sum_1^s E_i' W'}{\sum_1^s N' W'}, \tag{12}$$

where i runs from 1 to n.

These are the equations which define the combinations of statistics.

The significance should be noted of the fact that these conditions satisfy equation (8) independently of the values of $h_1, h_2, \ldots, h_n$.

This means that the problem of determining the h_i's, which is solved by means of the companies' claim records, is entirely independent of the problem of determining the E_i's, which is to be solved by data obtained from inspection reports.