Report of the Working Party on Motor Data Base and Statistics

The working party comprised the following members:

J.H. Beck
L.M. Eagles
P.H. Grace
G.B. Hey (Leader)
J.L. Manches
J.M. Taylor
A.K. Thomson

The working party met as a group on several occasion and separate sub-groups had further meetings. One problem was to decide on the scope of our report and as we were told that one aim was to record actual practice in companies with advanced D.P. systems and also to form the basis of a teaching document it was decided to cover the subject in considerable detail, even though some of it might seem very elementary. The first step was for one of our members (J.H.B) to set out a possible coverage which he did under the following headings:-

1. Topics to be considered -

   (a) Rating Statistics.
   (b) Transaction Statistics.
   (c) Statutory Returns.
   (d) Marketing Statistics.
   (e) Claims Statistics.
   (f) Profitability Statistics.
   (g) Miscellaneous Statistics.

2 Headings for consideration under each topic -

   (a) Contents of the data base.
   (b) Quality of the data base.
   (c) Typical problems and possible solutions.
   (d) Definitions of data items or complete records.
   (e) Effect of errors.
   (f) Effect of delays in the processing of information.
   (g) Kinds of reports required.
   (h) Effects of inflation or other secular changes.
   (i) Any other relevant aspects.

3. The use of statistics -

   (a) The effects of inflation on relative and absolute premium rates.
   (b) The measurement of inflation for various types of cost.
   (c) Experience rating.
   (d) Other rating factors.
   (e) Index of premium rates.
   (f) Incidence of expenses over the policy year. (This may be verging on management accounts).
   (g) The use of external statistics.
   (h) Group rating.
   (i) Area rating.

It was recognised that this could all be regarded as falling within our remit, but to treat it fully would require a text book in itself. It was eventually decided to compromise and to divide the work up into two main sections, one concerned with getting data into a computing system and the other with getting it out again for analysis. Each of our members undertook to write part of one section and duly did so. Unfortunately, whilst each was satisfactory in itself they did not form collectively a coherent whole and were subject to considerable overlap.

Experience in writing papers with several authors showed that even if they work in the same office the time scale in obtaining agreement on a coherent narative was not compatible with the need to report in time for the Windermere meeting. In the event therefore the various contributions were incorporated, amended or re-written by the working party leader who takes responsibility for what is said and absolves any of the members from responsibility should they disagree from what has been said or the way it has been said; although all have had an opportunity to register their views it was simply not practicable in the time available to get everyone to agree to everything.

It is also necessary to add a further disclaimer. None of us regards this paper as a polished offering. We are all very well aware that it is long, repetitious, in some areas very repetitious, poorly written and generally not of the standard we would like in a permanent paper. We do hope though that it will stimulate discussion and after suitable amendment and redrafting into a more coherent form might provide part of an introduction to non-life insurance for those actuaries - the bulk of the present members of the Institute and most new entrants-who have little or no experience of non-life business. Whilst we deal with United Kingdom motor insurance we must never, however, forget that this represents less than 10% of the United Kingdom Companies' non-life insurance market,and other classes and countries present even more formidable problems.

Readers will appreciate that whilst there is a certain amount of new information the bulk of the paper is concerned with elementary matters which are none the less of considerable importance,and that it is more intended as educative than advancing the state of knowledge

MOTOR DATA BASE AND STATISTICS

Introduction

This paper is divided into two parts. The first considers the problems involved from the moment that a vehicle owner decides to apply for cover on one or more vehicles, through the resulting documentation to the recording on the computer of some of the information as well as subsequent changes to the details of the cover, the incidence and settlement of claims and the eventual termination of the cover. The second deals with the analysis of the records so created with particular reference to marketing, to premiums and profits, and to statutory returns.

It will be assumed that a computer is used and that the data are to be held in a form that the computer can read, normally tape or disc. Experience has shown that very large amounts of data are required if any subsequent analyses are to give reliable indicators, and sample investigations by manual methods are not often used. It must be emphasised at the outset however that the detailed responsibility for creating and processing a data base will be that of the data processing professional and not that of the actuary or statistician. There are, however, problems in communication and if the actuary can have a general understanding of the D.P. processes commonly used and of the alternatives available, it is likely to promote a better mutual understanding and to lead to better and more efficient use of the computer. We shall consider this problem in more detail later. We shall also need to consider the administrative procedures that lead to the submission to the computer of data in the form it requires and for ensuring that errors disclosed by computer analyses are investigated and correctly adjusted.

The nature of Motor Insurance

Motor business differs markedly from life business and familiarity with the different statistical aspects applicable to motor business is essential in developing a motor data base. One particular feature is the frequency with which amendments to policy records are required and the importance of and the difficulty in ensuring that the corrections are applied to the policy which was intended. We shall cover this aspect in rather considerable detail as we think that much of it may be new to the actuary who has been concerned solely with life business.

The treatment of life business is an established actuarial concept dealing primarily with the probability of an event happening (death) with a pre-determined financial consequence to the Insurer (payment of Sum Assured). Interest and expenses apart, the concept of life business is further eased by the existence of published rates of mortality which are confidently assumed to remain stable from year to year and by the absence of partial claims in the events of a man's "half-dying". In any case the effect of changes in the rates of mortality is small in respect of most types of life business whereas changes in claim rates in motor may have very serious implications.

Whereas mortality tables normally take into account only sex and age, the statistics of motor business embrace many inter-related variables, complicated further by the absence of any general agreement on what variables to use as well as what definitions and what levels of them to use in analyses.

District, type of car, age of car, and age of policyholder are amongst variables popularly accepted, but motor business is at the same time a short term contract which may or may not be renewed at the end of 12 months, where the contract may, and often is, changed during its currency (e.g. on change of car) and where the order in which transactions are presented to a statistical system may not coincide with the order in which they actually happened (for example it is quite common for claims to be advised to the statistical system before the basic policy has itself been recorded).

A major complication in motor business is that the policyholder is commonly offered a discount if he makes no or few claims or, on the other hand he may find his premium loaded or renewal not invited if he makes too many.

Whilst it is not imperative for such information to be recorded on a computer file, there are very considerable advantages in doing so.

Different starting points will inevitably arise in considering the data base, for example :

a) A manual statistical system might already be in operation but a review of the data recorded (or retained) is being undertaken to ensure the maximum benefit being derived, and possibly integrate it with a computer system.
b) A computer system might exist but the form of the data or the way in which changes are recorded or old information retained may be such as to impede or prohibit proper statistical analysis and we may wish to remove these impediments.

We shall, however, consider the system we might design were we not inhibited by any existing system.

One of the primary considerations in developing a motor data base must be the purposes to which the system will ultimately be put and the different interests which it may be required to serve. It is essential to realise that emphasis must be placed upon the information that it will be wished to get out of the system and while the statistician may be acting as an independent user, or as a general coordinator, the system is likely to encompass several needs if it is to be used to the best advantage. The motor underwriters will clearly be closely involved but the needs of other users such as Administration sections, Accountants, Sales and Claim Departments will also be of importance. The statistician may well be the residual user having to make do with crumbs from the rich man's table, but it should need relatively little additional effort in a well defined system, to give him all he needs, compared with the little he can extract from a system designed solely for administration and accounting.

The needs of these groups will naturally overlap but each will have special interests which are briefly set out below.

UNDERWRITERS AND STATISTICIANS

It is difficult to generalise on the needs and functions of underwriters. In the past, they have been regarded as superior beings whose judgement cannot be challenged. It is, however, quite clear that whilst there are many areas where judgement of an experienced underwriter is the only tool available (insurance of risks where claims are few or very variable or the statistician may not have had time to accumulate data) there are other areas where enough information is available, subject to proper analysis, to enable the statistician to give valuable advice to the underwriter, advice which he will ignore at his peril. This position cannot be emphasised too strongly or too often and underwriters who think they can ignore the competent statistician in certain fields, merely reveal their profound ignorance of the true basis of insurance. Equally the statistician must not pretend to do more than his investigations allow and must ever remain aware of such things as moral hazard, selection and the realities of a competitive commercial environment if his advice is to have any value. We can say that the underwriter must seek risks where he can obtain a premium large enough to pay the claims and other outgo and that information to help him reach such a decision will come partly from statistical analysis and partly from accounting data, if they are prepared on a rational basis (which is not in present circumstances that normally adopted by accountants).

The information required by the underwriters and statisticians will extend to a large proportion of the information available from the file but they will be particularly concerned with assessing risk premium according to the various levels of rating factors and on the relative possibilities of obtaining different types of business, and will also be concerned with estimating processes which are essential to lead up to estimates of risk premium.

ADMINISTRATORS AND ACCOUNTANTS

They will require information for Company accounts, Department of Trade returns and for the calculation of reserves.

SALES STAFF

They will be concerned with effects of advertising campaigns, the cause of lapses, sources of new business and information in regard to agents.

CLAIMS SECTIONS

They will or may be concerned with the progress of claims, the possibilites of measuring administrative procedures in the economical settling of claims, the expenses of paying and agreeing claims, also the making (and possibly the adequacy of) estimates, although this point will be dealt with later.

The statistician may well be responsible for the preparation of most of the analyses and reports to aid all the above groups.

The development of our computer system must further conform to the requirements of the computer department who must be closely involved in any plans regarding either input, or output of statistics from the computer. There is clearly little value in issuing coding details before the computer has been programmed to accept them or in trying to add reinsurance information to a system which has been conceived only in gross terms.

It will become apparent that a considerable volume of information will be needed on policies for purposes quite apart from statistics and with a view to the ultimate size of the data files which will be produced some limitations may be imposed on the volume of space containing information required, or more likely retained, solely for statistical purposes.

The division between what constitutes administrative information and what is statistical is rarely clear. The main file will need to contain information for both statistical and administrative purposes and details are likely to include

Policy number
Branch or district office code
Agency code
Renewal Date
Description of cover (rating area, vehicle groups, age of policyholder and so on)
Limitations on cover (excesses, restrictions etc.)
Claims history

In the case of the administrator he may be satisfied with manual records of cover and claims, but such a system would impose an almost impossible barrier to statistical work and we must assume here that full details are recorded in the computer system. The statistician will require full details of the progress of each claim settlement, whereas the administrator might be satisfied with less. Both are likely to need to know the premium but it matters little to them (although it may be very important to the D.P. department) whether the actual premium charged or the annual rate is actually recorded or whether the data are available to enable premiums to be calculated as and when required, a point that will arise later.

The administrator will also wish to know more about commission arrangements (and so may the statistician at times) and to have details in regard to the name, address and phone number of the policyholder as well as the place to send renewal notices to (policyholder, branch, broker etc) and may wish to have names rather than codes for branches and agents. If both a name and a code are recorded, special steps must be taken to ensure they agree.

Processing of Changes

Any of the information is liable to be changed at any time and often there will be several changes in the course of a policy or calendar year each giving rise to a new "status" that will be in force for this period. Unfortunately, the dates when a status starts and ends are often weeks if not months before they are accepted by the computer so the file will always fail to record the true insurance position having many statuses or policies which are no longer in force and being unaware of new policies and changes that took place some time ago. This gives it a bias that will need to be examined later.

As time passes and the amount of information for a policy mounts up the computer file may eventually become too bulky to handle unless records are either dropped or pruned. Rather than lose statistical information in this way a separate statistical file might therefore be created in addition to the main file so that a record of all statuses that have been in force at any time within a named period can be developed and maintained unencumbered by other information needed only for administrative purposes.

An important principle in the success of the system, and one only made possible by the use of a computer is that statistical records should be generated and passed down to the statistical file automatically from day-to-day transactions. Branches advising a new policy or a change in terms will not thereby need to submit a special additional advice for purely statistical purposes and the normal renewal of a policy can lead to a statistical record being created without the Branch being involved at all. It is in fact highly desirable that all information used by the statistician be based on data maintained for business or accounting reasons as there will then normally be a strong restraining influence on the generation of data that cannot be shown to be wrong but where it is regarded as being "for statistical purposes" might receive less care than it should.

We shall from time to time refer to "files" but it must be emphasised that we shall regard a file in this paper as a logical concept rather than having a separate physical existance so that one physical data set (volume of tape or disc) may contain several logical files either "concurrent" or "consecutive".

For example a file of current policy statuses and one of statuses held "for statistical purposes" may comprise one physical file with a series of statuses for each policy covering its history during at least some stated period. Alternatively there may be two physical files, one of administrative statuses and one of statistical statuses. The latter may comprise only out-of-date statuses: if it contains all statuses processing may be eased but problems, likely to be severe, will arise in ensuring that the "in force" records on the administrative and statistical files correspond precisely.

The Transition from writing to computer

Having now discussed in general terms the sort of information we wish to have recorded and retained we must consider in some detail the problems of getting that information directly recorded. There is a further problem in ensuring the data are not lost or mutilated by accident after they have been properly recorded but we will return to this in the second part of the paper. It will be necessary to consider the problem in general terms rather than in relation to a particular system since there are very large differences between companies in their practice. We shall, however, restrict ourselves to systems in which all information required to calculate a premium is in the computer system. If we have to rely on manual systems for any basic information then the statistician will be very restricted in his analyses. Note that we do not necessarily require the actual premium charged to be calculated and recorded. It may well be a help to do so, but this is more a D.P. problem and the statistician will very often become concerned with matters other than the premium since he will prefer to measure his risk by something more reliable and stable.

The reasons for requiring the file to have information to enable the premium to be calculated are twofold. Firstly because the statistician is likely to want to analyse the results against all the factors taken into account in the premium calculation. After premium he may well wish to use other factors against which to measure performance but he at once comes up against the problems of maintaining reliable data. We shall return to this later. The other reason is to be able to apply some notional (or actually used) premium scale to all statuses and not to analyse merely the premiums actually charged: the purpose of doing so will be apparent in the second part.

THE PROPOSAL

The usual originating document is a proposal completed by or (undesirably!) on behalf of the policyholder. This asks questions relating to the vehicle to be insured, to the policyholder and any other likely drivers, and to their past insurance history if any and to convictions (where that question is legally permissible).

Our first problem is to find out whether the answers are true. Experience tells us that some are very likely to be true (the registration number, make and model, date of first registration). Some are at times liable to mis-statement (whether the vehicle has been hotted up, whether it is used for commuting or business purposes especially where the business use is by someone other than the policyholder) whilst other answers are frequently found to be wrong (driving by young persons, convictions, past insurance history, driving experience, mileage). Unfortunately it is costly to establish the truth and to do so is not conducive to obtaining business so that it is uncommon to reject statements unless the underwriter smells a rat. A good underwriter should have a good nose: in fact it may well be the best contribution he can make to motor insurance nowadays. In fact, the time spent on scrutiny of a proposal must be limited since the average premium in 1975 on new business is probably of the order of £50 net and the contribution to all expenses (other than commission) including claim settlement is likely to be under £10 which does not permit much time per policy if that time is to be spent by a senior official.

From the information on the proposals some computer record will be created either by card punching or direct to disc or tape. It may be necessary to code some information before punching (for example rating district) but much may be left to the computer. This input record will in due course be offered to the main file via some sort of validating process that will reject values that cannot (or are very unlikely to) arise. This process may also act as a crude form of underwriter, especially for such tests as limitation on cover or type of vehicle for young policyholder or those revealing bad claims or driving histories. To verify that the information is positively right rather than that the codes are valid is a much more difficult process.

All these validity tests will generate a series of messages which require human investigation and intervention. Normally they will require reference to the proposal papers, often to the proposer either direct or via branch or agent. This all takes time and to do it properly takes longer than to "get it on the computer". Meanwhile, the file is ignorant of the existence of the policy and, most likely, the policyholder will not receive his policy. He may get one prepared directly or indirectly from the proposal by manual means. If so, neither he, nor the company, can be sure that the policy issued agrees with the computer record when it is duly set up, or even for that matter if a computer record has been set up! Meanwhile, also, other things may be happening. There may be a claim, or even several claims, there may be endorsements, either because of a change of address or vehicle or merely because the proposer realises he has given wrong information. None of this can be processed until there is a record on the file to alter.

This is a very real problem - to neglect it can be far more disastrous than neglecting risk theory or other abstractions and we make no apology for labouring the point. As we said earlier, motor insurance is very different from life: in that case there are fewer obstacles to setting up a valid record and far fewer changes of importance. The most important, namely death, could well be coped with manually and the computer left to record the event in due course. In motor such an attitude might be disastrous.

CHANGES

When we deal with changes in a policy record we must begin by ensuring that we are changing the record we intend to change. Merely to quote a policy number is unreliable unless the number incorporates a check digit such as is used by banks in allocating account numbers. The subject of check digits and their efficiency has been widely explored although it is not easy to find reference to it in published work. One of the simplest systems is to multiply the digits of a number in turn by different prime numbers, add the products and find the remainder when the total is divided by 11. For example, using three, seven. one as multipliers, the check digit for 1234567 is found to be 1x3+2x7+3x1+4x3+5x7+6x1+7x3 = 94, The remainder = 6 (=94 - 88) so that such a policy number could be quoted as 12345676. This system will pick up any error in a single digit and most transpositions (e.g. 1324567, but not 1567234 which is a most unlikely type of error in normal cases). There are many variants of this arrangement which the reader can try for himself but it is probably difficult to improve on the above plan without being much more complicated. There is, as ever, a problem, namely that since the divisor MUST exceed 10, remainders will run into two digits unless some possible policy or claim numbers are not used - with a non-consecutive set of account numbers, this is not serious, but there is some reluctance to adopt it with policy numbers and claim numbers although this is a little irrational since one can readily ascertain the total number of policies or claims in a series from a knowledge of the first and last numbers allocated by means of a simple table which the computer can prepare. Another solution is to use a letter rather than a digit. One company has done this with claim numbers and found it a considerable help - although it probably increases the tendency, on replying to queries, to alter the quoted letter to "get it through the computer"! That is liable to make two more errors to investigate. It was also found in practice that letters were more often confused than figures, CL, AH, XK being favourite pairs, so that finding 11 distinct letters was a problem: this Company currently uses A B D E F M P R T W X. The simple solution of using a two digit remainder could cause problems in card or record capacity.

If one does not use check digits (or as an additional precaution if one does) it is useful to record some information other than the policy number which can be tested by the computer against the in force records. Agency code may be useful, especially if it is not associated with batches of policies. Vehicle registration number, renewal date, policyholder name are other possibilities, but beware of being too complex for the change record may itself be wrong in some of the check details and cause an excessive number of validity rejects. One case reported related to warranty repairs on new cars where validity reports threw off so many unnecessary errors that the whole system had to be abandoned. One must always realise that 100% accuracy in a computer file is not obtainable and the point at which one stops checking must be determined by cost and time. One must however, always bear in mind that sooner or later errors will get through and steps must be taken to see that they do not upset the analysis. For example, a single renewal notice with a premium of £250,000 can be stopped: in a total premium of £20,000,000 such an error easily gets lost. The only safe rules are to take reasonable care in seeing the records are right to begin with and especial care in program writing to test that nonsense does not arise and allow misleading results.

We do not propose to go further into the details involved in amending records, but we must refer to the problem of lost or mutilated records. Any well-designed system will provide means of going back two or three generations of file but from time to time one discovers that a complete record or set of records was lost some time ago and they have to be recreated manually. It must therefore be possible somehow to recreate these records by having the data available somewhere and also being able to add them to the file. These are very dangerous processes and are particularly liable to error either through carelessness or, perhaps more seriously, fraud. If the statistician is responsible for production of statutory returns, he must have regard to the possibility of unauthorised creation of supposedly "lost" data, a point considered later.

DELAYS

The time taken to record information on a file depends on many factors, including:-

- the method and frequency of updating,
- the frequency of, and time taken to correct, errors,
- delays in notification of changes which are liable to be particularly acute when brokers are involved and report only infrequently,
- delays in claim notification where the effects of Knock-For-Knock agreements are particularly noticeable.

It is important to monitor these delays: we consider this later,

PERIODS OF COVER

We have referred to a policy status, that is a condition which remains stable for some period. The simplest logical concept is of a status with a starting operative date and (if it is now superceded), an ending operative date. The recording on the computer file may enable this information to be extracted but may in some cases be in condensed form where a change is recorded only in the field to which it applies to, avoid repeating data that are not changed.

MULTI-VEHICLE POLICIES

If a policy covers several vehicles problems are likely to arise. If each vehicle is treated as a separate insurance and the records can be separated into parts corresponding to each vehicle, all is probably well, and it may be possible to analyse such cases to see if they behave differently from the general body of single-vehicle policies. If however, there are different arrangements in regard to no claim discounts or if the policy is to be treated as a fleet and rated on an experience basis it may be necessary to identify the policies so that they can be excluded from the normal analyses.

CLAIMS

On the claims side, where estimate run-offs are likely to be essential statistical returns, estimates must be differentiated from payments. Companies have different practices over the recording of claims on the computer, some restricting computer records to payments only, where others maintain a full claims history recording not only payments but also the various estimates placed on the claims throughout their currency.

These may be adjusted manually or automatically on the making of payment or they may be adjusted periodically for example at the end of the accounting year (note that care is needed where a payment is made near the end of the year to ensure that the making of the payment and the amended estimate are in step with each other).

There are advantages in maintaining an up-to-date estimate and the advent of quarterly returns to the D.O.T. will make this question more urgent. However there are major problems in regard to the frequency of estimating and the methods to be adopted which we shall not consider here. For the moment we must remember that at one extreme there will be estimates for every claim in which any amount may be outstanding and which are continuously updated: at the other extreme all estimating may be by statistical methods and few or even no claims(at least for the last year or two)may have individual estimates. Our treatment of these cases may range from the application of an average estimate on every open claim to ignoring them altogether on the claim record. The consequences of adopting the various options will be discussed later.

To record the claim payments alone will considerably restrict the scope of the system - to produce the experience on a cohort which includes outstanding claims it is necessary at least that the latest estimate on a claim be recorded. Retrospective information however, will often be required for year-end positions and can be provided for by recording on the statistical record for the claim the position at consecutive year ends. Nevertheless to obtain a complete retrospective experience giving the position at any point in time requires only that the current estimate should not be lost when a revised estimate is submitted and this can be simply achieved by the addition of estimate records to the statistical file: for any claim there will be a succession of payment and estimate records, each showing the date on which it was raised and enabling the retrospective position at any point in time to be derived from records raised prior to the required date.

e.g.	DATE	PAYMENT MADE	ESTIMATE OUTSTANDING
	1.6.72	-	1,000
	1.9.72	50	1,000
	30.6.73	-	10,000
	1.12.73	6,500	2,500
	1.3.74	2,300	-

By reference to the date of each record a run-off of the claim can be produced showing for instance at six monthly intervals, the current position as it then appeared.

POSITION AT	PAYMENTS TO DATE	ESTIMATE OUTSTANDING	TOTAL
30.6.72	-	1,000	1,000
31.12.72	50	1,000	1,050
30.6.73	50	10,000	10,050
31.12.73	6,550	2,500	9,050
30.6.74	8,850	-	8,850
31.12.74	8,850	-	8,850

Two remarks should be made. The first is that this sort of sequence of estimated final totals is by no means uncommon with liability claims namely a steady rise to a maximum followed by ultimate settlement at a figure rather less than the maximum previously attained. The second is that if claims are estimated statistically it is unlikely that this process can be usefully begun before the end of the second year or so from the notification of the claim. that is when most or all claims have individual estimates.

In practice few companies record specific estimates,(also known as 'case' estimates) on the computer for every outstanding claim. One practice is to allocate a standard reserve to every claim expected to be settled for less than a certain amount. This saves considerably on the administrative task of manually estimating every individual claim, although in operating such a system an additional run-off that will be wanted from the statistics(affecting as it does the overall level of outstanding reserves that the Company must hold at any time) is a review of the adequacy of the standard estimates by examination of the actual amounts for which these claims are ultimately settled.

It must be realised, however, that this can only be done on a bulk basis and not on an individual claim basis.

The use of standard estimates does, however, have underwriting drawbacks for classes of business such as motor fleet where "incidence" may be reported to the Company more because of the insured's internal accident reporting procedure than because any claim payment is likely. In any event such claims, if damage only is involved, may be quite unlike claims for individual vehicle policies since rating will probably not involve no claim discount, giving rise to many small claims. There may well also be a larger excess in such cases, causing further differences.

WHAT IS A CLAIM?

We have been referring on several occasions to claims and payments. Unfortunately, however, it is not easy to define either. So far as a claim is concerned practice varies from one Company to another, but in the United Kingdom we think it is fairly standard practice for any one incident to give rise to one claim at most for each vehicle involved, however many other vehicles, people, animals or property are involved. One car hitting a lamppost will give rise to one claim (at most) in the books of the insurer of that car. Two vehicles hitting each other will give rise to two claims (at most), one in the books of each insurer, although if one vehicle were clearly and solely to blame and there were no Knock-For-Knock agreements involved the insurer of the other vehicle might not set up a claim file or give it a number. When there is a multiple collision, things get complicated although if all damage happened more or less simultaneously, there would probably be one claim at most for each vehicle involved.

It is, however, a normal policy condition that an insured reports any incident that might give rise to a claim and many such incidents are reported either direct or through brokers. Such incidents may not lead to any liability on the part of the insurer either because no claim is made by a third party or the Insured merely reports (to comply with the terms of the policy) but does not claim in order not to jeopardise his N.C.D. There is a third case where no payment is made, because the damage caused by the Insured is settled under a Knock-For-Knock agreement by some other insurer and the Insured is covered for third party only. It is probably essential to record such "incidents" as claims as they will normally cause the policyholder to lose his N.C.D. exactly as if the Knock-For-Knock agreement did not exist and his own insurer was therefore called upon to pay.

The most important thing is to maintain consistency within a Company. Whether an incident is treated as a claim immediately on notification or later is not vital although so far as the statistician is concerned, it is better to create a record as soon as possible, a view doubtless shared by the auditor. Some companies are believed to defer counting a claim or allocating it a number until either payment is made or some estimate raised. Some are also believed to count as a new claim one which is reopened after being "closed" in the normal way.

This is an area where the administrator may wish to be heard, but it seems wise, in the light of what we have to say later, to require every notification to a company or to its agents or brokers to be recorded at once as a claim so that it can be counted and its progress watched until final settlement, and for any subsequent operations to take place under the original claim number. The convenience of the administrator may at times have to give way to the needs of the analyst who must have the fullest information available if he is to give reliable advice.

WHAT IS A PAYMENT?

This seems simple but it is not. Firstly if an insurer pays cash or draws a cheque relating to an identifiable claim or covering a number of identifiable claims then it should count as a payment on the claim or one payment on each claim (if more than one) of the relative amount applicable to that claim.

However difficulties at once arise with professional fees. If an insurer pays one surveyor for specific reports he can normally allocate the payment to individual claims. If however he pays the salary and expenses of his own staff for engineering reports or legal assistance, then it is not so easy to allocate the cost to individual claims and even if one does do so on an average cost basis it is likely to involve an amount of unnecessary work out of all proportion to the benefit (note that recent D.O.T. proposals may involve the industry in problems in this area). This applies also to payments in bulk to the B.I.A.engineers units and may apply to emergency treatment fees and police reports if they are paid in bulk out of an expense account.

One must have a standard practice in a Company and we hesitate to suggest what it should be although some of us think that the effort involved in separately charging, validating and analysing all payments under these headings, other than payment to a third party's legal advisor above some minimum sum, could conveniently be grouped with general claim settlement expenses with little, if any, loss of information. Similar problems arise with payments under M.I.B. and any other indemnity schemes. Once again consistency is more important that being identical with another insurer although the more that common practices are involved, the better in the long run.

Further problems arise with recoveries, but first we must dispose of a special kind of recovery, namely that from a reinsurer, whether excess of loss or, if such exist, stop loss, or in respect of treaty business. They must in every case be rigorously excluded from settlement details of any given claim. It is vital to work with gross liabilities. The fact that part of the liabilities are shared with another insurer in return for a share of the premium must not be allowed to interfere with the assessment of the full liability under a claim (although abnormally large claims may require special treatment and are considered later.)

Other recoveries fall into at least two categories. One comprises reimbursement of payments made for which the insurer was primarily liable but with a right to recover, or where, by concession a payment is refunded to preserve N.C.D. The other and more important class comprises payments received under Knock-For-Knock and claim sharing agreements: in the former case, the insurer may pay, but if he later discovers the insurer of a third party, and he has an agreement with him, he will claim reimbursement under the agreement.

It is desirable for some purposes to be able to separate payments and recoveries under claim sharing agreements in order that claim amount distributions may be examined at the gross and not the net level. Unfortunately, this is not possible under Knock for Knock Agreements covering damage to the Insured's own vehicles.

In some cases an insurer may seek to recover from an uninsured party by instalments over a period, and some of these cases may involve hundreds of individual recoveries of small amounts.

The treatment of recoveries is believed to vary considerably from one insurer to another but the position outlined above is thought to be the most satisfactory.

Type of Payment

It is normally considered desirable to provide a breakdown of net payments by the nature of the payment or the peril in respect of which it is made; common examples are:

Own damage
Fire
Theft
T.P.B.I.
T.P.P.D.

Legal Charges
Other (mostly PA)
Claims Sharing

This means that every payment made has to be coded. As usual the question arises, since to the claims clerk this is merely statistical fodder, how can we ensure accuracy? As usual the answer is we can't, but we can take some precautions. We can consider these at the same time as precautions to ensure that payments are debited to the correct claim and generally in the validating of data.

The use of check digits or letters has already been considered in relation to policy number: the same principles apply to claim number and if a claim number contains several "parts" (eg. claims or branch number, year of notification, as well as an actual number) they can all be accommodated in one check degit, or in more than one if one wants to take extra care. However, if the statistician has to return a notification of payment on account of a check digit showing a discrepancy, the originator may be faced with a problem, particularly if it is some time since the cheque was drawn. If the claim number has been quoted wrongly (and it may even turn out to be the wrong branch just to make life difficult) he may not easily be able to locate the correct number unless he knows the name of the payee and happens to remember the claim it refers to. It is tempting in these circumstances to alter the check digit to that for the claim quoted: the clerk must, however, be made aware that he should reply "don't know" rather than (in effect) 'don't care". In such a case the statistician has the problem of wondering where to allocate the payment. One office deals with this problem in the following way:

1. Cheques are normally prepared and issued through the computer. This will reject any check digit errors and since there will be an accompanying advice of payment to be sent out with the cheque, there is a good chance that the correct claim can be identified without difficulty. The occasional error still gets through but these are very rare. The computer also tests for valid payment codes and certain other information to see that the amounts are valid.

2. Recoveries are listed manually and processed fortnightly. Errors here are referred to claims offices and since they have the names of payees and Insureds, can normally find the correct claim number easily if the wrong one has been quoted.

3. Some cheques can be issued manually, or payment may be made by internal or bank transfer. These need to be treated like recoveries. They are sufficiently rare to cause few problems.

4. A payment or recovery following a closure must be coded "R" (for reopened) and the claim update system (see later) tests that "R" records are applied only to claims previously closed and that only "R" records are applied to those claims.

5. On closure of a claim the claim office submits to the statistician a note of the closure with details of the total payments and recoveries. These are compared with the cash on the tape which is derived from actual financial records. This process shows up residual errors in the recording of payments and recoveries and permits a reliable indication on the tape of the closure or reclosure of a claim.

6. A pool of records where the claim number is unknown is maintained and most are ultimately placed successfully.

The system has been working for several years and is thought to keep a tight check on the master tape. Even so it is surprisingly easy to find major errors creeping in through unforseen effects of program or system changes.

Coding of Information on the Master File

Descriptive Coding

It can be seen that there are many factors to be taken into consideration before ever reaching the finer coding details for any entry. The paragraphs above have dealt with the general form of the system and the information need to obtain the financial outcome of different periods of insurance.

The final step towards the development of the statistical system is the description of the type of business to which each transaction relates. In the first place such description will be a general business classification (Private Car, Fleet etc.,) influenced probably by a company's departmental structure, but further coding details refining the description will be equally appropriate. Clearly, however, the value of any finer description replies heavily upon having a sound base on which to build.

Whilst the actuary may be the expert in the field of statistics, it is the underwriter who is the expert on the business being written and his assistance is necessary from the start. Detailed discussion with him about the factors that are relevant to his underwriting will help to obtain a list of items that may subsequently be considered as the basis for statistical coding.

On the other hand there may be factors which the underwriter does not feel are significant but which may be material. It is important at this stage not to lose sight of the aim of producing a systematic definition of the main factors, and to seek to establish a simple list of items that can subsequently be broken down into appropriate codes. Policies are written on an individual basis and inevitably there will be many problems peculiar to certain types of policy; no statistical system, however, can embrace the many exceptions that will be abound and at this stage progress can more easily be made along the lines of what can be coded rather than what cannot.

Taking some of the factors which might be considered for coding on private car motor business, distinction might be made as follows:

Coding possible	Coding impossible or impracticable
Use	Annual mileage
Cover	Traffic density in area
Age of Policyholder	of common use.
Age of Car	Driving proficiency
Make and model	Colour of car
Rating Group	Annual salary
Area	Car roadworthiness
Driving history	Average number of passengers
(convictions etc.)	Moral hazards
N.C.D. (present and past)	Physical and psychological
Excesses (Voluntary/Compulsory)	conditions.
Occupation (within certain limits)	Extent of use for business.
Whether used for commuting	Extent to which the vehicle
Whether garaged at night	will have to be used even in
Membership of Institute of	adverse conditions.
Advanced Motorists.	
Number of vehicles insured	

It may well be that some of the factors in the second column have more influence on the risk than those in the first. Clearly, we use some of the data in the first column as a proxy measure of the real risk but not as a direct influence on it.

Discussions with the underwriter serve two necessary purposes. In the first place they give the actuary the opportunity to appreciate the more detailed aspects of the business under consideration, giving him an insight into the way that a risk is underwriten and the day to day problems that confront the underwriter. Secondly, and the importance of this cannot be over-emphasized, they draw the underwriter into the development of the statistical system and help to achieve a relationship in which the actuary can apply his skills for the benefit of the underwriter.

Agreeing the factors to be coded is a task involving months rather than weeks and starting from a list of the different underwriting factors that are taken into account for a risk, searching discussion and planning must be undertaken at this early stage. It must be borne in mind that branches will have to operate the system being designed, and that the longer the system remains in force without any subsequent change, the more successful it should prove. Subsequent deletions of parts of the coding will be unpopular amongst those who have spent hours recording unneeded information and it is important to avoid a tendency (and in some cases perhaps a tradition) to put into a system all conceivably relevant information on the basis that some day somebody may require it. Additions to the coding, are more easily made and are bound to be needed in time and a few unused coding positions will give the system a valuable flexibility for the future. One must remember, however, that with new codings the claims staff have to remember to make them, as well as to learn what they are. In one office a test is being made by asking claims staff to code specially claims that look, from the start, as though they may prove to be expensive. One needs an objective test for this and must ensure that it is applied regularly and uniformly. The results to date have shown that this result is not likely to be attained easily.

Certainly to start with, an underwriter's main interest will centre around more direct aspects of his account, such as the volume of business being written in different categories, together with the corresponding claims experience. Some data collection, perhaps through manual returns, will already be taking place prior to the development of the new computerised system, and a restriction that may be imposed is that the new system should be able to provide this data in the same form. Whilst this is sometimes an undesirable restriction, it does allow for continuity of familiar returns and an opportunity to check the accuracy of the computerised information before the manual returns are discontinued, and as such is a limitation that is often accepted in practice. It is much more likely to reveal, once the program is working, that the manual system was highly unreliable. The experience of some statisticians is that their early investigations tend to reveal a lack of control in existing systems and the need for much re-organisation before any progress can be made.

Sub-division into further coding factors is likely to be restricted at first, and whilst always available as a potential statistical development, it must be remembered that the value of any sub-division will be limited by the volume of homogeneous data existing in any one category. Five different coding factors, each of which may take ten different values, immediately give 100,000 different potential sub-classes.

Statistical factors which might be considered on the description of cover for each car are shown on page 12 and identification of each statistical record (Branch, policy number etc.,) would also be advisable.

Subsequent statistical analysis of the data will, however, certainly require a measurement of exposure for which the concept of vehicle-years is invariably used (one vehicle-year being defined as one vehicle exposed to risk for a period of one year). While it might not be necessary to actually record exposure on each statistical record, it is essential that this can be derived from the data base when needed.

In addition to purely descriptive details there are other items of a general nature which have been mentioned previously and which may also be appropriate -

Type of transaction (new business, renewal, etc)
Agency details
Premium
Date of inception of policy
Date of accounting entry
Period covered by risk (commencement and expiry dates)
Renewal frequency
Renewal date.

Coding of information regarding claims

It is very tempting to arrange for information on the origin of claims to be recorded on the file. It is, however, a temptation that should be resisted in the absence of some compelling reason. An insurer is in business, inter alia, to make a profit and that profit is affected by a £100 (or £100,000) claim equally whether it arose from a right turn, a cross roads collision, a left turn or a passing mini skirt. If one can relate claims cost to factors that can be used in rating, either directly or by a proxy variable, then a case could be made out for analysing claim cost by type of accident. If not there seems little merit, from the insurer's point of view, in recording type of accident. But even if the association could be established clearly and certainly it is difficult to forsee "are you inattentive when turning right?" or "are you abnormally susceptible to mini skirts?" as a question on a proposal form.

There may be a very good case here for investigation by the Transport and Road Research Laboratory but one of their problems is exposure for which they tend to have even less reliable measures than any insurer. If one found that an abnormal number of accidents happened, say when turning right, or in fog on motorways, then one could focus attention on them; the insurer is not likely to be able to do so in general, although one thing he might be able to do is to compare injuries when seat belts are used with those when they are not.* However, although this is a digression in this part of the paper, one must be very careful since there may well be a close correlation between the nature of injury accidents and the wearing of seat belts if both are at all closely related to the age of the driver or policyholder.

In fact the one piece of information in regard to a claim that might be useful to an insurer is age of driver and relationship, if any, to the policyholder, where the policyholder himself is not driving. The results may show the extent to which claims arise from driving by persons other than the policyholder, but once again if the results suggest that a high proportion of such claims arise, one cannot necessarily relate them to exposure since there is no measure of the relative amount of driving. Similar remarks apply to foreign use. If we take all vehicles for which green cards are issued and calculate the claim frequency per vehicle year of the cover we shall probably find it greatly exceeds the expected frequency for the group. So what? How many claims would they have had if the holiday or business trip had been in the United Kingdom and not abroad? The same? More? or less? We dont know and there seems little hope that we shall ever find out.

Such an enquiry might not be entirely useless if we found for example that on average a green card period of three weeks produced more claims (or more claim costs since foreign claims may prove very expensive) than 12 months (or even six months) of normal cover. If this was the case the existence of this problem would probably be known to claims staff and a special investigation could be mounted, possibly manually.

Before any special coding of information regarding the occurrence of claims is contemplated, the statistician must ask himself - "what use will the answer be?" If the answer is in any doubt whatever, it is unwise to consider the coding: to call for codes one cannot use will not merely result in careless completion; it is liable to be infectious and spread to a disregard of the need for reliability on the things that do matter.

We will now consider some practical problems; they may seem trivial to the analyst, but we cannot stress too much the need to control input information with the utmost care if one is to get reliable results.

*but there could be many cases where he is unaware that there would have been injuries had a belt not been worn.

Accuracy of data

Despite the most careful and exhaustive attempts to ensure the accuracy of the data thatis recorded, the resourcefulness of branch staff in finding ways round the checks imposed, albeit one would hope accidentally, has apparently no limit, and it is inevitable that errors will always be present - perhaps through a genuine mistake, a slip of the pen or possibly even because a young lady may have twenty policies to code up before she goes to meet her boyfriend in five minutes time.

However sophisticated the data base therefore, its success will depend heavily upon the accuracy of submission and unfortunately the value of the coding system and the continual need for accuracy will notbe easily apparent to branch personnel who may never see the results of their own labour.

It would be a shame, however, if the efforts that went into the development of a sound data base were not at least matched by further efforts in trying to protect it in operation.

Coding instructions should be set out very clearly in a suitable handbook which in the event of future changes to the coding should always be kept up to date.

The coding manual will, however, describe the procedures that should be followed - it can do little to encompass the procedures that should not. Depending upon the branch organisation and staff, some manual checking may be possible before coding sheets are submitted to the computer, but having done all that might be feasible in preventing error from arising in the first place, the only checks that can be applied on an automatic basis will be those carried out by the computer.

In truth, confidence in the accuracy of the data recorded will depend entirely upon the sophistication of computer validity checks, which can range from basic validity tests of acceptable values for each code to a complete premium check from the statistical coding. There is a tremendous value in the latter if it can be achieved and a decision may have to be made as to whether it is worth coding factors purely for the sake of a computer premium check. This decision may be aided by administrative considerations, such as policy and endorsement preparation by the computer and possibly too by the extent to which manual records are retained in addition to computer information.

Most of the validity checks that can be applied by the computer will be to ensure that the information falls within the ranges permitted by the statistical system, but certain other logical checks can often be constructed as well. Most of the checks will be very simple (and obvious) but experience shows that they are nonetheless worthwhile.

e.g.

Premiums Date of expiry of cover	-	not earlier than date of commencement of cover.
Types of business included on a single policy	-	compatible with one another (and perhaps with the policy numbering series).
Coded fields		permitted values (e.g. not blank)
Age of policyholder		not under 17 (or over 90 perhaps)
Excess (A.D.)		Zero for non-comprehensive policies.
Car Group		Compatible with make and model code.

Claims

Policy Number	-	not a claim on a non-existent policy (but care needed relapsed policies, late policy advices etc)
Date of advice of claim	-	not earlier than the date of occurrence - and not a date still in the future!
Date of occurrence of claim	-	not earlier than the inception date of the policy, nor a date in the future
Type of claim	-	compatible with the cover (e.g. no A.D. claims on non-comprehensive policies).

What happens when a validity test is failed?

Here there is likely to be some conflict of interest where accounting interests might demand that the premium or claims entry should be included in the regular accounting returns despite an error in statistical coding, and the statistical side would prefer to allow only valid entries into the statistical records. *

If the error can be corrected immediately both interests can be served but the recycling of submissions will probably involve some delay particularly if errors occur in any volume and if the computer files are being updated only monthly, and some workable compromise might be necessary.

Resubmission of errors in some form will be essential and the incorrect entry might therefore be accepted onto a subsidiary computer file, or be accepted onto the main file as a temporary measure before being overwritten or reserved by subsequent amendment. Care should be taken, however, to preserve consistency in cases where the premium itself is subsequently changed - a June entry for £10 which is changed in July to £100 creates an additional £90 premium which must be reflected at some stage in the monthly accounting returns.

In many cases it will be possible for the computer to indicate only that an error is present somewhere; where for instance a premium does not correspond to the coded statistics, the error could lie either in the coding or in the premium calculation, (or both).

An approach for defaulting such cases needs to be considered, and may lead to the entry being accepted in an amended form with some or all of the potentially suspect items being overwritten - generally, however, it will be difficult to default the actual premiums submitted without upsetting the accounts.

Some errors, however, will never be detectable despite the most stringent tests. Even the most exhaustive testing to confirm that a policy premium is compatible with the coded statistics will never reveal a policyholder's age as being 40 and not 50, or the difference between a Rover 2000 and a Triumph 2000 if the premiums for the two risks are identical.

Against this there is no protection and, being helpless to do anything about it or even to really know the extent of its existence in the statistics that are accepted, beyond making the occasional spot check manually, the actuary can do no more than optimistically ignore the possibility of these types of error.

Following a batch of branch submissions some computer listing indicating the outcome of the validity testing will need to be returned to the branch. At the same time as confirming the valid entries have been accepted, this return will also serve as a means of drawing attention to those submissions which appear inconsistent.

* It is probably better to incorporate invalid records in the system using a suspense account or pool of records whose validity has been challenged. This pool should be regularly watched to see that action is taken. Unless clerical

Error messages indicating the reason for inconsistency can then lead to a review of the relevant cases and subsequent resubmissions where appropriate.

Some validity tests will only be able to indicate a potential error which might, in practice, have been a deliberate act by the branch and not therefore incorrect (e.g. a special commission rate not catered for in the data base). Other validity tests may indicate that an error of some form definitely exists.

Not all error messages will therefore need resubmitting and can create a situation where only some of the offending entries are recycled. Recognition must be given to this possibility but at the same time cater for those cases where resubmission is necessary but for some reason is overlooked by the branch.

Different types of error message might provide an answer to this with certain failures being repeated in subsequent validity returns with a reminder that submissions must be made and have not yet been received.

Eventually, when the data base has been finally developed a pilot investigation at a branch is highly desirable. The coding must be explained in terms easily understood and comments from those who will operate the system may well bring to light elements of impracticability or ambiguity less apparent at this stage to those involved closely in the development. Quite apart from the comments this may elicit, the opportunity to criticise and to share in a new system also plays a part in carrying the goodwill of the branches.

Finally, when the coding system has been fully agreed the necessary computer programming finished and the coding instructions circulated, some indication should be given as to when it will be reasonable to expect to derive value from the new system.

Periodic monitoring to ensure that the system is being correctly applied is advisable. Sampling computer submissions is the most obvious approach to this and a small team could be set up for this purpose. It may even prove to be the case that the team pays for itself by correcting errors in premium calculations, since these are invariably biased towards under-charging as the insured will hardly draw attention to this himself.

Monitoring is particularly advisable in the early stages, but all changes need time to settle down and the first sensible information to emerge may well not come for a full year. By this time statistics will provide a detailed picture of the compostion of the account, but further time must pass before the corresponding claims will be recorded. Claims with little delay to settlement will start to provide meaningful figures in the second year of the system's operation, but long-tailed claims involving lengthy negotiations on, say, a large bodily injury settlement will be predominantly estimates to start with and a feeling of anticlimax can be avoided if there is some awareness of the wait that must necessarily ensue before the full benefit can be derived from the system.

MOTOR DATA BASE

We have now considered the elements of a data base for a motor insurance account and outlined some of the practical problems of which the actuary should be aware. These of course will vary from company to company. Essentially, however, a sophisticated data base should allow for direct access to computer records of the recent history of all policies that have been in force at any time during a specified period. The minimum such period that is likely to give the statistician all the information he requires is about six months, but longer periods are desirable. Unfortunately, in motor insurance changes take place so frequently that if the period of retention is lengthened the cost may become large in relation to the value if the period exceeds about two years.

If the statistician were able to specify exactly what he wanted to know this problem would not be serious but experience shows that very often one analysis gives rise to new questions and if historical records have been droppedit may not be possible to carry the analysis over a sufficiently large amount of data to give reliable results. It cannot be emphasized too strongly that very large amounts of data are frequently essential, large meaning at least one million and preferably ten to twenty five million vehicle years.

So far as claims are concerned, information regarding frequency should be kept for the same period as the in force data, but information regarding their settlement should be kept at least until all claims arising in a given calendar year are finally settled - that will be somewhere between ten to fifteen years. With well designed tape systems this is not likely to pose any serious data processing problems, but if the records are used, as they should be, for the production of statutory returns, then it must be possible to reconcile them with the insurer's audited accounts.

The first part of this paper has covered the setting up of a file with validated (but not necessarily correct) data; we must however, be continually alert to the possibility of inadvertently corrupting the data and be able to satisfy ourselves that all are complete at all times Experience shows that this can be a very serious problem. For example it is not unusual for the coding of a field to be changed in some way; in one application a field recording value of the vehicle allowed four digits, but when values in that field began to exceed £10,000 the decision was taken to redefine it as the relevant value divided by 10. However, many records of recently in force statuses were recorded on a separate statistical file and the need to amend this supplementary file was overlooked with disastrous consequences.

Another danger common to all data banks is that data may deliberately be recorded wrongly: a clerk may alter a figure "to get it accepted by the computer", but sometimes a code known to be wrong may be recorded simply because the correct value cannot easily be found and the wrong value will give the right result (at least for the immediate purpose). For example if the choice of a premium scale [that is between two or more scales in force at the same time, but worked on different bases, e.g. cars first registered before 1947, and those registered later], is determined by some date being "before X" or "after X" then if the true value is known to be for example "less than X" any value which is less than X will give the right answer for the time being but may be quite wrong if used for some purpose not originally envisaged.

In theory a well written program should detect impossible conditions but the ideal program has yet to be written and meanwhile eternal vigilance is essential.

We shall now consider the processing of information in the data base for the following purposes namely:

1) To establish that all information which should be recorded has in fact been recorded and that error reports have received proper attention and corrections have been duly made.

2) To check that no records have been lost or mutilated.

3) To test that monetary amounts agree with the insurers' books of account and in particular the premiums received, payments made and estimates raised agree at all times in total with known control figures.

4) To monitor the progress of work coming through the system.

5) To examine trends of new business, lapses and cancellations and to relate them to the levels of factors used for rating and the premiums charged. This is required firstly to examine past experience and secondly to estimate the likely results of our charging and commercial practices.

6) To examine the overall level of claims and their incidence in relation to the level of various rating factors.

7) To examine the absolute and relative level of claim amounts.

8) To examine the reliability of the estimating process and to extend the use of statistical methods for the setting up of reserves for outstanding claims and for unexpired risks.

Under each of these headings there are a number of practical problems; some of which will at least in part be affected by the particular form of data processing system employed including the methods used for updating files and the actual file organisation.

In the first paragraph of this section we referred to the need to retain data for a period. Information on a file at any moment can never in a large portfolio correspond precisely to the cover for which the insurer is on risk at that time.

Insureds may not inform the insurer of changes before they take place (although for a change of vehicle they will usually have to notify in advance in order to obtain a cover note), information given to agents and brokers will not be conveyed immediately to the local branch; thereafter the information will be coded, possibly vetted, punched onto card or disc, validated and the computer updated often at intervals of a week or more and discrepancies exposed by validity tests will have to be investigated and corrections made. Delays of two to three weeks are probably the minimum in many cases but delays of over 12 months are quite common. We shall refer to the information in force according to the file as the file position: the cover actually in force will be referred to as the insurance position.

The shortest delays will arise when a file is kept on disc and updated from terminals immediately on receipt of instructions from policyholders or agents. At the other extreme a tape or card file updated only occasionally, for example just prior to renewal, can clearly contain information as much as 12 months out of date.

The Insurance Position

It is, however, very desirable to be able to ascertain the actual insurance position as at some stated time. Even where the master file is continuously updated this time would have to be at least two to three months prior to the date of enquiry so that information on altered statuses must be stored for at least that time after they have ceased to apply together with a further period equal to the time covered by our investigation. For example if we wish to examine the inforce and claims

for the first quarter of a calendar year we may need to retain information relating to insurances in force at any time during that period until at least October.

The reason for the extra delay is that experience in one office has shown that claims are still being notified more than three months after they have arisen and a count of the first quarters claims in October has revealed 1½% to 2% more claims than were on the file at the July count and the percentages tend to vary with processing and other delays.

It must be emphasized that "retaining information" does not necessarily mean simply keeping it in its original form either on the main file, alongside the current record, or on a separate file of out-of-date statuses. There is, however, no doubt that for most purposes the retention of records in their original form.possibly slightly abbreviated. with a date (starting operative date or S.O.D.) on which they became operative and, if they are no longer in force a date (ending operative date or E.O.D.) on which they cease to apply has considerable attraction. Firstly, all the information is available and secondly it is all in the same format. However, this retention of records in their original format undoubtedly increases the size of master files and to some extent processing times and if either of these limitations cannot be accommodated within the office's computer system or the cost appears excessive in relation to the benefits alternative solutions are available. For example if changes are not very frequent one record might contain information at some inception date (for example the last renewal but two) and for any factors that have subsequently changed the new information and its starting dates. For motor insurance changes are probably too frequent to make this practicable, although it could be contemplated by an office writing a small amount of business but in other classes (for example domestic property) it can be "suitable".

Another method of retaining information is to analyse a record before it is destroyed. A simple example is the analysis of premiums for a DOT return. In this case we know that each premium must be analysed by:

Risk group (say 3 or 4 groups)

Starting operative month (at least 15 periods, but preferably 26 to cater for previous and future years of account and short period covers)

Amount of premium by year of exposure (at least 3, namely this year, next year and last year).

Country
and, optionally at present cover

This gives from 135 to 648 totals per country and there should be separate totals (in motor insurance only) for both the number of days of cover and the proportions of premiums by year. These totals can be accumulated whenever a policy status comes to an end and they can be amalgamated at the year end with similar totals for the statuses then in force.

There is one major objection to this procedure, namely that the user must be very careful to specify in advance exactly what he wants and be able to verify that the accumulating process is working correctly before any data are deleted.

Danger of dropping original records

Mistakes can never be rectified once the original record has been removed and whilst the risk of producing nonsense may be very small it is there all the time; if the information is required for accounting purposes or statutory returns the user must be aware of the risk he runs. What is more it is often found in statistical investigations that one's first tabulations give rise to more questions than they answer so that the statistician will want to try to access the original records for further data. If by then the original records have been destroyed he is apt to be helpless. Hence the procedure outlined above ought to be adopted only if full or sample retention would be unduly expensive or difficult.

INFORMATION THAT SHOULD BE KEPT LONGER

There are two sets of information that ought perhaps to be kept for much longer periods for any policy than any other information on it. One is the N.C.D. status, the other is claims history. Information on N.C.D. should include as a minimum the N.C.D. expected to apply at next renewal and that which applied at the last two renewals. It is often helpful to retain information for at least one or two more past renewals and it is essential that all the information be regularly updated whenever a claim is notified (unless it is known at the time that it will not affect N.C.D.) or whenever a provisional N.C.D. decision is altered. Ideally, these alterations should be carried through to out-of-date records but this may prove to be very complex and not very necessary. The claims history for several years will usually be important for administrative reasons as well as for the statistician; the latter will require not merely the record of claims on the policy but also the full details of the insurance in force at the date of the accident as well as the development and settlement of the claim.

The period of retention may well depend on the type of investigation. If the file is to be used for fundamental research, for example to create standard tables, a long period is desirable. If, however, we are likely to require merely to compare data with some standard then short periods of retention may be quite adequate.

The form in which claim development and settlement information is kept is primarily a data processing problem and the frequency of updating reference and processing are important factors in reaching a decision. It is, however, useful to examine this problem but it is best left until later after we have considered the sort of tabulations we shall or may call for and the frequency with which they are obtained.

Since the analysis of the information in the data base can best be illustrated from an actual system we approach this problem first by describing the system developed in the office of one member of our group which has been found to be very convenient in practice although with advances in the speed of reading and writing on tape it should be reviewed periodically. However, since the present system follows the separate logical concepts of in force statuses, obsolete statuses and claim developments, each of which has its own master tape, it is suitable for our purpose. It cannot, however, be too strongly emphasized that whether we keep three, two or only one physical file and whether they are kept on disc or tape or partly on one and partly on the other and whether they are updated instantly, daily, weekly or at longer intervals is primarily for the Data Processing Department and not for the statistician. Our use of the concept of three separate files in the description is intended solely to aid the reader.

DESCRIPTION OF THE SYSTEM

THE MOTOR MASTER TAPE

This file comprises a list of all policies known to the computer to be in force, together with some policies that have been cancelled prior to the normal end of the period of insurance. In other words this is the file position. Each record consists of a "status" giving details of the Insured and the cover, including some required only for administrative purposes, for example names and addresses of policyholders, and some which might be used by the statistician (for example occupation) if they could be reliably entered and subsequently coded. If any information (other than N.C.D.) is altered a new status is created with an SOD and this date is applied as EOD to the existing status which is then transferred to the supplementary master (see below). A claim is treated as an alteration (with EOD on the record equal to the accident date). The old record is given a distinctive record type (L) so that a count of type L records on the supplementary master gives full information on the number of reported claims and their accident dates. This notification also acts as a signal to output a record (claim rating record) of the insurance position as at the date of the accident which is used to build up the motor claims master file (see below).

THE SUPPLEMENTARY MASTER FILE

This contains all obsolete statuses removed from the motor master tape and they are retained for at least nine months; provision is made for increasing this period to as much as nine years if special investigations are envisaged. In practice it has been found that the number of records transferred to this file in a year is in the order of 50 to 70% of the number on the motor master file.

THE CLAIMS MASTER FILE

This file is normally opened by the claim rating record mentioned above. To it are added the details of all payments and recoveries (other than for the excess of loss reinsurance) estimates, closures, re-openings and re-closures. No time limit is put on the retention, but it would seem, based on experience, to be unwise to drop any records for a year of notification (or accident) until the whole of that year's claims are finally settled. This is likely to be of the order of 10 to 12 years so that with a claim frequency of 1 in 7 this file would ultimately have about twice as many records as the motor master file if the portfolio is stable, although most of these records will be quite short.

THE ORDER OF FILING

Whilst primarily a problem for the Data Processing Department it is helpful to consider one or two problems. The motor and supplementary masters are in policy number order and with minor exceptions, policies are numbered sequentially as they reach Chief Office. This is probably the most satisfactory order since policy number is normally used as the main identifier. The claims master is currently filed by number of claim within year of notification within claim office number. Whilst this was originally convenient before tape systems were fully developed there would now be some merits in sorting by claim number within claim office within year of claim. Neither is ideal but with the development of output on to disc or tape at the end of each office and year for subsequent sorting either by office or by year or both, the difference in advantage is small.

SOME PRACTICAL PROBLEMS

It will be seen that the claim rating details are specified as being those as at the date of the accident. Normally these will be the details on the motor master when the claim is notified to it, but there are times when this is not the case. A change of status may have been unduly delayed in being notifed to the office or in being processed, or the claim may not have been notified until long after the accident (for the office in question one claim in six is notified over a month after it occurs and one in twenty five at least three months later) or its processing may have been held up for want of information or been rejected by the validity tests. Strictly, therefore, the system should search the supplementary master file (probably in all cases) but in the present system this is a separate file and it would involve major changes in the present updating routine. In the office in question these changes would have been impossible when the system was originally set up and reliance on the actual status as found on the motor master (i.e. the file position) is not thought to have involved any serious errors and will not do so unless delays in recording claims data differ markedly from those in recording other factors.

A similar problem arises when a change is notified to the file relating to a date before a claim or when changes come along with operative dates before the SOD of the current status. Exact solutions to these problems are clearly possible, but the great variety of possibilities makes programming complex, and therefore, likely to be unreliable at times. One solution to changes out of order is to amend the operative date to the date after the start of the current status as on the file.

The danger of having an EOD before the SOD are real, the consequences may be trivial or catastrophic depending on how the program treats them (whether or not it has been told to expect them!)

There is also the problem of amending claim rating details in the cases described at the end of the last paragraph. Whilst the system described has provision for over-writing existing rating details it has not been the practice to do so automatically and there is no reason to think that this has seriously affected the results.

FREQUENCY OF UPDATING

The motor master file is in effect updated weekly with information regarding new business, lapses, cancellations, endorsements, claims, claim settlements and N.C.D. decisions.

The supplementary master is updated monthly.

The claims master is normally updated monthly, although the interval is more flexible and more frequent processing is possible when necessary, often at the year end. We would emphasize that this is not the only or necessarily best system and as we have said before it is used solely for illustrative purposes. We can now proceed to show how the statistician can use such a system for the eight purposes referred to earlier. These are now considered in order, being grouped where appropriate.

1. To establish that all information that should be recorded has in fact been recorded.

2. To examine that no records have been lost or mutilated.

3. To test that monetary amounts agree with the Insured's books of account including in particular that payments made and estimates raised agree in total.

So far as the motor master and supplementary matter are concerned this is primarily a master for the Data Processing Department, although the accountant will be involved in regard to the receipt of premiums. The statistician should receive brief summaries of the numbers of records read and written and the number of changes advised to the file as well as a breakdown of the statuses advised to, written on to, and dropped from the supplementary master. He will not be able to check the numbers that ought to be on the file exactly but should be able to see that they are reasonable.

A claims master is a different problem and several monitoring programs have been found essential. The normal order in which details are advised to that file are:-

a) Opened by a claim rating record

b) Note of payment (s)

c) Note of recovery (ies)

d) Note of estimates

e) Note of settlements

f) (b) - (e) repeated if the claim is re-opened.

However, whilst this is the normal order it can easily happen that records under (b), (c), or (d) arrive first and rather than have them in suspense it has been found convenient to allow any of them to open a record. Records of type (e) are never allowed until (a) is present together with any (b) and (c) type records to correspond with the details of settlement advised by the regional claim offices. (Note: Settlement is defined by this insurer as being the occasion when the regional claim office sends its papers to Chief Office for filing and sends a memorandum to the statistician recording the amount for which it was settled the date of "settlement" and the N.C.D. decision. Definition of settlement will of course vary from insurer to insurer since there is no absolute definition of it, but a satisfactory definition should be reasonably objective and independent of other processes, for example any end of year clearing of files). The statistician may need to verify from a preliminary sample that the definition adopted fits these criteria and has been applied uniformly. The tabulations that have been found necessary are:-

a) A list of records that have been opened but have no claim rating record.

b) A list of gaps in the sequence of records.

c) A list of claims which have been dormant for some time (say six months), in case the settlement advice has gone astray or the claim has been overlooked by the claims office.

d) A list of re-opened claims, that is where payments or recoveries have been recorded subsequent to a settlement (Note: These have to be coded by the claims office making them, otherwise the computer will reject them on a claim already "settled")

e) After claims estimating (in this Company on 31st December for a sample only of claims) lists of claims where it appears there should have been an estimate or where one was raised without evident need. (Note: The former usually arise where the settlement has not been notified and not picked up under (c) and the latter are usually re-opened claims where the estimates have not been correctly coded by the claims office

f) Totals of all payments and recoveries analysed by three two-way breakdowns, namely:

Claim office	: year of claim
Claim office	: year of payment
Year of claim	: year of payment

These are compared with total cheques and cash debited and credited to the Company's banking account. Differences are found by claim office or year of claim as necessary, otherwise the detailed breakdown totals are not required (the analysis by claims office would of course not be required in a company where all settlements were made by its Head Office, and the statistician must consider the particular administrative framework of his own office and also the extent to which he can supply the administration with useful information)

g) Totals of estimates recorded on the file analysed by claim office year of claim and year of estimate (for the last four years) which are then agreed with the known totals reported in the Company's accounts for those years. These estimates can be compared with ultimate settlement amounts; a separate note is to be found below.

The frequency of tabulation varies. (f) is normally done after every updating run, but it is vital to see that it is done before any earlier tapes are scratched. (a) and (b) should be done every other month and (c) and (d) quarterly, whilst (e) is desirable immediately after the estimates (or adjustments) have been notified to the file and (g) at the same time as (e) if not more often. This may sound like a formidable task, but groundwork of this nature is absolutely imperative in this work and it is quite wrong to expect the Data Processing professional to do it all; unless he is instructed very carefully or is unusually searching in his enquiries he may well be unaware of all the odd things that go wrong or get omitted in practice.

Relations with Data Processing

The last remark brings us to a problem that is always present but has no entirely satisfactory solution, namely the division of responsibility between the data processing, motor clerical and claims staff and the statistician,and the channels of communication between them which is of particular concern when considering validation. So much depends on the size or the organisation of the office and on the particular individuals concerned that no general statements can be universally true. Usually data processing will be regarded as a service department with the motor department as one of its "customers", to whom they owe a duty of providing a reliable service, but both are liable to regard the statistician as a bit of a nuisance and the quality of the information supplied to him as being less important than that for the accountant or policyholder; this is likely to be especially true of claims clerical staff if they think their work is only for "statistical purposes" and they suspect it is not, or cannot be, checked. Even when the data processing staff do their very best to understand the other parties it has been found that misunderstandings can occur quite by accident. For example, to the statistician "month of last renewal" would normally mean the last renewal dateof an individual policy before the date of investigation, whereas to the data processing man it would probably imply the month of the last renewals that had been processed at that time which could be a date two months in the future!

As mentioned before the nature of statistical investigation is such that one can rarely specify in full at the outset exactly what will be required. So often the first results show up other questions with the result that a stream of amendments and new programs is needed, some of them being one-off, in the sense that once an answer is obtained it may not be needed again. Work of this nature tends to fit badly into a data processing department and the statistician may find himself facing long delays and frustration. One solution, adopted in the office in question, is for data processing to update and process the master file and supplementary master files, using motor administration staff to deal with validity rejects and submission of revised data. For the claims master, however, data processing deal only with the updating of the master and the staff of the statistician are responsible for the submission of data to it, the examination of validity queries, corrections and re-submission and the writing of programs to read and analyse the tape file. The latter is accomplished by means of a short basic assembler program written by Data Processing staff which reads the master file and for each claim record outputs into a COMMON area of about 800 fields of information from the records, in fixed format. This COMMON area can be linked to FORTRAN programmes written as sub-routines. Many such programmes have been written, some of them run monthly and others less frequently. The monthly run can handle up to 15 separate sub-routines which are called using parameters in lead cards; they include all the programs (a) to (g) and others. The overall running time per 1,000,000 claims is about 30 minutes C.P.U. plus times of a few seconds to two or three minutes per sub-routine so that the total C.P.U. time rarely exceeds sixty minutes and the programme being almost wholly process bound the elapsed times may be as low as 70 to 80 minutes.

One great merit of this system is that if special investigations of claims are required they can be provided at very short notice. Such programs have been used to obtain details of claims on policies with

a given number or those with a given agency. If the program is purely temporary the saving in effort in specifying and documenting is considerable whilst regular programmes can also be created at short notice where necessary.

An arrangement of this nature can work well, but as with any form of joint effort by two or more departments, its success depends on willing co-operation and adequate computing experience. Knowing what one wants and where it comes from and the peculiarities of the data can, however, offset a good deal of any of the ill effects of using part time programmers rather than full-time professionals.

Having validated our data base and checked that it is complete we can at last consider some analysis.

4. To monitor the progress of work coming through the system.

Most if not all systems will, as we have remarked before, involve some delay between the happening of an event and its notification to the computer. In part these will be delays in notifying the insurer and in part delays between that time and the moment at which the information on the event is correctly recorded in the computer system. A first step in analysing the data is, therefore, to examine these delays. One way would be to extract the interval from date of change to date of recording and to analyse these intervals into groups sub-dividing the business if different parts are thought likely to suffer different delay patterns. We shall now give a general description of the nature of transaction statistics which we will illustrate by reference to a system similar to that described above which will incidentally consider the extent of recording delays.

TRANSACTION STATISTICS

INTRODUCTORY

Transaction statistics are concerned with the measurement of changes to the insured portfolio. Such information is of use in many areas, for example:

1. Motor Department. Transaction statistics provide basic information for the management of the motor account. They show whether the account is growing or declining , and at what rate. They shed light on the normal structure and development of the portfolio. They may enable us to project workloads, and to identify sources of delay, bottle necks or backlogs.

2. Marketing. Here transaction statistics enable us to measure the effect of changes in premium rates, of advertising campaigns or of changes to the commission rates. This is clearly a pre-requisite to our ultimate aim of being able to predict such changes
   The figures may also help us to direct our advertising to the correct audience, to identify sections of the portfolio where cancellation rates are very high and to show up particular risks where our premium rates are out of line with the rest of the market. (But beware of expecting too much as it is not easy to relate cause and effect).
3. Accounts. Our figures may enable us to help in this area by identifying Agents or types of Agent who are slow to pay, by making projections of cash flow and by providing more accurate estimates of earned premiums and of premium reserve.

4. General. The figures may provide information which is of help in the allocation of expenses. A knowledge of transaction volumes may assist in the design of computer or manual procedures. Finally, these figures may help in other areas of statistics, by showing how long we need to wait before the majority of transactions have been processed.

PROCESSING DELAYS

As mentioned before in some detail it is an unfortunate fact of life that it normally takes some time for a company to find out that it has gone on risk or come off risk, or that a risk has changed. These delays constitute the biggest single problem in the measurement of transaction volumes. The delays to which transactions are subjected are very variable, and different effective dates get mixed up, with the effect that a given batch of computer input will contain a wide range of effective dates. It may even be possible for a transaction to be processed at the computer in advance of its effective date. Suppose we analyse the new business processed in December according to month of inception. We will find some with inception date in December and some in November and so on. If we repeat this analysis in January, February and so on we can build up a picture of new business commencing in December, showing how much of it was processed in December, how much in January, how much in February and so on until no more is received. Such a picture we can call a "delay distribution" or"pipeline". Clearly we could have used quarters or weeks or any other convenient grouping of dates.

Appendix I shows the kind of information which might have been revealed from 12 monthly analyses as described above. Looking at January's new business, in appendix I, we see that although over 85% of it is on file by the end of March it continues to trickle on to the file until nearly the end of the year. One could argue that this form of analysis is unnecessarily complex, and that one could simply count the new business processed in a month, on the grouds that in the long run new business processed has got to be equal to new business incepted. Indeed appendix I shows that in January, February and March, the new business processed

each month is more or less the same as the new business incepted in that month. However, when we come to April we see 895 new business cases processed in April , and yet by December less than 500 cases had been received with the commencing date in April. Similarly, in May the number of new policies commencing in that month is clearly about 500 whereas the number processed is 653. So we see that when the new business rate changes it takes some time for that change to be reflected in the number of policies being processed at the computer. This illustration shows the danger of using simple counts of transactions processed in any given period. If we are going to make sense of these figures, we must work with effective dates.

If the computer file retains effective dates and old policy statuses, we need not carry out monthly analyses but can simply wait until all transactions have been processed, and then count the final number of transactions with a given effective date. This is the only practicable solution to the problem of delays when we are trying to measure exposure. It would clearly be impracticable to establish a delay distribution for every single category within our rating structure. However, the data base may be deficient in this information, or we may wish to establish the shape of the pipeline in order to be able to predict the number of transactions which have been effected but not yet reported to the computer. Since the shape of the pipeline may depend upon a number of factors such as the scope of cover, type of policy or the type of Agent, it is a very difficult matter to decide which categories to use in a pipeline analysis of this type.

Once we have established the shape of the pipeline and whether it is seasonal, we can use it as a predictive model. For example, when a company increases its premiums, it will want some early indication of the effect on its lapse rates, yet we have seen that it will take some months for this effect to be revealed in the number of lapse being processed by the computer. A combination of a pipeline analysis and previously established model, enable us to predict the number of lapses still in the pipeline and hence give an early indication of the lapse rate. Clearly, the same technique can be used to predict future workloads, and future cash flow. We would not of course make the mistake of assuming that the pipelines are of the same shape when measured by cash and by volume.

Another aspect of the impact of delays is that the money associated with these transactions is being reported late to the company's accounting system. This obviously affects the premium debited in the year and the company's estimate of earned premium and unearned premium reserves. Be re-sorting these transactions into effective month, we can refine the estimates of written premium, earned premium and unearned premium reserves. This refinement can be carried out in retrospect once the pipelines have cleared or at the end of the accounting period by using the predictive model. The precise effect of these delays on the company's accounting information depends upon balance between new business and cancellations. Since this balance can vary widely during the year depending on the incidence of re-rates the problem of delays can cause significant distortions in the company's estimate of earned premium for any given year.

LAPSES

It is convenient to distinguish failure to renew, which we can call 'lapses', from cancellations during the policy year. Among the reasons for drawing these distinctions are:

1. Lapses which occur in a particular period can affect only those policies which are due for renewal in that period.

2. The effective date of lapse is implicitly known.

3. Reasons for lapsing are different from the reasons for cancelling mid-term.

Ideally we should like to know the reasons for lapsing. For example the policyholder may be giving up motoring altogether or may be getting a company car. He may have been dissatisfied with some aspect of service, or he may be changing companies to get a cheaper rate. Unfortunately, experience shows that it is almost impossible to get meaningful information in this area. There are many problems in the way of accurate measurement of lapses. Some of the more important are mentioned below:

1. The effect of delays has already been described. This means that any batch of lapses being processed by the computer must first be re-grouped according to effective date before any other analysis is attempted. Normally this will involve either waiting until the pipeline has emptied or making an estimate of the volume of lapses which have occurred but have not yet been reported. However, if it can reasonably be assumed that the shape of the pipeline is independent of the factor or factors being used for analysis, then the analysis can be performed on lapses reported to date and the results scaled up to allow for the estimated proportion of lapses still in the pipeline.

2. The Actuarymay find that a certain proportion of policies are re-written at renewal, and this process is probably carried out by lapsing the old policy and issuing a new one. Provided both the lapse and the new policy are counted, the net affect on a file will be NIL. However, the apparent rate of turnover will be inflated, and if there is any tendency for such rewrites to be undertaken in otherwise slack periods, then the figures may be distorted on this account. The obvious course is to keep separate statistics of such internal turnover. For the purpose of measuring external lapse rates, such internal turnover should be treated as renewals accepted.

3. It may be that a proportion of renewals printed by the computer are wrong. For example, an endorsement or change of branch may have been forgotten, or have occurred too close to the renewal date to affect the renewal. The company's practice may be to return such renewals for deletion and to prepare fresh renewals manually. It is important to ensure that these returned renewals are not treated as lapses and that they are not counted twice in the renewals invited.

4. Another problem is that some mid-term cancellations will occur too close to the renewal date to prevent the issue of renewals. This sort of error artificially swells both the numerator and the denominator of our lapse rates, and we need to be able to identify the issue and subsequent return of such renewals, in order to correct raw data on lapses.

5. Processing delays may prevent some new policies from getting on to the computer file in time for their first renewal. Clearly, this is more likely if the policy was issued for an initial period of less than one year. Such renewals will probably be issued manually and the Actuary should ensure that they and any lapses to which they give rise are counted and included in his main figures.

6. The Actuary should consider the effect on his figures if the wrong policy is cancelled and then subsequently reinstated and the correct policy cancelled.

7. It may be possible for policies to be lost accidentally, or to be removed from or added to the file by special action, possibly to correct a previous error. Such problems may not affect the main figures unless the special action is used wrongly but they will bedevil the Actuary's attempt to reconcile his transaction counts with changes to the number of live policies.

8. The Actuary should establish what will happen if the policy being cancelled is not live when the cancellation is processed. This situation can result in cancellations being counted more than once, and it may be better to count the effect on the file, rather than the transactions themselves.

9. The Actuary needs to decide whether he wishes to count policies, vehicles or money. Unless his statistics are based on vehicles, or unless the file contains only single vehicle policies, the Actuary may find that his transaction counts do not reconcile with changes in his measurement of exposure.

10. If his statistics are based on vehicles, the Actuary needs to consider how substitutions of vehicles should be treated. He should also confirm that if the last live policy is deleted, then the policy itself will be cancelled.

11. If the Actuary follows through the development of a particular batch of renewals, identifying those which have definitely been renewed and those which have been advised as lapses, he may find a residue at the end of the year for which the company has received neither money nor the return of the Certificate. The Actuary needs to decide what to do about such cases.

12. The Actuary may find that lapse rates depend on a number of factors, such as type of policy, cover, N.C.D., whether a claim occurred last year and if so whether it resulted in loss of bonus, and also if the renewal notice was issued before the file could reflect the claim. Suspicions exist that a policyholder might in such a case, transfer to another insurer and preserve his discount by failing to disclose the claim - whilst this is dangerous and immoral there is some evidence in one company that transfer of bonus cases are worse risks than their N.C.D. category would predict.
The Actuary may find that he needs techniques similar to those he would use in the analysis of frequency, possibly culminating in the development of a standard table to minitor changes in lapse rates.

The purpose of all this complexity clearly is to relate lapse rates to their determinants which probably include price or the increase in price over the previous year, commission and advertising, economic conditions and so on. Some at least of these include information external to the company which may limit the precision with which they can be measured. These topics are covered under the heading of 'marketing statistics'.

MID TERM CANCELLATIONS

It seems reasonable to suppose that all policies are exposed to the risk of mid-term cancellations; thus a reasonable measure of the mid-term cancellation rate for a particular month may be the number of cancellations occurring in that month divided by the exposure of that month. It may be, however, that mid-term cancellation rates depend upon the time of year and the time since last renewal. For this and other reasons we probably want to measure the mid-term cancellation rate in any particular month, on each of the 12 monthly cohorts separately. Here the need to re-group by effective date to overcome the effect of pipeline delays results in a very complex analysis.

Many of the problems associated with the measurement of lapses also apply to the measurement of mid-term cancellations. There are in addition one or two problems peculiar to mid-term cancellations:

1. It may be the Company's practice to cancel a policy after a total loss. Do we want to give these special treatment?

2. The Company may deal with certain endorsements such as a change from non-comprehensive to comprehensive or a change of branch, by cancelling the old policy and reissuing another. Again we probably want to keep separate records of such internal turnover.

3. The Actuary should decide whether he wants the end of the period of cover of a short term policy to be treated as a mid-term cancellation or separately counted.

We may reasonably hope that mid-term cancellations will be less dependent on price, commission etc., than our lapse rates. If we can establish that this is so, we can probably make do with considerably less detail in the measurement of mid-term cancellation rates than in the measurement of lapse rates.

NEW BUSINESS

The problems are more difficult here, as there is no obvious base to use to convert the number of new policies into a new business rate. For comparison with cancellation rates we can construct a new business rate as the number of new policies commencing in a particular month divided by the number of renewals accepted in that month. We should recognise, however, that there is no obvious relationship between the number of new policies commencing in a month and the number of renewals invited or accepted in the same month. In framing the above definition we have implicitly assumed that new business is written for an initial period of one year. This is not always the case, and we should keep cohorts distinct. There is also the problem of short period new business which will lapse automatically within the next year.

The problems of measuring new business volumes are similar to those of measuring mid-term and renewal cancellations, and in particular we need to allow for the effect of processing delays. In addition we need to consider the possibility that a policy can be added to the computer file without being treated as new business. This could be done to reinstate a policy cancelled in error, or to take on the portfolio of another company in the event of amalgamation or rescue.

If new business is more difficult than lapses in one area, it is easier in another; we can establish where it came from. There may be merit in at least distinguishing between new entrants to the insurance market and transfers from other companies. Also an analysis of transfers to show the company of origin may help us to make more sense of the relationship between price and new business. It may also be that changes in the number of licences current, or the numbers of new registrations would form a suitable base for the construction of new business rates for new entrants to the market.

GROWTH RATES

So now we come to the stage or putting together the measures of new business and cancellations. Once we have established, either by waiting or by estimating pipelines, the numbers of new business and cancellations, we can calculate the change in the Insured's portfolio over a period such as a month. Because of delays this will not be the same as the change in the number of live computer records over the same period, and indeed these two figures can move in opposite directions. An obvious measure of growth is thus the ratio of this net change, to the number in force at the start of the period.

This seems straightforward, yet Appendix 2 shows a situation in which all transaction rates remain stable yet the portfolio shows a seasonal variation in size. The converse can also occur, that the actual in force remains constant when the transaction rates have changed adversely. The explanation of this phenomenon is that for constant volume, the erosion of each cohort by mid-term cancellation needs to be replaced at the renewal date by an excess of new business over lapses. If the number

renewing is not evenly distributed over the year, this process of replacement will not balance month by month.

We can try to allow for the uneven incidence of renewals, or we can try a different approach. Consider one monthly cohort. The number which commenced a year of insurance a year ago was:-

Renewals invited less lapses plus new business.

This number will have been reduced by mid-term cancellations and increased by new business (written for an initial period of less than a year so as to join this cohort). These survivors will then be offered for renewal. By some means we can establish the number of lapses and number of new cases, and the net total becomes the number commencing a year of insurance in the current month. The difference between this figure and the corresponding figure a year ago can be said to be the net growth, and the ratio of this difference to the earlier figure can be taken as a growth rate. Clearly for this purpose we should exclude short period new business.

Over the long term these two ways of looking at growth must reveal the same picture, but in the short term they can move in opposite directions. The first of these is probably the more useful if we are interested only in changes in exposure but if we are interested in trends, probably related to pricing policy, the latter may be more useful. In some cases the latter may be easier to establish if we are content to infer mid-term cancellations as a balancing item.

There can be difficulties in reconciling these two approaches with each other and with changes in the computer file. One problem may be the occurrence of endorsements changing the renewal date and hence transferring policies into and out of the cohort of interest, without appearing as either new business or cancellations. Another difficulty depending on whether we are looking at the whole portfolio or a sub-set of it is that endorsements (such as change of cover) may be transferring policies into and out of the sub-section we are examining.

Reverting to the system describedearlier in the paper we should explain that there is a monthly run (about the 20th) on which details of policies due for renewal in the next but one month are extracted and on which the new business lapses and other changes processed during the previous four or five weeks are collated and a new master tape written out.

At the same time tab ulations are produced for renewals invited (i.e. those extracted) new business, lapses and cancellations. These are analysed along the general lines already described by several factors including:-

Operative month (except for renewals)

Administrative division

and for private cars, comprehensive and non-comprehensive separately.

Rating area

Rating group.

Age of policyholder.

Age of vehicle.

In addition private car new business is analysed according to whether it is a revival by a former policyholder of the Company, a transfer from another insurer or someone entirely new to insurance.

If we now consider lapses then for month x we issue R_x renewal invitations and in months y (where y will normally start at x - 1 and continue at least up to x + 12) we process $L_{x,y}$ lapses. We shall ultimately have lapsed $\sum_y L_{x,y}$ L_x lapses giving a lapse rate of $1000 L_x / R_x$ per thousand renewals invited (Note the use of "per thousand" . If we work in percentage rates we need to keep one decimal place, but per thousand we can work in integers. There are very few cases where rates per thousand to the nearest unit are too crude or too fine and in those cases we can work by 10,000 or 100.) This lapse rate is a real rate since the lapses relate exactly to the renewals invited and our tabulations will show how the lapses are processed. A typical table can be set out conveniently as shown in Tables 1 and 2. Table 1 relates to the numbers, Table 2 converts the numbers into rates per thousand renewals invited.

Table 1. Lapses - class of business all - numbers of lapses.

Operative month	7.75	6.75	5.75	4.75	3.75	2.75	1.75	12.74	11.74	10.74
Processing month										
10.74										350
11.74									150	3100
12.74								180	1800	900
1.75							210	1600	600	150
2.75						200	2460	1300	210	70
3.75					840	3800	2100	605	40	0
4.75				400	4900	2000	150	40	0	0
5.75			315	3400	3990	120	55	20	0	0
6.75		240	3240	2750	350	30	0	0	0	0
7.75	195	2965	2395	330	120	0	0	0	0	0
Total lapses to date (Nos.	195	3205	5950	6880	10200	6150	4975	3745	2800	4570
Renewals invited (No.)	40000	40000	45000	50000	70000	40000	30000	20000	30000	50000
Lapse rate to date per 1,000	5	80	132	138	146	154	166	187	93	91

Table 2 Lapses - class of business all - rates per 1,000 renewals invited.

Operative month	7.75	6.75	5.75	4.75	3.75	2.75	1.75	12.74	11.74	10.74
Processing month										
10.74										7
11.74									5	62
12.74								9	60	18
1.75							7	80	20	3
2.75						5	82	65	7	1
3.75					12	95	70	30	1	0
4.75				8	70	50	5	2	0	0
5.75			7	68	57	3	2	1	0	0
6.75		6	72	55	5	1	0	0	0	0
7.75	5	74	53	7	2	0	0	0	0	0
Total lapses to date (Nos.	195	3205	5950	6880	10200	6150	4975	3745	2800	4570
Renewals invited (No.)	40000	40000	45000	50000	70000	40000	30000	20000	30000	50000
Lapse rate to date per 1,000	5	80	132	138	146	154	166	187	93	91

It will be seen that the number of lapses processed each month varies quite considerably. The actual numbers being:-

Month	10.74	11.74	12.74	1.75	2.75	3.75	4.75	5.75	6.75	7.75
Lapses processed	9000*	6250*	3080*	2595*	4235	7370	7490	7885	6605	6060

Allowing for columns not given in this table.

The figures given are realistic, although the variations in numbers of renewals per month and the effect of premium increases (and the subsequent approach to the old rates of lapse) are a little exaggerated.

These data which are similar in form to those earlier are given here to facilitate the explanation that follows in regard to lapse, new business and growth rates in sub-groups of the portfolio. Obviously, if we can sub-divide each sub-group by operative month we can proceed exactly as for the whole portfolio. However, this involves the output and retention of very large data volumes and in the system we are describing this was quite impracticable when it was originally designed. An alternative method was adopted assuming that delay patterns were uniform across the whole portfolio; there is no evidence that this is not acceptably near the truth.

If, therefore, we wish to estimate the movement rate among group r vehicles, we may proceed as follows:

let

M^r_{xy} = Movements M for group r, with operative month x, processed in month y.

R^r_x = Renewals for group r, invited in month x.

Also, let

$$M^r_y = \sum_{x=-\infty}^{+\infty} M^r_{xy}$$ (Total movements for group processed in month y)

$$M_y = \sum_{r=1}^{P} M^r_y$$

$$M_{xy} = \sum_{r=1}^{P} M^r_{xy}$$

$$R_x = \sum_{r=1}^{P} R^r_x$$

Then we calculate weighting factors f^y_x from the all-group totals by

$$f^y_x = \frac{M_{xy}}{R_x} \Big/ \sum_x \frac{M_{xy}}{R_x}$$

In practice, if the bulk of the movements are processed within n months of the operative date, it is adequate to calculate f^y_x for $x = y-n$ to y only, $\sum_x M/R$ being summed over the same range of x. (In the above example, it would be adequate to consider only the 6 most recent months for lapses).

The adjusted rate for movement M, for group r processed in month y, is then given by

$$1{,}000 \times M^r_y \Big/ \sum_{x=y-n}^{y} R^r_x \; f^y_x$$

In other words, we regard the movements as relating to a weighted mean number of renewals, the weights being the proportions of movements relating to renewal month x to the total movements processed in month y, from the same range of x.

Calculation of the adjusted rates for new business and for lapses may be done on the above basis. For cancellations, a rather simpler calculation may be adopted because of the wider spread of cancellations over the policy year.

A formula which is suitable for cancellations is

$$1{,}000 \times M^{r}_{y} \Big/ \tfrac{1}{12} \sum_{i=1}^{12} R^{r}_{y+1-i}$$

Depending on the definition of processing month used by an individual company, which will be affected by its own processing cycle, a variable number of processing days may be included in successive months' figures. It may be necessary, therefore, to apply simple scaling factors to each of the above rates to make them comparable from month to month.

The rates for both new business and lapses brought about by this process have proved extremely reliable and sensitive. Care in interpretation is, however, needed since there are cyclical patterns (for example the younger policyholders are noticably less seasonal than the older ones and there is a vast batch of renewals in March arising from the collapse of the V & G which are so different from the normal spread as to cause considerable distortion). It is for this reason that figures for about 15 to 18 months are regarded as vital if the reader is not to be misled.

A typical table arising from these tabulations is set out below:-

Table 3

Relative adjusted lapse rates - private car comprehensive.

Processing month	A	B	C	D	All (rate)
5-74	99	100	110	80	145
6-74	100	102	109	78	160
7-74	100	98	111	79	170
.					
.					
.					
5-75	104	105	95	80	150
6-75	104	105	96	80	170
7-75	105	106	95	80	180

Note: The actual adjusted lapse rate in 7-75 area A is $\frac{105}{100} \times 180 = 189$

These figures which are typical but not actual would indicate[a] set back in areas A and B with an improvement in C.

Tables of this nature are essential to enable us to measure the effect of our pricing and selling practices, but the experience of two offices suggests that it is far from easy to measure cause and effect so that if we are to use the result for forecasting we must do so with great care. Not merely do the effects depend on our own actions; to a large extent they may depend even more on the action (or inaction) of other insurers. At least, however, we have some insight as to what has been happening even if we do not know exactly why it happened that way, we also know fairly reliably just what we do know and, even more importantly what we do not know.

One use of the tables has been to make old style comparisons more meaningful. The earliest figures available may well be counts of the pieces of paper being received for processing, which have traditionally been compared with last year's figures. This is all very well if last year was normal and we know full details of it. If we have a comparable analysis of last year then so long as we are made aware of any hindrances to normal working we may find such preliminary figures useful especially when major changes in rates or rating structures come into effect.

However, we have now reached the field of marketing statistics and it is useful to provide a more detailed survey of this subject.

MARKETING STATISTICS

I Introduction

Rating statistics should provide information about the cost of claims, in total, and for each category, and the analysis of expenses should provide information about the costs of commission and administration in similar detail. However this information is not sufficient to construct a premium table as at least some of the expenses will be fixed and these overheads and the anticipated profit introduce an element of arbitrariness into the premium rates, but there is room for manoeuvre on how overheads and profits are allocated. It is in this area that a company's premiums can (and may have to) react to market pressures, and any relevant figures produced by the actuary to assist can be conveniently labelled "Marketing Statistics".

The ultimate objective will be to obtain estimates of:

(a) the elasticity of demand for motor insurance or for various sub-sections of it,
(b) the effectiveness of advertising and
(c) the effectiveness of intermediaries with special reference to commission arrangements.

There is a clear link here with the Transaction Statistics we have just considered and one can probably express that link by saying that Transaction Statistics measure the effect on the portfolio and marketing statistics measure the strength of the stimulus. To date much less use has been made in insurance marketing of the weapons of advertising and commission than the price mechanism. The principle should be fairly similar, but of necessity the rest of this topic is developed around the subject of price.

II Elasticity of Demand

Before considering, in a traditional way, the elasticity of demand, it is absolutely vital to bear in mind the structure of the total market for motor insurance. The only persons who will buy motor insurance are those wishing to run motor cars. Since insurance is still a minor part of the cost of motoring, changes in its price will not affect, to more than the most trivial extent, the demand for an insurance policy of some sort. The total number of policies therefore may be regarded as fixed externally and not capable of being altered by the effect of marketing strategies. What can be achieved is a transfer of policies between insurers and from less cover to more or vice versa.

In 1969 many insurers seemed unaware of this fact and tried to increase their premium income by offering rather less cover for substantially lower premiums. This was, naturally, and forseeably by any sensible person, followed by a competetive price cutting with disastrous results. Collectively the market had to do a lot of work in transferring policyholders from one company to another, to give no increase in total business with little relative change between companies but with a considerable drop in the total premium income (note; the natural and expected increase in the number of vehicles on the road offset this to some extent but had nothing to do with the price war).

The situation in motor insurance must therefore be very clearly distinguished from many other markets, for example air travel, where major cuts in price can produce very much larger than compensating increases in sales and profits. The motor market is almost inelastic in numbers, very slightly elastic in the amount of cover whilst the airline market is extremely elastic. We now consider the position in general terms according to normal economic theory.

This involves building a model linking volume to price.

A pre-requisite is adequate information about each of these and the former is covered by transaction statistics. The latter involves some kind of index, as rating structures are too complex to permit comparison in each cell. This is covered in the next section , but here we look at the much easier question of what would that elasticity need be in order for a given price rise to produce a balancing loss of volume and so leave the profit unchanged.

Suppose we have a portfolio of N_0 policies at a price of P_0, and we believe we can maintain this volume at this price. The claims ratio we take to be 70% on the current premium and the total claims cost of the policy to be unchanged after the price increase. The commission rate is assumed to be 10% and the expenses to be all fixed so that no saving in expenses can be made if the volume falls. This is not of course entirely realistic but will do for the present purpose. Suppose we want to put the price up 10% and we wonder if we will be better off. If the volume which could be maintained at $1.1P_0$ is N_1, then we can compare the contributions to overheads and profits before and after the proposed change.

At present prices and volumes the contribution is:-

$$P_0N_0 - 0.1P_0N_0 - 0.7P_0N_0 = 0.2P_0N_0$$

At the new price and volume the contribution will be:-

$$1.1P_0N_1 - 0.1 \times 1.1P_0N_1 - 0.7P_0N_1 = 0.29P_0N_1$$

Now $0.29P_0N_1 = 0.2P_0N_0$ $\therefore$ $N_1 = \frac{20}{29}N_0$

In other words we can afford to lose 9/29ths or about 31% of our existing volume before we are worse off.

This can be an interesting way of looking at the problem. If the Company can maintain a portfolio of say 100,000 policies at its present price level, the question is would it be able to maintain a portfolio of more than 69,000 if it increased its premiums by 10%. If the claims ratio had been at a higher starting figure, a greater loss of business could have been tolerated before the position became worse.
For example it the current claims ratio had been 80% then the company would have been better off following a 10% increase in premium provided it could maintain more than 52½% of its current portfolio at the new price.
One must take care however: If a company is losing money on each policy it is obviously "better" to have fewer policies unless the overheads are completely irreducible! Other considerations apply in the case of an office such as an industrial life office selling through its own full-time agents.

This is clearly an over-simplified situation and applies more to long term pricing strategy than short term tactics. Nevertheless if more companies had thought of looking at prices in this light it may be that premium scales would have been on a much more coherent basis than was current during the early 1970's.

III Market Index

There is no problem in constructing premiums for hypothetical risks for any number of companies. The problem is in the vast number of such hypothetical risks which one would need to examine to make a proper comparison.

What is needed is some kind of index, but since rating structures are far from uniform, there are serious problems in constructing a reasonable system of weights. It would be possible to use one's own portfolio, but one would need to re-code some rating factors, and add information which is used for rating by some companies but not others. More to the point if brokers are doing the job they are paid for, each company's portfolio should be heaviest in those areas where its premiums are most competitive.

Thus a company's own portfolio should yield a lower value for that Company's rates than would a combination of the same rates and any other portfolio. Comparison of a company's portfolio with M.R.S.B. statistics might provide some support for this thesis.

At the moment it is likely that comparisons are made in one of two ways:-

1) By comparing rates in a large number of separate categories.
2) By inadequate indices based on a few rating categories and probably with arbitrary weights.

A crude measure of market distribution by each rating factor separately can be gained from M.R.S.B. statistics by those to whom they are available, and these could be used as weights in constructing an index of the rates of a company and of any competitor. However where there is much variety and bases of scales, index comparisons may not be very valuable.

IV Multiple Policies

One of the arguments sometimes put forward against price increases which are otherwise obviously required is the possible cancellation of other policies held by those who hold policies of the type under consideration. For this and other reasons it is desirable to obtain information about multiple policy holdings. Two forms of analysis have been found useful in this context:

(a) An analysis showing simply the number of policies (and their premium value) which are held singly, jointly with one another, in holdings of 3, and so on.
(b) A complete concordance showing each type of policy along the top and down the sides of a square matrix whose elements show the number of policyholders who have a policy of each of the types appropriate to that row and column.

It may come as a surprise to find out how few policyholders have more than one policy with your company. This analysis must, properly, extend to all classes of business not merely motor. It may be thought commercially desirable to run some classes as "loss leaders".

V Agency Statistics

There should be little doubt here about multiple holdings, but it may happen that agents and brokers specialise so that a large proportion of say motor business is through Agents who provide very little else. More to the point, the kind of information which is of interest here is the premium volume of each agent possibly analysed by major types of business. It may indeed be useful to categorise agents and brokers by band of premium volume, so that inspectors and other staff are better able to concentrate their time on agents who provide a worthwhile volume of business.

It seems desirable to consider the quality or profitability of business introduced by an agent rather than its mere volume but one of the problems with this sort of procedure, even allowing that the company can link together the codes for associated agents or branches of national agents, is that agents' accounts are normally too small to yield reliable profit figures. In theory one could work out a level of notional "reinsurance" for each agent, above which losses would not count towards his results. Of course some part of the premium would need to be appropriated notionally to pay for this notional reinsurance. In effect the comapny is saying to each broker that his account will be too variable on a year to year basis unless his losses are limited in any particular case to some specified value, and that the company will insure the excess above this value for

a specified premium.* An alternative to this type of stratification would be to fix a lower premium volume (depending on the portfolio mix) below which profitability would not be measured. In order to bring smaller agents into the scheme they would need to have several years experience amalgamated.

Similar arguments apply to large fleet risks which are experience rated. The group has a suspicion that in this area underwriters often make inadequate allowance for the costs of very large claims.

VI Other Statistics

Other areas where the actuary or statistician may be able to help are:-

1. By comparing the distribution of certain rating factors in his own portfolio with those in M.R.S.B. data. This comparison will be more meaningful if he can add a comparison of his own rating structure.

2. Make use of Government or other published figures to help assess market share and potential markets.

* This is similar to top-slicing as discussed on pages [redacted].

65 to 67

THE MEASUREMENT OF EXPOSURES AND CLAIMS

At long last we are approaching the point at which we can do a real statistical investigation. The fundamental aim is to measure expected claim costs and to seek to establish whether there is a relation between the levels of various factors that might seem to affect the risk and the relative expected claim costs. The factors include the traditional ones of the size and type of vehicle, the place where it is normally kept, the use to which it is put and the characteristics of the policyholder or regular driver.

It should be made clear at the outset that we are looking for associations between rating factors and claim costs (or "risk" for short) not necessarily causal relationships. An example might be colour of vehicle; suppose that extrovert drivers were found to be bad risks and to be particularly likely to purchase gaudy vehicles, then colour of vehicle might be useful as a rating factor. Whilst this is perhaps a little far fetched, there is at least a suspicion that age of vehicle may be a less unreliable indicator of risk than mileage. In the language of the economists one readily ascertainable fact may be a good proxy for another fact which is the true source of risk but which may be incapable of accurate assessment

A list of factors that we may readily examine will be limited in general to those from which reliable information is available on the master tape. If we wish to examine experience by some other factor we must make special enquiry either of the whole file or a large sample in which we can compare the levels of the factors or, if we are able to forcast expected risks, we may select only those cases with particular levels of the factor in which we are interested. The techniques for such estimation those described by Johnson and Hey (1). The fact which we are likely to have recorded on the tape and which are reasonably liable to be accurate (if only because they affect the premium calculation) are:-

Make, model and cubic capacity of vehicle (or plated weight for goods vehicles).

Date of first registration
Place of Garage
Class of use

N.C.D. Category and/or claims history

Age of policyholder (and possibly some note if young drivers other than the policyholder are likely to drive the vehicle)
Restrictions on driving
Excesses
Other Covers
Special Loadings

Other information which might be available, but which is likely to be difficult to verify, even after a claim, include:-

Actual or expected mileage.
Use for commuting
Use for business purposes, rallies etc.

In some countries, for example U.S.A., premium may depend on the sex or marital condition of the policyholder and for the younger ones it may even depend on school or college records. Little thought is needed to realise the difficulty in maintaining this sort of information up to date even if it is used in rating, unless premiums are revised immediately on any change or notification is made a condition of continued cover. As a rating factor occupation presents problems since it is difficult to define and code in a way that might seem to be related to risk, although most insurers enquire about certain occupations which are felt to be particularly hazardous, although whether these are associated with additional risk, moral hazard or underwriter's prejudice is often hard to discover.

In view of the small numbers often involved the most we can probably hope to achieve with one company is to see if cars loaded by underwriters whether for occupational risk or other reason, do on the whole seem to merit the loading applied. However there are serious difficulties involved as will appear later. Some companies give special rates to members of certain professions (civil servants and teachers get lower rates, doctors and insurance outside staff higher) but a company not currently doing so will have great difficulty in testing whether the practice is justified simply through lack of information on exposure.

Finally there are characteristics that may well affect the risk but are incapable of measurement except possibly by proxy. There include moral hazards, medical conditions, temperament, happiness and so on. The amount of risk unexplained by conventional rating suggests, that collectively, factors of this nature are quite important and, for reasons explained in the Johnson and Hey paper, can only very imprefectly be measured by examining the claims experience.

THE ACTUAL MEASUREMENT OF EXPOSURE

If we were examining experience by only one factor, say private cars by age of policyholder, the situation would be very close to the measurement of exposure under one class of life insurance, say endowment assurance on male lives. There is one major difference and it is vital to remember it however. In life assurance exposure is likely to be classified by age last birthday on the census date, whereas in motor insurance, where the premium is normally based on age last birthday at renewal (even if the policyholder has his birthday on the following day and therefore, is one year "older" all the remaining 364 days), age at any point, whether on census date or the date of an accident should be taken as age last birthday at the renewal prior to the event. Other definitions, if consistent might be used, but since inter alia, we want to measure risk in relation to the premium charging process the above definition is likely to be much the best. The same rule must apply to all other factors, so that our risk measurement is in line with the premium charging process.

The actual definitions may therefore need to depend on the practice of each insurer. Some companies will amend the premium rate on any change which brings the vehicle into another rating category such as change of place of garage, change of vehicle, change of use and so on. Others may continue cover for the rest of the period at the old rate. In our investigation, one possible approach is to follow our company's practice, so that if a change in rating area is ignored until next renewal we shall define rating area as that in force at the last renewal (although this involves the assumption that such changes are uniform from year to year); if the premium is altered we shall define rating area as that which applies at the date of any event (census or claim). The latter will give a better indication of the effect of an area, which has considerable attraction but will not so readily be translated into a premium scheme as practised by that company.

It seems that N.C.D. will always be that applying at last renewal, but problems arise with adjustment of N.C.D. decision some time after the claim date. Once again, so long as we are consistent, it will not matter too much, although the practice adopted by some insurers of an instant, but provisional N.C.D. decision will help since it means that subsequent changes should be fairly few in number.

Let us now consider the simple case of a single rating factor, namely age of policyholder. Experience tells us that the"calendar year" type of method is essential. Whilst it is customary to use an actual calendar year any period of time would be satisfactory so long as all policies are exposed to the whole of the period chosen or for such part of it as they were in force in; for example, experience can be built up quarter by quarter and can be merged into summer (May to October) and winter (November to April which have quite different types of experience. One common method of obtaining exposure is the census method, counting at the start and end of each period (at least quarterly, if not more often) the policies in force classified by age next birthday at renewal (or inception) prior to the census date and the number of claims classified by age last birthday at renewal (or inception) prior to the accident date. If the file is counted about three to four months after the end of the quarter it is likely that most alterations and claims will have been recorded and the results will be reliable. Alternative methods are:-

1) Before a status record is removed from the file it should be counted if it was in force on a census date which is going to be investigated in the future and the totals of such statuses should be accumulated and added to the count of the current statuses at the end of the quarter where those statuses were in force at the census date (this last count can be made only some time after the end of the quarter).
2) To count the number of days which each status has been in force during the period of investigation (which need not now be as short as a quarter year) by the age (or ages if it has been renewed during the period) according to which it has been classified. This can be done either on a file containing the necessary retained statuses or by the method described in 1) above.

Any of these methods will give the number of claims Cx at age x and the corresponding vehicle years of exposure at that age Ex, from which a claim frequency may be derived, namely

$$1000 \times \frac{Cx}{Ex}$$

This is a crude rate and is liable to the same sort of distortion as crude death rates. We must consider how to deal with these distortions.

It has long been clear that there are associations between the levels of various rating factors, some of them quite strong. For example, a 17 year old policyholder is unlikely to have earned any N.C.D. (Note that we say unlikely,not impossible - he might have owned a car at a much younger age but have been excluded from driving it and have employed a chauffeur: it is very difficult to visualize circumstances of this nature which could never arise) whilst a 50 year old is very likely to be on maximum N.C.D. (about 90%; with more than 50% having 9 or more years claim free). In the case of motor cycles the association between age and N.C.D. is so strong that it is almost impossible to separate the effect of the two factors. Other associations noted have included age of vehicle with cover (strong) and with rating area and age of policyholder (weak); also with age of policyholder,voluntary excess (strong) and driving restrictions (moderate). It should, therefore, be clear that some steps have to be taken to deal with these associations. Just to give one example it was shown clearly in a large pooled sample that ten year old vehicles insured for third party only, gave rise to a higher claim frequency than ten year old vehicles insured comprehensively. The reason is the higher proportion of young policyholders in the former group.

The basic method for dealing with this situation is described in the paper by Johnson and Hey referred to earlier. Briefly this consists of a standard table which expresses the expected claim frequencies for any policy as the sum of 9 or so parameters, one for each level of each factor thought to be associated with the risk. These parameters are used in two ways, firstly to estimate the expected number of claims in a group, secondly to adjust the actual claim frequency in the group so as to eliminate the distorting effects of factors associated with the one under investigation. This will give a relative claim frequency for that factor alone; these are similar to independent rates of decrement in a multiple decrement life Table.

The method was developed by observing 3, 4 and 5 - way breakdowns and noting that the claim frequencies seemed to follow quite closely an additive pattern. It is thought that the process is very reliable for private cars but for motor cycles and commercial vehicles, where the variation of risk is very much greater, the purely additive model with a single table of parameters is not entirely satisfactory.

It will be seen that we have referred to relative frequencies. It is known that there is a marked seasonal pattern in claims experience, fewer claims in summer and more in winter, whilst the parameter table gives an overall figure for the year. Practice has, therefore been to multiply all numbers of expected claims by a factor so that the expected number for the whole portfolio equals the actual number. This gives a measure of quarterly experience, and trends in quarterly factors can be most informative.

It might be argued that we should add a constant rather than multiply by a constant: it is not intuitively obvious which is the better although a multiplier seems to have worked well in practice.

In a small account detailed investigation of the claim frequency by each factor is probably not possible. However if a standard table is available an overall frequency can be derived and the relative frequencies for each factor calculated assuming that they follow the standard table. This technique is broadly similar to that used in the examination of small life or sickness insurance funds.

Having described in outline the sort of tables we wish to produce we now go into a little more detail on the methods used to produce them in the office system previously described. This was set up using the census method and unless there are major discontinuities in the flow of business (for example following the collapse of another insurer) seem to have been thoroughly satisfactory and to have provided a useful by-product in the form of trends in the number of policies in force in various categories (Note that the methods of Tables 1 and 2 et seq are deficient in one major respect, in that they do not deal with endorsements, and in particular mask a substantial transfer by endorsement (or "cancelled by new" which are ignored) from non-comp. to comp. cover or vice versa.

The procedure adopted is to examine once a quarter the master and supplementary master files and to select every status that was in force on the relevant census date (about 3½ months ago) and every claim in the quarter ending on that date. For each such status we calculate the expected claim frequency using our standard table of parameters, the net premium (using some uniform scale, preferably the one that was in force throughout the quarter or throughout most of the year to which the quarter belongs) and accumulate the totals of these quantities and a count of the number of policies by 23 separate breakdowns. The output is taken on tape and accumulated for a calendar year. Each quarter there is produced an analysis in the form shown in Table 3 for each of the 23 breakdowns for both comprehensive and non-comprehensive separately and smaller analyses are made for other classes of business. Half yearly and yearly analyses are also produced, the former in October and April and the latter in April relating to the whole of the previous year.

In addition to this information on claim frequency the system stores a further standard table of parameters for the expected amount of claims which it uses in the same way as the parameters for frequency. The intial derivation of the amount parameters was by a very elementary method but as experience has been accumulated, evidence of the average amounts of claim has made it possible to test, and, where necessary, amend,the parameter table. These tables are described later in the section dealing with claim amount.

THE MEASUREMENT OF EXPECTED CLAIM COST

We have so far tacitly assumed that we should investigate the claim frequency and claim amount separately, and it is now essential to examine the basis for so doing. Traditional practice was to compare claims cost with premiums collected (the claim ratio). The objections to this are many and the most important are:-

1. Unless the results are to delayed for several years they must depend to a greater or lesser extent on estimated outstandings. This is a notoriously unreliable process and if, in order to maintain the right overall level of cost, adjustments to estimates are required then fluctuations arising from the adjustment of past errors as well as current errors in estimating are likely to produce major errors. Further the claim ratio was often computed on the basis of a stable portfolio i.e. claims paid and outstanding in an accounting period divided by premiums received in that period.
2. The premiums actually charged, even within one company, are likely to be a mixture of more than one scale and where frequent adjustments of scale are coupled with the existence of a discretion to local staff to quote special terms, then total premiums are likely to be a poor indicator of actual risk and, particularly, changes in premium may be an even worse indicator of changes in risk.
3. Except in very large groups the incidence of individual large claims is likely to cause the claim ratio to fluctuate substantially.

It was thought during the 1960's that the best way to avoid all the difficulties was to ignore the actual premiums charged and to measure the risk in some other way and this approach has still much to commend it. At the same time the basic process could be analysed by examining separately the probability of a claim and the distribution of the amount of the claim once one had happened which really assumes that there is no correlation between frequency and amount. On the assumption that expected total claim amount (at least expected total relative claim amount) is relatively stable (apart from the effects of inflation on the actual overall average) then one can use amount data based on other more fully developed cohorts, where the claims are sufficiently developed to give a reliable answer, coupled with up to date claim frequencies.

If this hypothesis is justifiable it makes statistical investigation much easier. At present it is known, or at least strongly suspected, that some influences cause claim frequencies to move in one direction and the average amou ntin the opposite direction, giving a smaller proportionate overall movement in total cost than in either component. Whether different groups behave sufficiently differently as to upset comparison based on current relative frequencies and historic relative amounts is not clear. Where caution is most needed, however, is in assuming that increasing claim frequency necessarily means an increase in claim cost or vice versa.

Pursuing this line of investigation we may use our standard table of estimated claim costs allied to the actual number of claims occurring to calculate an expected total cost for all the claims in the group. We also have the premiums for the total exposed to risk in this group (on a selected scale which is used throughout the year) and obtain a calculated claim ratio from these two figures. Since absolute values of claims and the absolute level of premiums earned in the period are not readily available we express all our claims ratios for every group in terms of the overall claim ratio calculated on this basis and multiply the result by 1,000.

So if we see a ratio so calculated that exceeds 1,000 it means that the group is being relatively undercharged whilst if the ratio is below 1,000 it is being relatively overcharged. No allowance is made in this calculation for expenses that are not exactly or nearly proportional to the risk premium.

We can now set out a table showing the presentation of some actual results: we have chosen a voluntary excess as it illustrates many of the points we have been making.

TABLE 4

MOTOR CLAIM ANALYSIS FOR THE 4TH QUARTER 1974 - BREAKDOWN BY VOLUNTARY EXCESS

EXCESS £	EXPOSURE FINAL	EXPOSURE TOTAL	TOTAL PREMIUM £000	EXPECTED CLAIM COSTS	RELATIVE C/R	CLAIMS ACT.	CLAIMS EXP.	A/E.
0	326550	82120	2784	1296	983	11449	11481	997
5	23112	5861	165	68	863	646	721	896
15	132854	33259	1055	521	1042	4733	4644	1019
30	56939	14149	507	253	1051	2178	2160	1008
TOTAL	539455	135389	4511	2138	1000	19006	19006	1000

EXCESS £	RATES ACT.	RATES ADJ.	CLAIM COSTS	AV. PREMIUM	"GROSS" * PREMIUM
0	139	142	113	33.9	33.9
5	110	125	105	28.2	31.8
15	142	136	110	31.7	35.6
30	154	131	116	35.8	42.7
TOTAL	140	140	112	33.3	-

This is the premium before the discount allowed for the relevant excess and gives a crude measure of the relative risk.

Some comments on this table from the Company that provided it may be useful:-

1) The £5 excess is no longer offered except to policyholders who had it in 1971 or earlier. The discounts for £10, £15 and £30 in 1974 were about 11%, 11% and 16% respectively. The excess levels are no longer appropriate.

2) The final exposure is simply a count of the in-force as on 31 December 1974. Trends in these figures are most useful.

3) The total exposure = 1/8th final exposure + 1/8th exposure as at 30 September 1974.

4) The total premium is calculated on an old scale which had subsequently been increased by flat percentages. The absolute level, in this tabulation, is of no importance so long as the correct relativities are used. This premium is 1/8th of the total of the premiums on the in-force at 30 September and 31 December 1974 which comprise the exposure.

5) The expected claim costs come from the standard table. The units are not given but in 1975 are probably of the order of £1500 to £1700. Like the absolute level of premium this figure has no effect on the tables.

6) The relative C/R is $1000 \times \left(\frac{E}{P}\right)_{group} \div \left(\frac{E}{P}\right)_{total}$ where E is the expected claim cost and P the earned premium for the group total.

7) Under claims, "ACT" is the number recorded in the quarter, "EXP" is the value from the standard table multiplied by a factor to make the totals of ACT and EXP equal to each other.

8) A/E = 1000 x actual/expected; it will be seen that except for the £5 excess the ratios are very nearly equal to 1000. The short-fall in the £5 excess is partly a random fluctuation and partly a result of the parameters not allowing quite enough for this rather unusual group, the average for the 12 quarters has been about 950 but it has been slowly falling as the group gets more and more extreme.

9) Rate "ACT" = 1000 x actual number of claims ÷ total exposure. Rate "ADJ" = ACT $\pm$ an adjustment calculated from the standard table to allow for the make up of vehicles insured in this group.

10) Claim cost is an average in units of 1/1000th of those for total claim costs = expected claim cost ÷ claims ACT

11) Av. Prem = total premium ÷ total exposure. It is the same sort of £'s as the total premium.

12) It will be seen that total exposure is a little greater than ¼ of final exposure indicating a slight growth in the portfolio over the quarter.

13) The relative claim ratio is much below 1000 for the £5 group. This follows from relative overcharging of some older policyholders or those driving older cars. For the £15 and £30 excesses the relative C/R is above normal, as a result of slight relative undercharging of younger policyholders (to whom the larger excesses have the greater appeal) and the fact that the premium scale assumes that the administrative cost is a smaller proportion of the premium than for the whole portfolio.

14) It will be seen that the actual claim frequency for the £30 excess is greater than for the "no excess" cases (154 compared with 139) in spite of the fact that the existence of the excess will mean that some claims will not even be reported. However the adjustment gives the adjusted claim frequency for the £30 excess 11 points lower than for the no excess. Similarly it will be seen that the expected claim amount for the £30 excess is higher than for the nil excess; this is in spite of the fact that the parameter for no excess is plus 2 and for £30 excess minus 9½.

The differences are accounted for by the fact that the £30 excesses tend to be taken by the higher risk cases and especially by the younger policyholders who have the highest average claim amounts. This is in line with the average premium in the £30 excess, namely £42.7 before allowing the 16% discount compared with only £33.9 for the cases without excess. By contrast the £10 excess, which will apply to only long standing policyholders has an equivalent gross premium as low as £31.8 and a very low actual claim rate as well as a low adjusted rate.

Tabulations of this nature provide a very satisfactory test of the performance of the portfolio, and if the claim ratios and A/E are observed over a period of years it is easy to see at a glance whether or not the situation is under control.

Tables can be provided for any factor or combination of factors that are recorded on the file and the current processing schedule provides for private cars, 23 breakdowns by single factors by fairly fine subdivisions (for example the 146 rating districts under the Motor Conference recommendation and 24 age groups), all possible combinations of 7 factors two at a time in broader groups of not more than 6 levels per factor, as well as, for administrative purposes, an annual tabulation by each of over 200 offices through which policies are sold.

It should be noted that throughout this section the practice has been to analyse the claims by date of accident, not by date of report. This is, of course, theoretically correct, but we have previously referred to the need for some checking by recounts at a later date. It is however most unlikely that such recounts will affect the relative ratios.

THE ANALYSIS OF CLAIM AMOUNT

The problems here are quite different from those associated with claim frequency. For one thing the question of getting the rating factors right, whilst important, is not so vital. With claim frequencies it is absolutely imperative to make sure that the numerator of the frequency (the number of claims) and the denominator (the exposure) are calculated so as to correspond precisely. With amounts, however, the effect of moving a few claims from one group to another will be marginal. For example if we have two groups (1 & 2) with mean claim amounts of 100 and 110 the position is:-

	CORRECT			50 CLAIMS OF EACH KIND WRONGLY CLASSIFIED		
GROUP	NO.	AMT.	AVGE.	NO.	AMT.	AVGE.
1	1000	100000	100	1000	100500	100.5
2	500	55000	110	500	54500	109.0
ALL	1500	155000	103.33	1500	155000	103.33

Since 50 is 10% of the smaller group it will be seen at a glance that the risk of drawing wrong conclusions from anything but gross errors in classification is likely to be remote. The need, for example, to go back several years to correct for a reversal of a N.C.D. decision is not great and this is probably the most serious single source of error. Its effect will probably be to slightly understate the average for the nil N.C.D. group, as these are most likely to be revised into a higher group with a lower real average. On the face of it all we need to do with claims is to keep our claims tape up to date and ensure that the amount data are reliable and agree with books. We then merely need to add up the total cost in each group and divide by the number of claims which have been added into that group.

There are however four major problems, namely:-

1) What to do about claims not yet closed (the problem of defining closure or settlement has already been considered in an earlier section)
2) What to do about large claims.
3) What to do about associated factors which produce effects similar to but usually smaller than those arising with frequency.
4) What to do about inflation.

One system adopted has been to produce at the end of each quarter an analysis of average amounts of claims then settled at least once. The treatment of claims re-opened is a matter for individual choice but it is probably best to adopt one of two possible courses.

1) To ignore altogether transactions not included in the first closure and subsequently to make a separate analysis of all such payments and recoveries. If the amounts are large enough one can investigate whether they arise in a random manner or not. Much will depend on the current practice of the office in regard to the speed of the original settlement. In the office in question it was found some years ago that about 1 claim in 10 was re-opened at a cost of about £3½ extra, that is 35p per original claim - then about £80. Recently this seems to have increased to over £1 on an original £180 but for most purposes the amounts are too small to matter. Much will depend anyway on the definition of "settlement".

2) To replace the original settlement by a later settlement, ignoring any extra payments or recoveries until the later settlement is actually recorded. This has the merit of giving a truer average settled claim and a truer claim amount distribution.

The progress of overall average settlement shows a steady rise in the average amount. The precise curve these averages follow will depend, amongst other things, on the rate of inflation and the mix of business (comprehensive cover should, however, be kept separate from non-comprehensive at all times even though the various classes of non-comprehensive can probably be treated as a single group). A typical rate of increase of average settled claim for a yearly cohort by date of notification, NOT by date of accident, in quarters, quarter one being at the end of March of the year of notification, and the rate of settlement for private car comprehensive is:-

TABLE 5

QUARTER	AVERAGE SETTLED CLAIM AS A % OF ULTIMATE AVERAGE	PERCENTAGE SETTLED BY NUMBER	PERCENTAGE SETTLED BY AMOUNT*
1	27	6	1
2	35	19	7
3	42	36	16
4	47	59	28
5	52	78	46
6	57	86	55
7	64	91	60
8	70	96	65
12	77	98.7	78
16	85	99.4	84
20	91	99.7	91
24	93	99.8	93

* Excluding payments on claims not treated as settled. The actual amount PAID will be greater, often much greater.

ϕ That is when only about one quarter of the cohort has been notified and less than a quarter of these notifications has been settled.

(In this table the average settlements are the average monetary amounts so that the effect of inflation has been left in. An alternative approach would be to express the settlements in constant pounds by choosing a base date and dividing each subsequent settlement amount by the ratio of the Retail Price Index of the settlement date to the base date. This concept is expanded later)

It will take many years before it can be shown that this type of table gives reliable estimates of final average settled claim. Not enough is yet known about the behaviour of the figures, although in times of increasing inflation it is likely that the figures in the second and fourth columns will decrease slowly. There will of course be fluctuations in the results but it seems reasonable to expect that unless there are major changes in the mix of the portfolio or in the rate of inflation, the relation between the second column and the third and fourth columns will be reasonably close. If this is true, then knowing:-

(a) The amounts paid on claims already settled

(b) The percentage by number of claims already settled

(c) The amounts paid on outstanding claims

might be sufficient to make a reliable estimate of the amounts still to be paid on the outstanding claims. This method should be subject to fewer distorting influences than assuming that the build-up of payments over time follows a fixed curve, which is the basis of grossing up on chain ladder methods.

Unfortunately, the practical difficulty is that after the end of quarter eight the number of claims outstanding is very small (1¼% at three years, ½% at four years, 0.2% at six years) whilst the amount unpaid is still large (22% at three years, 15% at four years and 7% at six years) so that small errors in the number settled can cause large errors in the estimates of the percentages paid. It would also be extremely difficult to detect by statistical methods any artificial distortions in the pattern of settlements.

An office is principally interested in the assessment of ultimate claim amounts for two distinct purposes. The first is for setting future premium scales and here abnormal fluctuations in the contribution of large claims or other abnormal features (such as a surge of minor or major claims arising from exceptional bad weather) should be ignored, except to the extent that they need to be spread over several years. For this purpose an estimate of claims cost can best be made after two or three years. Any benefit from waiting longer and thus having more settlements is offset by the additional waiting time and by the disturbing effect of large claims.

The second purpose is for accounting and solvency. Here, abnormal features have to be taken into account in the year of occurrence but must not be allowed to over-influence reserves for later years of claims, nor future premium scales unless there is good reason to suppose that they are likely to be repeated. The problems of solvency estimating outstanding claims and large claims are inextricably mixed up with each other.

In the office referred to previously it is the practice for claims officials to individually estimate as at 31st December annually all claims more than two years old, and so from quarter 12 onwards projected average claim amounts on the basis of settled claims plus payments on account and estimates of the outstanding claims may be calculated at yearly intervals. It is not obvious that the result is any more reliable than working on the proportion settled and the corresponding ratio of average settled to average ultimate.

Analysis of the outcome of claims that have been individually estimated has revealed large savings on claims settled during the year after estimating, but large deficits on claims which have been re-estimated a year later. Apart from the disturbing effect of inflation this variation occurs mostly where third party bodily injuries are involved. It may take several years for the extent of the injuries and the degree of subsequent recovery to become clear and there may also be dispute over liability. Ultimately the Court may be required to give judgement on both counts. For these reasons it would be wrong for each claim to be estimated at its maximum potential liability since gross over estimating would result in total. The examination of estimating practice is considered later.

Except for one important aspect, estimating will in no way affect the outcome of past business or the present true solvency margin unless "profits" have been prematurely distributed. In this, the situation is analagous to the emergence of surplus in life insurance where the rate of emergence has no effect on the ultimate amount unless there also it leads to premature distributions of apparent "profit". The exception is the psychological behaviour of claims staff who may feel they are doing their job well when they settle claims for less than the last estimate. This may resulteither in economical settling, or over-estimating followed by settlement within the estimate, but still more than it might have been. This is an area where it is difficult for the statistician to monitor progress reliably. Comparing branch claims sections with each other is one method but requires great care due to variations in mix of business, claims consciousness of the local population and the major variations in claim amount by geographical area. There is also the knowledge that the statistician is measuring performance against the estimates.

At first it may appear that estimating is an essential process in maintaining the solvency of an insurer, and in helping him to fix adequate premium scales for the future. It is not, however, certain that this is so. The use of the relation between percentage settled and the ratio of average settled to ultimate average has been mentioned. A similar but more sophisticated method is being tested in one office.

This makes use of a table showing, for each of five sections of the business (private motor comprehensive, private motor non-comprehensive fleet, other comprehensive, other non-comprehensive) the expected build up of the number of payments and the proportions expected to have been paid at each point in time monthly over five years. This table is used to calculate month by month the expected number of payments and the expected total amount of payments analysed by year of claim. The ratio of the actual number of payments to the expected number of payments and the actual amounts to the expected amounts can then be calculated. The latter effectively gives the average ultimate claim in pounds. Preliminary results are encouraging and since the method is not likely to be very sensitive to changes in the mix of business it would give quite reliable indicators within a year of the closing of a cohort: much more experience is required and since this will reflect the inflation of 1973 to 1975 it should be a good test of the method.

If a process of this nature works it may help solve the problem inherent in all estimating of finding out what hasn't been done. Since what hasn't been done, with claims over two years old, represents perhaps one or at the most two claims in 100, this is bad enough for the honest insurer, it is worse for the auditor or supervisor who can never be sure that ninety and nine sheep safely filed away as "closed" do not in fact contain the black one that is merely lying dormant.

Efforts to make and verify estimates by mechanical processes have been in hand for some years, starting with Scurfield (3). In the hands of an honest insurer who knows what he is doing and is aware of the dangers they have saved a lot of pointless work. Without that special first hand knowledge, however, they are liable to be very dangerous processes indeed. If we can project claim frequency from recent experience and obtain fairly recent average claim amounts and project them in the light of assumptions on inflation etc., then we can get and test premium scales on a basis that can be verified by supervisors and other outsiders, and they can apply control at the proper stage, namely to ensure that an insurer collects adequate premiums for the risks he assumes (he must of course preserve the money but that it is not our concern).

EFFECT OF INFLATION

Let us now turn to the effects of changes in the rates of inflation.

Model Distribution

Let us assume for simplicity that following a long spell of steady conditons with no inflation we find that a portfolio produced 1024 claims in a year, costing £131,072 (that is £128 each) and that payments are made on the 30th June in the year of claim and in each subsequent year in accordance with the following table.

YEAR	AMOUNT PAID
c	65536
c+1	32768
c+2	16384
c+3	8192
c+4	4096
c+5	2048
c+6	1024
c+7	512
c+8	256
c+9	128
c+10	128
	£131072

Let us now make the following alternative assumption:-

1. Inflation starts on 30th June in mid-year c at a steady rate of 6% pa indefinitely.

2. As in 1, but inflation becomes a steady rate of 25% from year c + 6.

3. As in 2, but the 25% rate lasts for only two years, after which it reverts to 10% and thereafter 6% pa.

4. Inflation starts on 30th June in mid-year c at a rate of 5% pa increasing by 5% pa indefinitely (i.e. c+1 to c+2 is 10%, c+2 to c+3 is 15% and so on).

Risk Premiums

Now let us calculate the amount that will be paid in years c + n up to c + n + 10 on 1024 such claims notified in year c + n (n = 0,1, ...8) assuming everything else is unchanged and let us look at the risk premium we require, assuming that

1) a premium scale applied for 12 months commencing 1st July until the following 30th June,

2) since claims occur over that year and the following year they are regarded as happening all on the 30th June at the end of the period during which the scale has been in operation,

3) the premium year c relates to the period c - ½ to c + ½.

Then the risk premiums required are - (in £000)

	Assumption				Percentage increase on previous year			
premium year	1	2	3	4	1	2	3	4
c	136	137	137	145				
c+1	144	146	145	162	6.0	6.5	6.3	12.1
c+2	153	156	155	192	6.0	7.0	6.6	18.0
c+3	162	168	166	238	6.0	7.8	7.2	24.0
c+4	172	184	180	309	6.0	9.4	8.2	30.1
c+5	182	206	198	421	6.0	12.2	10.1	36.3
c+6	193	242	224	600	6.0	17.1	13.4	42.5
c+7	205	302	267	894	6.0	25.0	19.1	48.9
c+8	217	377	308	1389	6.0	25.0	15.1	55.4
c+9	230	472	332	2249	6.0	25.0	7.8	61.9
c+10	244	590	352	3790	6.0	25.0	6.0	68.5

The detailed calculations are shown in the attached tables. Premiums have been rounded off to whole pounds.

The two most noticeable features of the table are the astronomic rise under assumption 4 and the extent to which future inflation can affect risk premiums for many years before it happens.

For example in assumption 3 where we have been going on a steady 6% basis until mid-year c + 6 the two years at 25% cast their shadow back to year c + 1; in no year is an increase of 25% required in risk premium. If we first become aware of the 25% at the start of year c + 7 (which is the most that can be hoped in real life) we might then give effect to the increased risk by raising premiums for mid c + 7 (i.e. premium year c + 8), having previously only increased premiums by 6% each year which up till then would have appeared adequate.

If we obtain an extra 25% premium for both c + 8 and c + 9 and drop the increase to 20% in c + 10 and then to 10% the risk premium and the charged premiums using year c as 100 are (based on the full table figures):

Year	c	c+1	c+2	c+3	c+4	c+5	c+6	c+7	c+8	c+9	c+10
Risk	100	106.3	113.3	121.5	131.5	144.8	164.2	195.6	225.1	242.7	257.3
Charged	100	106.0	112.4	119.1	126.2	133.8	141.9	150.4	188.0	234.9	281.9
Deficiency		0.3	0.9	2.4	5.3	11.0	22.3	42.2	37.1	7.8	-24.6
Accumulated deficiency		0.3	1.2	3.6	8.9	19.9	42.2	87.4	124.5	132.3	107.7

Under assumptions 2, 3 and 4 the chances of ever getting the right premium at least by the methods currently in use, seem very remote indeed.

This demonstration does show quite dramatically that the effects of inflation on large claims on the scales currently being experienced make errors in forecasting total costs arising from chance variations of a few large claims likely to be quite insignificant in relation to other errors.

ASSUMPTION 1

Year of Payment	Year of Notification								
	C	C+1	C+2	C+3	C+4	C+5	C+6	C+7	C+8
C	65536	-		-	-				
C+1	34734	69468	-	-	-				
C+2	18409	36818	73636	-	-				
C+3	9757	19514	39027	78054	-				
C+4	5171	10342	20684	41369	82737				
C+5	2741	5481	10963	21925	43851	87702	-	-	-
C+6	1453	2905	5810	11621	23241	46482	92964	-	-
C+7	770	1540	3079	6159	12318	24635	49271	98542	-
C+8	408	816	1632	3264	6528	13057	26114	52227	104454
C+9	216	433	865	1730	3460	6920	13840	27680	55361
C+10	229	229	458	917	1834	3667	7335	14671	29341
C+11		243	243	486	972	1944	3888	7775	15551
C+12			258	258	515	1030	2060	4121	8242
C+13				273	273	546	1092	2184	4368
C+14					289	289	579	1158	2315
C+15						307	307	614	1227
C+16							325	325	650
C+17								345	345
C+18									366
TOTALS £	139424	147789	156655	166056	176018	186579	197775	209642	222220
AVERAGES £	136	144	153	162	172	182	193	205	217

Year of Payment		Year of Notification								
		C	C+1	C+2	C+3	C+4	C+5	C+6	C+7	C+8
C		65536								
C+1		34734	69468							
C+2		18409	36818	73636						
C+3		9757	19514	39027	78054					
C+4		5171	10342	20684	41369	82837				
C+5		2741	5481	10963	21925	43851	87702			
C+6		1453	2905	5810	11621	23241	46482	92964		
C+7		908	1816	3631	7263	14526	29051	58103	116205	
C+8		567	1135	2270	4539	9079	18157	36314	72628	145256
C+9		354	709	1419	2837	5674	11348	22696	45393	90785
C+12		443	443	887	1773	3546	7093	14185	28370	56741
C+11			554	554	1108	2216	4433	8866	17731	35463
C+12				693	693	1385	2771	5541	11082	22164
C+13					866	866	1732	3463	6926	13853
C+14						1082	1082	2164	4329	8658
C+15							1353	1353	2706	5411
C+16								1691	1691	3382
C+17									2114	2114
C+18										2642
TOTALS	£	140073	149185	159574	172048	188203	211204	247340	309175	386469
AVERAGE	£	137	146	156	168	184	206	242	302	377

ASSUMPTION 3

Year of Payment	Year of Notification								
	C	C+1	C+2	C+3	C+4	C+5	C+6	C+7	C+8
C	65536								
C+1	34734	69468							
C+2	18409	36818	73636						
C+3	9757	19514	39027	78054					
C+4	5171	10342	20684	41369	82737				
C+5	2741	5481	10963	21925	43851	87702			
C+6	1453	2905	5810	11621	23241	46482	92964		
C+7	908	1816	3631	7263	14526	29051	58103	116205	
C+8	567	1135	2270	4539	9079	18157	36314	72628	145256
C+9	312	624	1248	2497	4993	9986	19973	39945	79891
C+10	331	331	662	1323	2646	5293	10586	21171	42342
C+11		351	351	701	1403	2805	5610	11221	22441
C+12			372	372	743	1487	2973	5947	11894
C+13				394	394	788	1576	3152	6304
C+14					418	418	835	1671	3341
C+15						443	443	885	1771
C+16							469	469	938
C+17								497	497
C+18									527
TOTALS £	139919	148785	158654	170058	184031	202612	229846	273791	315202
AVERAGE £	137	145	155	166	180	198	224	267	308

ASSUMPTION 4

Year of Payment	Year of Notification								
	C	C+1	C+2	C+3	C+4	C+5	C+6	C+7	C+8
C	65536								
C+1	34406	68813							
C+2	18924	37847	75695						
C+3	10881	21762	43524	87048					
C+4	6529	13057	26114	52229	104458				
C+5	4080	8161	16322	32643	65286	130572			
C+6	2652	5304	10609	21218	42436	84872	169744		
C+7	1790	3581	7161	14322	28644	57289	114577	229154	
C+8	1253	2506	5013	10026	20051	40102	80204	160408	320816
C+9	908	1817	3634	7268	14537	29075	58148	116296	232592
C+10	1363	1363	2726	5451	10903	21805	43611	87222	174444
C+11		2112	2112	4225	8450	16899	33798	67597	135194
C+12			3380	3380	6760	13519	27039	54077	108155
C+13				5577	5577	11153	22307	44614	89228
C+14					9481	9481	18961	37922	75844
C+15						16591	16591	33182	66364
C+16							29864	29864	59728
C+17								55248	55248
C+18									104970
TOTALS £	148322	166323	196290	243387	316583	431358	614844	915584	1422583
AVERAGE £	145	162	192	238	309	421	600	894	1389

Effect on claim Distributions

We now go on to examine the effect of inflation on percentages paid and on the results of grossing-up methods.

If we follow the proportions paid year by year under the different inflation assumptions, for successive years' cohorts, we get the following cumulative build-up (per 1000 total claim cost in each case)

Assumption	Year of Development	Year of notification (N)				
		c	c + 2	c + 4	c + 6	c + 8
No inflation	N	500)				
	N + 2	875)	ditto	ditto	ditto	ditto
	N + 5	984)				
Assumption 1						
	N	470)				
	N + 2	851)				
	N + 5	978)	ditto	ditto	ditto	ditto
Assumption 2						
	N	468	461	440	376)	
	N + 2	847	836	796	758)	ditto
	N + 5	973	964	951	943)	
Assumption 3						
	N	468	464	450	404	461
	N + 2	848	840	814	815	848
	N + 5	974	969	969	972	977
Assumption 4						
	N	442	386	330	276	225
	N + 2	801	741	670	592	510
	N + 5	946	914	869	813	744

These clearly show the deferment of total liability paid at any duration due to the inflation of later payments. Even so, looking at Cohort years c and c + 2 the differing assumptions of severe inflation in subsequent years do not have a marked effect on the different build-up patterns at this stage.

It is quite a different matter, however, when we look at the cohort years c + 6 onwards, under assumptions 2 - 4, when high inflation is affecting immediate payments. In each case the pattern has now changed radically from that pertaining to cohort year c, a much larger percentage of total liability being deferred due to the inflation content.

Only under conditions of an indefinite level rate of inflation do we get a stable pattern of claim build-up (Assumption 1, and Assumption 2 from year c + 8)

It is clear from the above therefore that, unless we get a stable rate of inflation over a long period of time, then we are not going to have a fixed pattern of claim build-up, all other things being equal. The pattern of build-up is particularly influenced by the rates of inflation in the earlier years of a cohort, and any projective work should therefore be principally concerned with getting these rates right.

Effect of G.U.F. Methods

It is also clear that under periods of changing inflation we are unlikely to predict the outcome of cohorts properly by simply relying on past cohort patterns without correction, since these will be unlikely to reproduce the future under different inflation conditions. This can be demonstrated very well by using the distribution from cohort year of notification c to project subsequent years' cohorts at different durations, under each inflation assumption. (The total liability forecast in each case is expressed per 1000 actual total liability for each cohort).

Assumption	Year of Development	Year of Notification (N) c + 2	c + 4	c + 6	c + 8
Assumption 1	N	1000)			
	N + 2	1000)	ditto	ditto	ditto
	N + 5	1000)			
Assumption 2	N	985	940	803)	
	N + 2	987	940	895)	ditto
	N + 5	991	977	969)	
Assumption 3	N	991	962	863	985
	N + 2	991	960	961	1000
	N + 5	995	995	998	1003
Assumption 4	N	873	747	624	509
	N + 2	925	836	739	637
	N + 5	966	919	859	786

Quite serious errors in the total forecast liability occur from the use of unadjusted past distributions in this way, when inflation is changing. The error in total liability narrows with increasing development of the cohort as one would expect, but conversely, although the outstanding amounts decrease, the percentage error in the outstanding liability increases, due to the gearing effect at later durations.

In the case of Assumption 4 the method breaks down completely, although in the severe inflation assumed for this model it can be argued that conventional insurance would cease to be viable in any event!

Corrections for Inflation

We are therefore led to the conclusion that if we are to use GUF methods of projection based on past cohort distributions, then we must make an attempt to correct for past inflation in the base data and also to make suitable allowance for future inflation, at least for those years of development immediately in the future. Otherwise, we are accepting projected values which are inconsistent with known or assumed inflation conditions.

The following section describes how this may be attempted in practice, with application to the projection of a partly developed claims cohort.

The correction for inflation falls readily into two areas:

(i) Correction of base data for past inflation
(ii) Allowance for inflation appropriate to the particular claims cohort in the resulting standard distribution.

Correction (i) involves deflating the base payment values to a common datum year, by reference to a suitable inflation index for each year of payment. In practice several years' data will be used for this purpose. This produces base data in 'real value' terms, from which a standard distribution table free from past inflation is produced.
On the assumption that the real distribution pattern is stable over time, we can then use this distribution table for projective claims cohorts, by reflating the distribution using inflation rates appropriate to the particular claim cohort (some of these rates will reflect past rates to date of projection, the rest will be projected rates).
We can see how this would work in practice by considering a claims cohort for year c, developed to year c + n.

Year	Actual Payments			
		Distribution excl. inflation	Inflation Index	Distribution incl. inflation
	(1)	(2)	(3)	(4) = (2) x (3)
c	S_c	A_0	r_c	$A_0 r_c$
c + 1	S_1	A_1	r_{c-1}	$A_1 r_{c+1}$
c + 2	S_2	A_2	r_{c+2}	$A_2 r_{c+2}$
⋮	⋮	⋮	⋮	⋮
c + n	S_n	A_n	r_{c+n}	$A_n r_{c+n}$
⋮		⋮		⋮
∞		A_∞	r_∞	$A_\infty r_\infty$

The actual payments made to date will themselves have been subject to past inflation. If we assume that these have followed the standard distribution to date including the allowance for inflation, then we have

$\sum_{i=0}^{n} S_i$ = actual payments to year c + n

$\sum_{i=0}^{n} A_i r_{c-i}$ = standard payments to year c + n, including past inflation

$\sum_{i=0}^{\infty} A_i r_{c+i}$ = standard payments (total), including past inflation and assumed future inflation

and the projected total liability is given by

$$L = \sum_{0}^{n} S_i \times \frac{\sum_0^\infty A_i r_{c+i}}{\sum_0^n A_i r_{c+i}}$$
$$= \sum_{0}^{n} S_i \times f_{cn} \text{ (say)}.$$

from which by deduction the outstanding payments are given by :

$$L - \sum_{0}^{n} S_i$$

A series of GUF ($f_{c,n}$) would be calculated for each partly developed cohort to be projected, incorporating appropriate inflation values for each cohort.

The above method makes the implicit assumption that there is a reasonably stable distribution of payments for each cohort over time. This could be disturbed by such things as changes in office systems, strikes, work back-logs, etc., or by changes in portfolio, type of claim etc., and would have to be taken into account when deciding whether the distribution was appropriate for projective purposes. This emphasizes the difficulty of applying statutory supervision techniques which do not have this knowledge of internal office changes.

Cost/Claims Methods

From the same considerations as above it is evident that cost per claim techniques based on projection of a base cost per claim value to a future mean claim point, will be an inadequate estimator of that claim cost in periods of changing inflation, unless proper allowance is made for inflation both in the base value and in the projection period.

In particular, it is insufficient to project a base cost-per-claim value forward to a future mean claim point merely by inflating the base value by the rate of inflation over that time interval, since we have still carried forward the original base assumptions for run-off inflation into the new calculation. These of course may be quite inappropriate in the projected run-off period.

(e.g. a cost/claim value determined at the end of 1974 on a then assumed 15% run-off inflation rate becomes inadequate when projected to mid 1975 if a 25% inflation rate is then anticipated!)

The Development of related average claim amount over time

If we take any cohort of claims and consider the averages in various groups as developed from time to time, we can observe some interesting trends. During the first two to three years we can consider only claims settled to date since the estimating process is unlikely to be sufficiently reliable before then. This has been done for several cohorts of private cars insured under Comprehensive policies and for periods up to five years. Results for separation into five groups by area in which the vehicle is garaged (Rating area), are shown in the table below. The figures are the ratio of the rating area average to the average for the whole comprehensive portfolio multiplied by 1000. This grouping into areas has been chosen because it is now fairly clear from the results of several such analyses that there are no significant differences between areas: if there are any actual differences they are probably of the order of five in one thousand or less. This means that any differences between area averages in any one year consist almost entirely of random variations.

A typical set of results is given in the Table below :-

Relative average amount of claims

Basis: claims notified in 1969
settlements with n months from 1.1.69.
ULT = position at 31.12.74 for all claims including estimates
Private Car Comprehensive

Rating Area	Months of Development (n)												
	3	6	9	12	15	18	21	24	30	36	48	60	ULT
A	1409	1112	1060	1116	1087	1069	1050	1035	1003	1022	1020	1013	949
B	997	1034	997	984	992	1000	999	998	992	1014	1030	1021	1011
C	859	938	968	982	1000	998	1006	1006	987	997	998	1004	1007
D	995	983	997	990	980	982	985	990	976	984	974	968	979
E-G	1173	1071	1065	1036	1015	1012	997	996	986	1007	1000	1012	1028
range	550	174	97	134	107	87	65	45	27	38	56	53	79

It will be seen that the random variations decrease fairly steadily for about 2 - 3 years and then begin to increase again. The decrease is a natural consequence of the steadily increasing numbers on which the figures are based as time passes, whilst the increases are the result of a few large claims whose settlementsbegin to appear in 2 - 3 years' time. Obviously, these later settlements give us, in one sense, more information than we previously had, but it is not so obvious that this information is of value for predictive purposes. Clearly we would be wrong to conclude that because in 1969 a large claim had occurred in say, area A but not in any other area, it was more likely to happen that way also in 1970.

It might seem reasonable to divide the investigation into two parts, the first being concerned with relative average amounts of claims based on the smaller, or earlier, or truncated settlements. In view of the fact that rates of paying claims and practice in regard to declaring then "closed" does vary both within and between insurers, a truncating (top slicing) method seems more objective, and if we top slice at £1500 the later part of the above table becomes

Relative average amount of claims settled

Basis claims notified in 1969
settlements within n months from 1.1.69.
ULT = position at 31.12.74 for all claims including estimates
Amounts are the actual amount of settlement or £1500 whichever is the lower

Private Car Comprehensive

Rating Area	MONTHS OF DEVELOPMENT (~)			
	48	60	72	ULT
A	1036	1040	1035	1032
B	1015	1016	1017	1023
C	998	1000	997	997
D	974	971	971	968
EFG	1015	1013	1021	1019
range	62	69	64	65

It will be seen that the range is now steady, in fact settlements after about 2 - 3 years are so few in number that they have very little effect on these figures.

Incidence and Trends in Amounts

The second aspect of the investigation is concerned with the incidence of the larger claims and on trends in amounts. The grossing up factors can be calculated for various sub-groups, such as Rating Area, and for various cut off points, and for a series of claim cohorts - preferably each for a calendar year, since there is a noticable seasonal pattern which is possibly related to factors such as age of policyholder, age of car and Rating Area, which appear to be related to the kind of exposure giving rise to the risk of large claims.

LARGE CLAIMS

Large claims are an unmitigated nuisance. Not merely do they make serious inroads into the fund (or the fund of an excess of loss reinsurer), they also make a statistical analysis difficult and uncertain. Even when they don't happen, they are still a potential nuisance since the prudent insurer will assume that they will appear sooner or later, maybe in the development of claims that arose from events many years ago; his difficulty is in knowing how much to provide for them, or, if they are reinsured, whether the premium he is asked to pay is reasonable or not.

Let us first set out the nature of the problem and consider some of the difficulties. It will help, to begin with, to examine typical claims amount distributions arising from U.K. Private Motor Insurance.

The figures below are taken from a group of just over 80,000 claims notified during 1971 on a portfolio of about 700,000 private cars, and are based on payments to the end of 1973 plus estimates for those still open at that time. When the claims are finally settled it is likely that the distribution will be a little flatter, but more widespread, as claims are settled without payment, or, on the other hand, for more than was expected at the end of 1973. This, however, will merely emphasise still more the observations to follow.

Claim amount bands	Assumed average amount in band	Number of claims per 10,000 claims in the type of cover		
		T.P. only	T.P.F.T.	Comprehensive
£	£			
0	0	6000	6017	1466
0-5	3	457	444	241
5-10	8	187	193	588
10-15	12.5	365	331	866
15-20	17.5	310	281	534
20-30	25	644	566	736
30-40	35	310	331	587
40-50	44	238	252	534
50-70	57	395	416	941
70-100	80	355	357	993
100-150	117	253	292	962
150-250	182	212	262	839
250-500	333	169	152	538
500-1250	750	79	76	150
1250-3000	1900	24	23	22
over 3000	5000	2	7	3
Total		10000	10000	10000

The average amount in each band is based on the actual averages observed over a number of years. In the case of the last three bands, however, the small number of claims is liable to make the actual average differ from these assumptions, particularly in the last group, as will be seen later. The effects of such differences will be demonstrated.

The distributions for TP and TPFT are very similar and the two columns have been kept separate to show how sensitive the results are to random fluctuations in the numbers in the last group. Differences will arise from the fire and theft claims but these are quite small in both number and amount and also possibly from the tendency of the TP only class to contain a larger than normal proportion of young policyholders. The comprehensive distribution is quite different from the TP and TPFT as it is dominated by damage claims in the £50-£500 range, whereas in this range the non-comprehensive claims are scarce. Many of the smaller non-comprehensive claims will be for payment of a third party's excess on his own policy or of an emergency treatment fee.

We shall now calculate the mean and the variance of these samples taking all the cases within one particular amount band as being equal to the average in that band but making four different assumptions. Two of these take the distribution as shown, one including zero claims, the other excluding them. The remaining two are the same as before except that one of the £5000 claims has been replaced by one for £50,000. It must be noted that from a much wider claims experience the expectation of a claim of £50,000 or more in every 10,000 claims is not entirely unreasonable, the expectation probably being of the order of 0.2 to 0.3 so that the assumption we are making is quite realistic and the modified sample could well have arisen in practice.

THIRD PARTY ONLY CLAIMS

Claims	Non zero claims only		All Claims	
£50,000 claim included?	No	Yes	No	Yes
Mean	85.1	96.4	34.0	38.6
S.D.	216	814	143	517
S.D. of Mean	3.4	12.9	1.4	5.2
3 x S.D. of Mean ÷ Mean	12	40	12	40

TPFT CLAIMS

Claims	Non zero claims only		All Claims	
£50,000 claim included?	No	Yes	No	Yes
Mean	92.7	104.0	36.9	41.4
S.D.	275	834	179	529
S.D. of Mean	4.4	13.2	1.8	5.3
3 x S.D. of Mean ÷ Mean	14	38	15	38

COMPREHENSIVE CLAIMS

Claims	Non zero claims only		All Claims	
£50,000 claim included?	No	Yes	No	Yes
Mean	97.8	103.1	83.5	88.0
S.D.	178	566	168	524
S.D. of Mean	1.9	6.1	1.7	5.2
3 x S.D. of Mean ÷ Mean	6	18	6	18

Note: The units in this table are in currency which is probably about £2 in 1975 except for the last line of each block where the figures are percentages.

The first thing to notice is that if we consider only non-zero claims, there is just a small difference between the non-comprehensive mean claims and that the T.P.F.T. is nearer to the comprehensive than the TP only, the relationship being even closer when the £50,000 claim is present than when it is absent. The standard deviations of the non-comprehensive are larger than those of the comprehensive, the difference being very pronounced when the £50,000 claim is present. Finally it will be seen that the difference between TP and TPFT are small except for the numbers in the largest band and this is almost certainly a random fluctuation.

If however we consider all the claims the mean comprehensive is very much larger than either the non-comprehensive but variances are much the same for all three. This is because far more non-comprehensive claims are zero claims. It is by no means obvious whether we should use all the claims or merely non-zero claims: there is some merit in using the all claim figure since it may take a very long time to establish exactly how many non-zero claims we shall finally end up with and experience has shown that the proportion of claims which are zero at settlement does not dwindle very quickly as time passes.

We regard a study of these figures as being absolutely vital to any serious discussion of techniques for detecting the differences between groups and among the points to be noted are:

1) The figures correspond to 10,000 claims, which is about 100,000 vehicle years exposure in the non-comprehensive classes and 60-80,000 vehicle years in the comprehensive class (for some high risk groups it requires rather fewer policy years to produce 10,000 claims but this is not normally important.) The largest U.K. portfolio includes about 1.5 million private vehicles, there are about five portfolios in the range ½-1 million and the rest are all smaller than ½ million. These figures are for all covers combined: for the largest portfolios it is likely that at least two-thirds of the policies will be comprehensive and under one third non-comprehensive. It will be seen therefore that very few investigations other than for pooled experience or covering several years will give groups over 10,000 claims if there are more than 3 or 4 different groups even for comprehensive cover, whilst groups of this size will be virtually unattainable in non-comprehensive.

   In other words observed means are going to be more variable than those in the table and differences considerably less likely to be found to be statistically significant.

2) The estimate of the mean claim amount derived from the sample is increased by between 5% and 13% with the appearance of a single £50,000 claim whilst the estimate of the standard deviation of the population is increased by 200-250% in that event. It follows that estimation of the S.D. of the population from a sample of claims is probably almost meaningless One solution would be to adopt coefficients of variation (that is the ratio of the S.D. to the mean) as follows:-

	COVER	
	Comp	Non-comp
All Claims	5	10
Non-zero claims	4	7

   No calculations have been made in regard to third or fourth moments. If any reader thinks they have any meaning oh relevance in an enquiry in these circumstances he would be well advised to find some non-statistical occupation!

3) If we use the figures in the last paragraph we obtain a table of mean (M) and 3 S.D. of mean (=D) as follows for 10,000 claims:-

	COVER			
	Comp		Non-comp	
	M	D	M	D
All claims	85	12	37	12
Non-zero claims	100	12	90	12

   In other words for a difference between means of two groups of 10,000 claims each to be significant at about the 5% level we require that it be at least

£12 which is from 15% to 35% of the mean. It should be remembered that these figures are in terms of 1971 £'s and should be substantially increased for 1975 . Another way of putting this is to say that the unreliability of the amount of any such difference is quite marked; an apparent difference of 15% could correspond to a true difference in the range 10-20% at the best and 5-25% or sometimes even worse.

It should now be clear that conventional methods of dealing with large claims are unlikely to be helpful, firstly because of the very long time it takes to establish the amount of them and secondly because even when they are established, the resultant means and variances may be poor predictors and too variable to permit reliable estimates of differences between group means.

Various methods of overcoming the difficulties have been tried. The simplest is in effect to ignore all amounts over a specified limit and to spread the excess rateably over all claims. Symbolically

$$\sum_{n=1}^{N} C_n'' = \sum_{n=1}^{N} C_n' \quad \frac{\sum_{n=1}^{N} C_n}{\sum_{n=1}^{N} C_n'}$$

where $\begin{cases} C_n \text{ is the claims amount} \\ C_n' = C_n \text{ if } C \leqslant L \\ C_n' = L \text{ otherwise} \end{cases}$

Claims from A to B form one group; there are N claims in all

The factor $\sum C_n \div \sum C_n'$ is the grossing up factor (G.U.F.); if we are concerned only with <u>relative</u> claim amounts it can be taken as unity.

The tacit assumption is that the expected proportionate contribution of large claims does not vary from group to group. This assumption may be good enough for some groups, but research has shown that it is probably not universally true; for example young policyholders seem to make a higher proportion of large claims so that they probably should have higher grossing up factors.

A paper setting out this research in detail is in preparation, but some idea of the size of grossing up factors relating to all claim payments is:-

	<u>L= £250</u>		<u>L = £1000</u>	
	<u>Comp</u>	<u>Non-Comp</u>	<u>Comp</u>	<u>Non-Comp</u>
Young policyholders	1.8	*	1.3	*
Older policyholders	1.4	*	1.15	*
All	1.5	2.5	1.2	1.8

* not enough data are available yet to give reliable differences

Discussion has taken place in regard to a suitable value for L. Something in the region of £500 to £2,000 seems to be indicated, and if the best value seems to be at all critical, it would point to the need for further examination.

It must be realised that problems of estimating large claims prior to settlement mean that the grossing up factor will not be reliably known for several years so that this is little help in establishing absolute levels of average claim at least until several years data are available, most of them fairly fully developed. It should be noted that after about two years, contributions of C' to the final $\sum C'$ are negligible (as they are very few in number and cannot exceed L each, which will be relatively small). After this time changes show up in the GUF' not in the curtate total $\sum C'$.

A similar attack on the problem is by weighting, whereby claims over some limit (probably less than in the straightforward "top slicing" method just described) are weighted with weights that decrease in some defined way. In such cases C' increases monotonically, but there are merits in not having a discontinuity as in top slicing especially in times of rapid inflation. Various weighting methods are described in Hey (2). It must be realised however that we have to choose between giving a claim so much weight that a very big one distorts the result, and so little that it has too little effect and too much information is lost. Gains over a top slicing method of the limit of say, £1,500 may be quite small but further investigation is needed.

Some typical figures are given in the following table:

Grossing up factors to apply to curtate totals of claim amounts

Basis: Private Motor Comprehensive
Claims notified in 1967 to 1969 developed to 31.12.72
Claims notified in 1969 to 1972 developed to 31.12.74
Amounts paid plus those estimated to be outstanding at the date of development.
All payments net of recoveries other than for excess of loss reinsurance

	Year of notification							
	1967	1968	1969 to 31.12.72	1969 to 31.12.74	1970	1971	1972	Average
Limit £500								
Age of P.H.								
Under 21	1.39	1.35	1.37	1.59	1.42	1.49	1.47	1.44
21-25	1.24	1.27	1.30	1.31	1.34	1.42	1.64	1.36
26-45	1.24	1.19	1.26	1.27	1.20	1.32	1.26	1.25
46-65	1.23	1.22	1.23	1.25	1.27	1.26	1.31	1.25
Over 65	1.06	1.18	1.35	1.21	1.27	1.22	1.35	1.23

	Year of Notification							
	1967	1968	1969 to 31.12 72	1969 to 31.12 74	1970	1971	1972	Average
Limit £1500								
Age of P.H.								
Under 21	1.18	1.18	1.19	1.35	1.23	1.27	1.24	1.23
21-25	1.10	1.15	1.15	1.16	1.19	1.26	1.41	1.20
26-45	1.14	1.11	1.15	1.16	1.10	1.18	1.13	1.14
46-65	1.12	1.11	1.12	1.13	1.15	1.13	1.17	1.13
Over 65	1.02	1.09	1.24	1.11	1.13	1.09	1.19	1.12

Basis : As above except that amounts are based on payments and estimated outstandings for third party liability only and exclude payments and recoveries under own damage, fire, theft and miscellaneous covers and also under claim sharing agreements and excess of loss reinsurance, but after any other recoveries

	1969	1970	1971	1972	Average
LIMIT £500					
Age of P.H.					
Under 21	2.65	2.32	2.41	2.55	2.48
21-25	2.10	2.30	2.69	3.52	2.65
26-45	2.14	1.84	2.43	2.07	2.12
46-65	1.94	2.07	2.08	2.29	2.09
over 65	1.76	1.93	1.75	2.35	1.95

	1969	1970	1971	1972	Average
LIMIT £1500					
Age of P.H.					
Under 21	1.80	1.62	1.68	1.65	1.69
21-25	1.55	1.67	1.91	2.30	1.61
26-45	1.61	1.39	1.78	1.51	1.57
46-65	1.47	1.54	1.52	1.65	1.54
over 65	1.43	1.44	1.34	1.70	1.48

The total number of claims for each year of notification varies between 60,000 and 70,000 (500,000 to 600,000 vehicle years of exposure); it will therefore be evident that it will require very large exposures to give reliable results. At a limit of £1500 the number of claims in various years have been approximately as follows:-

BASED ON TOTAL PAYMENTS

Age of P.H.	1967	1968	1969	1970	1971	1972	TOTAL
Under 21	25	25	20	11	10	10	101
21-25	33	28	42	36	42	66	247
26-45	82	67	90	89	116	134	578
46-60*	84	85	84	100	92	109	554
Over 60*	2	6	24	45	50	67	194

* 65 in 1967 and 1968

BASED ON T.P. PAYMENTS AS ABOVE

Age of P.H.	1969	1970	1971	1972	TOTAL
Under 21	18	10	11	10	49
21-25	36	33	34	58	161
26-45	77	70	89	115	341
46-60	70	85	82	88	325

Investigations so far suggest that age of policyholder may be the only factor to give rise to differences of any importance between the grossing up factors for various groups, although this could be the result of a lack of data.

The variation from year to year in the grossing up factors suggests that any departure from a uniform spreading of large claims over the whole portfolio is likely to be difficult to justify: we must not however overlook the effect of Knock-for-Knock agreements which operate to transfer crude claims costs from the higher risk policyholders to those with lower risk as well as from non-comprehensive to comprehensive cover, but the transfers are concentrated on the smaller payments, namely from own damage. It seems that further investigation is justified, and that until results are available insurers are unlikely to suffer by adopting a policy of uniform spreading; if any one does something different he has a good chance of being wrong.

Finally we come to the problem of associated factors. Tabulations have been produced for each of the 23 breakdowns for which the claim frequency analysis is available, and for many breakdowns in other classes of business. These give, for years of notification 1969 onwards for each line of each breakdown:

The number of claims

The total amount based on payments to date plus estimates (note:where only a sample of claims has been individually estimated, the actual estimates have been multipled by the sampling ratio, namely, 2 or 10)

The total parameters based on the standard table (excluding the parameters for the breakdown in question, when it corresponds to a parameter).

Total curtate amount, namely actual amount or £1500 whichever is the smaller

From these we obtain:-

Average actual amount

Average curtate amount multiplied by the grossing up factor for the entire portfolio (1.25 was used for comprehensive cover and 2.0 for non-comprehensive cover for the latest tabulation although in fact these figures ought to have been slowly increasing over the years).

Relative average adjusted amount = $\frac{1{,}000 \times \text{Group average-group parameter}}{\text{(overall average - overall parameter)}}$

Relative curtate average adjusted amount.

The tabulations afford evidence for the validity of adjustment of the standard table. For latest years of notification the group averages of actual amounts are liable to fluctuate arising from the estimating of a sample only of the claims and the curtate amounts are affected by the sampling: for example one claim of £15,000 will replace ten of £1500 each and if nothing has been paid on any of these claims the one estimated claim will be treated as £1500 whereas it should be 10 at £1500 thus giving rise to an underestimate. However with curtate claims we are mostly concerned with the relative amounts so that this is not very important as the effect will be the same (more or less) in every group.

We might observe here that during the first two years of development there are very few claims over £1500 so that curtate and actual averages differ very little; after two years there are so few settlements that even if most of them are large they have surprisingly little effect on the curtate averages.

There is probably very little that curtate averages can tell us if we cut off in the £1,000 to £2,000 region that we don't already know from 2 years development of actual figures. The latter have one advantage in that they make us aware of how little we do know, whereas with the curtate averages almost everything is hidden by the final grossing up factor, of which we are likely to be ignorant for a long time.

ESTIMATING AND RESERVING

We now come to an area which is only partly the concern of the statistician and where exisiting practice leaves a great deal to be desired. There are a number of problems, mostly connected with each other and mostly dependent on the proper estimating of liabilities for outstanding claims and for claims to be incurred in respect of cover which the insurer is committed to provide either in respect of policies in force or those which will be written before future discussions on premium scales can become effective.

PROFITABILITY

An important figure which is a natural consequence of claims estimates is the measure of profitability. Conventionally "profit" is measured from a Company's annual revenue account, and the extent to which the return will reflect the underlying experience for that year alone will depend upon the extent to which included reserves are correctly valued and also the extent to which adjustments relating to earlier years are absent. It must be realised that the claims incurred in the year will take several years to settle, and in the meantime the outstanding claims included in that years accounts are still very susceptible to continuing inflation beyond the company's control as well as other uncertainties which arise in the settlement of liability claims. The true profit emerging from the years experience cannot therefore be calculated exactly until the last payment or recovery has been made from the resultant claims. In the meantime any prior release of surplus should be regarded as only a component of the final profit, and it will be dependent on the accuracy of the estimates on the outstanding claims liability. Certainly at the end of the year of account with about half the claims liability still outstanding, any profit determination at that stage must be very provisional.

Another feature which further complicates the calculation of profit on the years experience is the inclusion in the account of changes in liability stemming from earlier years outstanding claims although this may well tend to disappear if the recommendations of the Sandilands committee are adopted.

Where the estimate brought forward for earlier years outstanding claims differs from payments made on these claims during the year plus the estimates carried forward at the end of the current account for those claims still outstanding it will give rise to either a release of liability or an imposed additional liability. Strictly these should be allocated back to the original years of account and certainly should be so allocated if the true profitability for each year of account is ultimately to be measured.

For formal presentation, however past years claims movements are traditionally included with the current years claims liability. If these movements are substantial their inclusion can lead to a considerable distortion of the apparent profit arising from the current years experience only. It is therefore highly desirable to distinguish in the years account between current year liabilities and savings or losses on estimates included on earlier claims when measuring profitability of the current account year. Even then it is necessary to ascertain whether the current years estimates are on a strict sufficiency basis or whether, as is often believed to be the case, there is a deliberate built-in element of future saving in the stated estimates.

The apparent rate of profit emerging from any year's account may be further affected by the treatment of reserves made in respect of claims liability and unexpired risks.

Claims liability relates to all events up to the balance sheet date. We have already dealt with the estimates required for notified claims, and additionally we require to make an IBNR claims reserve for claims incurred up to that date. Although the principle is clear enough, there is scope for variation in the value placed on the IBNR reserve, depending on the Company's assessment of notification delays, internal work flow situations, claim frequency trends, and cost/claim characteristics for IBNR, (which may be different to the portfolio as a whole). Additionally the Company may maintain some level of claims equalization reserve, to smooth out the impact of singular large claims in any year.

Premium reserves can be regarded as the unearned premiums carried forward at the end of the accounts period, plus a further unexpired risk reserve where the unearned premiums are inadequate to meet the liability likely to arise from them. The method of calculation of UPR is standard enough, most companies now using a '24ths' method, although some variation exists in allowance for initial expenses in this calculation, and premium back-logs in the system can also have an effect on the value. With the experience of current (1975)inflation in mind, however, is there not a case for an 'index-linked' 24ths method (ie. the spread of earned premiums would be weighted for inflation, so that later months would be allocated a higher proportion of the premium than under the standard 24ths method)? Much greater variation in thought exists over the question of an additional reserve for unexpired risk. This falls between those who would allow for the deficiency in the year in which the premium is written (ie. the current accounts year), and those who would allow the deficiency to appear in the year of exposure (ie. the following accounts year in this case). There is further scope for variation on the question of whether the additional reserve should cover only the excess claims risk or whether one should go further and bring associated expenses into the calculation.

We can therefore see that the concept of accounting 'profit' is subject to considerable variation, and illustrates the fallacy of attempting to use the accounts as a measure of profitability when considering rating adequacy. The revenue account, even having excluded the distortion of earlier years' claims, is a measure of the experience exposed in the current accounts period. As such it contains an amalgam of premiums exposed during the period which, in a 12 month accounts period, will have starting dates spanning a period of 24 months, and probably covering more than one generation of rating series. In this situation it is quite possible for the accounts to show an aggregate profit in the accounting period, whilst disguising the fact that premiums written towards the end of that period were incurring substantial losses, requiring immediate remedial action.

This is an important point, as it illustrates the quite different purpose served by the accounts in measuring an amalgam of profit in a given period, whereas what we should be trying to ascertain for rating purposes is whether we are currently and (perhaps more important) prospectively writing risks at a rate of premium adequate to cover all outgoings likely to arise from those risks. The trouble is that, because of the carry-forward of risk premium at the end of an account, the accounting profit will always lag in detecting a trend on current experience, and relying on the accounts to re- rate we shall do it too late, and by the wrong amount.

This therefore suggests that, for rating purposes, we should adopt an approach whereby we measure the run-off of cohorts of premiums written in the same rating series, both for measuring current profitability, and also for projecting over the expected duration of a rating series, since we should aim to make each rating series pay for itself during its lifetime. If we achieve this as a primary objective, then the accounting profit will ultimately tend to the same levels.

In practice, the measurement and application of cohort profitability in this way has difficulties, since we require to allocate all items of outgo back to the policy-start-date cohorts, and to follow the complete run-off of items attaching to these cohorts. If we are to act quickly enough on the emerging information, then we are inevitably bound to include a considerable element of projection and estimation for the unexpired portions of risk.

Where the company has adequate computer systems, however, it is possible to measure profitability in a current period by re-calculating the entire portfolio at the current premium rates, applying that against current claims experience. By an extension of this technique, rating profitability can be projected to given future periods, both on run-off and accounting bases. These methods can provide important 'early warning' information for management and also allow alternate strategies to be tested.

STATUTORY STATISTICS

This section must to some extent overlap the work of the group dealing with claims estimating and solvency but it is an integral part of the "statistics" part of our remet. Insurers are required by law to submit accounts and statistical returns of two general kinds.

First is the normal kind of accounting information, comprising revenue accounts leading up to a profit and loss account and balance sheet which are to be submitted (in the case of companies) in accordance with the Companies Acts, to shareholders and to the Department of Trade, Companies Division. Non-company insurers submit similar accounts to the appropriate body (the Committee of Lloyds or the Registrar of Friendly Societies).

Secondly are the more detailed returns (often referred to as the statistical returns) submitted under the Insurance Companies Act to the Department of Trade Insurance Division, which Division also lodges copies of all these returns with the Companies Division. The latter requirement can be dispensed with at the Insurance Section's discretion.

At the present time (1975) both forms of return are to be made annually within a limited period, but with power to the Department to allow a longer period. Companies listed on the Stock Exchange must also supply quarterly returns and under the Insurance Companies Act, quarterly returns will in due course be required for the Department of Trade.

We must consider in relation to these returns:-

What are they?
What are they for?
How are they compiled?
Do they achieve their objects?
Will changes now being considered go any further to achieving their object?
How could the system be
a) simplified?
b) made more effective?

WHAT ARE THEY?

WHAT ARE THEY FOR?

The accounting returns comprise the accounts of the various insurance funds, a profit and loss account and a balance sheet. They are primarily intended to give shareholders an idea of the profits of the Company and secondly to allow them to estimate the capital position, their own security and the possible need to raise further capital. They are not, primarily, statistical.

The statistical returns are intended primarily to establish solvency or trends in the surplus of free assets and to detect unfavourable trends in an insurers experience in ample time to take remedial action.

The form of accounting returns will be familiar to all actuaries and they are similar in kind to those for a life office; the problem of measuring *profits* depends on estimates of liabilities for outstanding claims and unexpired periods of cover and this has been explored elsewhere in this paper. The form of the statistical return is decribed briefy below.

STATUTORY STATISTICS

Subdivision of Business

The regulations call for returns in respect of each of the statutory classes of business namely liability, marine aviation and transport (MAT) , motor, personal accident, pecuniary loss and property.

Returns are required not only in respect of each class of business but also in respect of each country in which risks are undertaken. Companies are further required to subdivide the classes in each country according to the type of risk within the class (for example private cars and motor fleets would usually be two separate "risk groups"). The number of risk groups is currently between one and five for any class and country but this is likely to increase under the regulations now being discussed. There are provisions to exclude small returns, but whilst they apply to a small class or a small country, they do not extend to small risk groups within a class.

The rest of this note relates solely to motor business.

CLAIM FREQUENCY STATISTICS

A return is prepared in respect of each group; this should show:

the gross premium and the number of vehicle years of cover granted during the year,
- that were brought forward during the previous year*
- that were commenced during the present year
- that were carried forward to the next year

the number of claims in that year that

- were in respect of incidents in earlier years
- were old claims reopened
- were in respect of incidents in the current year
- the estimated number that occurred in the current year but which will not be reported until next year

claim frequency for the year

An analysis of a claim frequency return is shown in Appendix 3.

* strictly those that originated in the previous year: this will differ from the brought forward since the cover actually provided in a year will differ from that brought forward at the year end whenever there is an endorsement or mid year cancellation. In the new forms of return this situation should be clarified.

CLAIM SETTLEMENT ANALYSIS

A return is prepared in respect of each risk group for years of claim from 1970 onwards giving a historical record of the claims settled, and outstanding.

These schedules show for the year of claim and each year up to and including the year of account the following figures:-

-number of claims closed without cost
-number of claims closed with payment
-number of claims outstanding at the end of the year
-amount paid in the year
-cumulative amount paid in respect of the years to date
-amount reserved at the end of the year
-total amount and reserve at the end of the year.

An example of a claim settlement analysis is given in Appendix 4.

APPLICATION OF STATUTORY STATISTICS

Claim frequency statistics recorded over a period of years were intended to indicate trends in respect of the company. These trends may be the direct result of

- national trends, e.g. effect of legislation, changing traffic densities, safety measures etc.

- change in underwriting policy of the company which might for example alter the mix of third party/comprehensive business

Without background information on the Company's underwriting policy many of the figures are liable to be misleading.

The figures of total expected liability were intended not merely to show whether a company was previously reserving on an adequate basis but also whether its current estimates should be amended. Unfortunately it would be quite unsafe to assume that the practice of a company is unchanged over a period and it is a pity that realisation of this fact seems to have come so slowly.

HOW ARE THE FIGURES COMPILED?

A great deal of uncertainty surrounds methods used to prepare returns and it is known that some companies have had to be allowed considerable latitude especially in the early years of the new returns. The members of the working party, however, are very well aware of problems that have arisen that make compliance with the spirit of the regulations extremely difficult and for reasons that should now be apparent, they cannot be "right" if they are also to reconcile with accounts prepared on traditional lines. The problem of definitions of reliability and of reconciliation are very severe.

DO THEY ATTAIN THE OBJECTS ABOVE?

In a word NO. In fact it would not be unfair to say that the interpretation of legal requirements as to accounting has traditionally been such as to lead one to suspect that its main aim is to conceal rather than reveal. In a paper read to the Institute of Actuaries of Australia and New Zealand in July 1975 Roger Sawkins has examined the extent to which accounts on conventional lines may completely distort the picture and his paper is worthy of scrutiny by anyone interested in the subject. The principal objection which was noted by the Sandilands committee is that in the revenue accounts and in the profits and loss account receipts and payments for the current year which frequently include substantial amounts based on estimates of claim of which little or nothing is known, are mixed with transactions and adjustments relating to the previous years, and, possibly, the provision for future years not reflected in premiums charged, so as to produce a hybrid figure which may or more likely may not be a reliable indicator of the year of account. Many of the current proposals unfortunately do not seem likely to affect this position very much.

Quarterly figures are an example of the triumph of hope over experience. They can be little more than guesswork and in many types of insurance with seasonal patterns (hail, wind storms) could well be meaningless. Far from their being required from the Stock Exchange or even from the Department of Trade a good case could be made for banning them if they seem to affect share prices or the assessment of companies without good cause.

The trouble largely stems from the legal requirements to establish within three months of the end of a period the results of an insurer whose activities may be world wide. The short answer is that the problem is insoluble. Like a life fund, true profit or surplus emerges only over a period and the account of a period should be regarded as a single entity not to be combined with those for any other period. Outside the company world this is recognised by the Lloyds system whereby each years accounts (which relate to premium written on dates within the year and the relative claims arising from those premiums) are kept open as a separate entity two years after the end of the year at which time the outstanding liabilities are reinsured and a balance struck.

Whilst this is better than annual accounts it still sweeps a good deal under the carpet after two to three years, although, as has been explained elsewhere, what happens after these accounts are closed may be of little value in predicting future experience, where for example it is related largely to a few big claims.

It is not in one sense a matter for serious concern by the non-life actuary if companies insist on preparing or are required to prepare meaningless accounts. When we come to the statistical returns, however, the position is entirely different and we must first explore in greater depth what they are for.

THE PURPOSE OF STATISTICAL RETURNS

The primary concern of these returns is solvency rather than profitability and the projection of experience to enable future premiumscales to be judged adequate. Within these concerns they are intended to test the adequacy of reserves for outstanding liabilities but without reference to the assets held to cover them the idea is that if $A - L > kP$, where A is the "value" of the assets, L the "amount" of the liabilities (both outstanding claims as well as unearned premium reserve and unearned risk reserve), k is some constant set by law and P the retained premium for the year of account then the insurer is solvent and can continue to trade; otherwise he must be wound up.

In the United Kingdom k depends to some extent on P, but will normally be a little over 0.1 with the likelihood of a rise to 0.18 or thereabouts in the near future to conform with EEC rules. The relation between A, L and P depends very much on the class of business. For a liability class L may exceed 3.5P for motor it will be in the order of 1.2P unless an abnormal amount of non-comprehensive business is carried on, whilst for property and personal accident it may be as low as 0.6P. Since A - L in practice tends to be of the order of 0.3P, values of A will be 0.3P greater than L, that is will range from 3.8P to 0.9P. The idea is that k will provide a margin to cover flucuations from all causes. Now in 1974 variation in A was a fall of perhaps 40% whilst values of L were reassessed with rises of 10 to 20% or even more. Fluctuations in the value of A - L were from nil to a fall of as much as $3.8 \times (0.4)P + 35 \times (0.2)P = 2.22P$, so that $k = 0.12$ in a liability account is not going to be much of a defence in itself. Fortunately with rare exceptions this situation is exceptional and time is a great healer. It will be some time before L is ascertained and if companies are allowed to put their premium up long enough before the liabilities have to be settled they should be able to remain in business.

This arithmetic is performed to demonstrate that the theory behind the solvency margin deserves some scrutiny, although if realistic assumptions (i.e. those corresponding with what actually happened in 1974 and early 1975 are adopted) the size of k can reach alarming and virtually unattainable heights. There is obviously scope for examination of the variances of A and L arising from causes other than inflation and for fixing a seemingly more rational basis for solvency margins, although if recent experience is a guide the results might do more harm than good!

As we said earlier, the idea of finding a solvency margin and examining its trends is, seemingly, to detect adverse features in time to take action to avoid loss. Even from what little has been said so far, it should be clear that however much effort is expended on estimating the expected value of L, in many cases the additional accuracy will be small compared with errors and uncertainty arising elsewhere. When it is realised that returns cannot be reasonably expected until 5 or 6 months after the end of an accounting period (if one insists upon having them earlier it will be at the cost of more guesswork and unreliability) and that it will be a further 2 or 3 months (even with the aid of a computer; currently it is 6 to 12 months) before the result can be intelligently scrutinized,

it is obvious that at least a year will have elapsed before action can be initiated by a supervisory authority and longer still before it can become effective. If the company is losing money fast enough to have brought on this position the year will have made things worse; if it is profitable, however, things will be getting better and the need for intervention may have gone.

In other words whilst we are fundamentally concerned with solvency the important influence is profitability. A company making real profits will not become insolvent (although if it overtrades on its existing capital it may be technically insolvent on the statutory basis) whilst one losing money will be come a candidate for action long before the statistical returns reveal the fact.

Unfortunately we are to some extent arguing in a circle. Profitability on a cohort basis depends on the premium collected and the outgoings, including commission, expenses and claims, the claims for any recent year depending to a greater or lesser extent on outstanding liabilities and so to a lesser degree do the expenses that will be incurred in settling them. So far as one can perceive there have been two tests for liabilities. One is a comparison with the experience of other companies, the other is an examination of the run-off of payments compared with earlier estimates. Experience has shown that whilst the 1968 regulations seem to have had both objects in mind, the practical value is slight and in spite of improvements in detail proposed for 1976 and later, are likely to remain slight. So far as comparisons with other companies are concerned it is clear that they differ greatly from each other and the experience of one insurer may well be a poor guide to another with a different type or source of business. Insurers do differ and will continue to differ notwithstanding the repeated claims of the D.O.T. that they must treat them all alike. Since we do not unfortunately share Alice's Wonderland merely saying that they are the same either three or three hundred times makes no impact on the obstinate fact that they are different. Experience has shown that company practice varies for many reasons, and the fact that a company under-estimated in 1970 (maybe quite properly refusing to believe in 25% or 30% inflation which was going to overtake them before the slower claims had been settled) does not imply that it must still be under-estimating in 1975. The upholders of the value of run-off statements have a duty to explain their value· they have not yet done so and there seems little likelihood that they will ever be able to without more knowledge of the basis used in setting up a reserve. The subject was dealt with more extensively in a paper by Abbott et al JIA 101 part 2, PP217-283. That paper was largely concerned with criticising the definitions adopted in the regulations and the way they had been interpreted or misinterpreted by insurers rather than discussing the statistics which should have been collected, whilst the debate on the paper with one or two exceptions was concerned with methods of fixing liabilities. New regulations currently under consideration by the Department of Trade will go a long way to remedying the defects of detail noted in the Abbott paper, at the cost of a large increase of volume, but whether the statistics thereby collected will be of any more value in detecting a move to insolvency before the insolvency actually happens is very much open to doubt. Statistics do not necessarily increase in value when they become more numerous or more detailed.

HOW COULD THE SYSTEM BE

a. simplified

b. made more effective

This is where there is the greatest danger of overlapping with the work of the other groups. However the first thing that must be observed is that information in relation to a small insurer may require considerable detail and in particular be concerned more with the net position after reinsurance than with the gross position.

If a small company reinsures under stop loss or excess of loss contracts the bulk of the risks it writes, then what happens gross is of little interest. On the other hand for a large world wide company the gross position is likely to be the more important and many of the smaller risk groups or smaller countries are likely to be of little interest or value (assuming the currency risk is absent). Regulations that treat both insurers identically are bound to fail; if they are adequate for small companies they will give rise to a wholly excessive volume of information from the larger ones whilst if the volume from the larger companies is to be managable the smaller returns will be uninformative.

This position has been recognized in the United Kingdom by the Bank of England which has different rules for different sizes and kinds of bank and in the United States of America where the Insurance Commissioners now seem to recognise that supervision should be aimed more at the smaller companies if time and assets are not to be wasted. It is a fundamental rule of statistical investigation that the data collected should, in addition to being accurate and unbiased, be suitable for the object in view. Until this is recognised officially, statutory statistics will remain a waste of time and assets on the part of all concerned by producers and consumers and there seems little point in examining the 1975 proposals in detail.

QUARTERLY RETURNS

It is clearly important to be aware of adverse movements as soon as possible and the statistician must first ascertain as best he can what might be described as a normal situation inorder that he can measure whether, and if so by how much, there is a move away from normal. This poses some serious problems since we can only find out how far one has got in recording information on a file or in settling claims by knowing how far one still has to go. We have considered in an earlier section the problems caused by delays in recording information on a file; it is always to some extent an article of faith to assume that what is still to come is a normal back-log and whilst one can have doubts one can rarely be quite sure. Similarly with claim settlements one can never be sure how much has not been paid and neither statistical calculations (e.g. percentage paid) nor individual claim estimating is wholly reliable.

So far as premiums are concerned one can, in the absence of a severe interruption in work (strikes or computer breakdowns) form fairly reliable quarterly estimates within a few weeks of the end of a quarter but care must be taken to deal properly with adjustments to estimates in respect of previous quarters since such adjustments could very easily obscure trends.

For claim settlement, however, the position is much more difficult and case estimating on a quartedy basis is liable to be too expensive in relation to its value. It is possible to have continuous estimating of claims which are over say one year old but the uncertainty in the shorter and more recent claims makes forecasting absolute figures difficult and forecasting trends virtually impossible. One method being developed is to use standard tables for the rate of settlement of claims in terms of both numbers of payments and proportionate amounts, with separate tables for each class of business that seems likely to produce different patterns. At present five tables are in use, one for each private car comprehensive, private car non-comprehensive, fleet, other comprehensive and other non-comprehensive. By applying these tables to claims actually notified we can, for any cohort and any given time of development, calculate the expected number of payments and the expected amounts. The ratio of actual to expected will then give:-

for numbers : an absolute measure of progress in settlement

for amounts : the projected average claim amount.

Obviously any major change in practice (for example in the frequency of calling for police reports or the payment of emergency treatment fees in batches) could give a misleading indication, but preliminary results suggest that the method may be very useful. This is scope for much research in this area. It must be emphasised very strongly that merely collecting quarterly figures and looking at the trends from the last one or two quarter's figures is liable to be extremely misleading unless one knows precisely what one is doing and what disturbing influences may have arisen in each of those quarters.

THE COMPARISON OF ESTIMATES WITH ULTIMATE SETTLEMENTS

Run-off statements will show over a period whether the estimates for a cohort have been correct in total. It is however sometimes useful to monitor settlements compared with estimates at more frequent intervals although there are a number of difficulties involved.

If statistical estimating takes place we can compare settlement with estimate for any claims that have actually been estimated on an individual basis, and if for example a sample of 1 in m claims has been estimated we can also compare total settlements with m x estimate for such of the settled claims as came into the sample. Experience with one large portfolio suggests that in that case the number of estimated claims settled was neither large enough or sufficiently representative to give a fair test. It was also found that surprisingly few claims with larger estimates at the start of each year were actually treated as settled in that year, largely because a small outstanding payment for charges or the expectation of recoveries (other than under excess of loss reinsurance) remained even though the major payment had been made in the course of theyear.

As a result of these factors it was found that settled claims showed a large profit, but that payments on account and closing estimates were substantially inexcess of opening estimates. The net balance was much smaller than either the apparent profits or losses. The reasons for this situation are probably quite numerous and varied, but they have shown quite clearly that even with an adequate total reserve the actual reserves on individual claims were often most unreliable. Whilst this should not be regarded as a defect (since the aim is to set up reserves which are adequate but not excessive, and if every claim is estimated on a conservative level the result would be a gross overestimate), it does show that estimating every individual case does not necessarily improve the accuracy of the total estimate compared with statistical methods.

CONCLUSION

We have dealt with some aspects of recording and analysing information realting to a motor portfolio. We have tried to show some important features that ought to be incorporated in any system. We have also referred to the major differences between companies in what they record, how they record it, and how they analyse it.

There are large areas we have covered only briefly and we have indicated large areas where the present state of our knowledge is less than we would wish. In other classes of non-life business e.g. General and Employers Liability even greater practical difficulties arise. We need more data, lots more data , and more critical analysis of them. If we have succeeded in this paper in stimulating discussion and obtain offers of further data we shall be satisfied.

References

1. Johnson P.D. + Hey G.B. JIA 97 p199
2. Hey G.B. J.R. Statist Soc 133 p56
3. Scurfield H.H. JSS 18 p207

APPENDIX 1. TYPICAL NEW BUSINESS PIPELINE

PROCESSING MONTH (1974)

Month of Inception	JAN.	FEB.	MAR.	APR.	MAY.	JUN.	JUL.	AUG.	SEP.	OCT.	NOV.	DEC.	Proc. By Dec. 74
1973 JAN.													*
FEB.													
MAR.	5	1											
APR.	4	3		2									
MAY.	4	4	1	2	1								
JUN.	5	2	3		2								
JUL.	9	4	5	2		3							
AUG.	25	8	5	4	1	2	1						
SEP.	42	18	12	7	2	1	2		2				
OCT.	70	35	20	8	7	2	4	1					
NOV.	175	65	37	23	10	5	3	2					
DEC.	485	185	55	35	15	8	5	2		4			*
1974 JAN.	210	470	185	55	37	20	8	5	4	3	3		1,000
FEB.		220	460	175	70	30	18	12	6	3	3	2	999
MAR.			215	475	175	60	35	20	6	6	4		996
APR.				105	230	92	35	20	9	3	1	1	496
MAY.				2	103	242	80	33	18	11	3	3	495
JUN.						100	237	88	32	20	13	2	492
JUL.						1	110	232	95	25	17	9	490
AUG.								107	238	85	32	18	480
SEP.									100	235	92	33	460
OCT.										107	235	88	430
NOV.											110	230	340
DEC.												105	105
TOTAL	1034	4015	998	895	653	566	538	523	510	502	513	419	-

* The 1974 figures are not very meaningful. If those for 1973 were included we should have a picture of the flow of business

APPENDIX 2

The effect of non-uniform distribution of renewal dates.

The table below illustrates the effect of a non-uniform distribution of renewal dates on an otherwise stable portfolio.

The figures assume:-

(a) that new business is stable at 28.0% of renewals offered.
(b) that lapses are stable at 15% of renewals offered.
(c) that mid-term cancellations are stable at 3% per quarter of business in force at the start of the quarter.
(d) that new business and lapses take place at the end of the quarter.

The portfolio is analysed into quarters of renewal, and is examined quarterly.

Period of Account 1974	Q1 Renewals	Q2 Renewals	Q3 Renewals	Q4 Renewals	Whole Portfolio
In Force at 1.1.74	91,267	94,090	97,000	200,000	482,357
Mid-Term Cancellations (3%)	2,738	2,823	2,910	6,000	-14,471
Renewals Offered	88,529	-	-	-	(88,529)
Lapses (15%)	13,279	-	-	-	-13,279
New Business (28%)	24,750	-	-	-	+24,750
In Force at 1.4.74	100,000	91,267	94,090	194,000	479,357
Mid-Term Cancellations (3%)	3,000	2,738	2,823	5,820	-14,381
Renewals Offered	-	88,529	-	-	(88,529)
Lapses (15%)	-	13,279	-	-	-13,279
New Business (28%)	-	24,750	-	-	+24,750
In Force at 1.7.74	97,000	100,000	91,267	188,180	476,447
Mid-Term Cancellations (3%)	2,910	3,000	2,738	5,646	-14,294
Renewals Offered	-	-	88,529	-	(88,529)
Lapses (15%)	-	-	13,279	-	-13,279
New Business (28%)	-	-	24,750	-	+24,750
In Force at 1.10.74	94,090	97,000	100,000	182,534	473,624
Mid-Term Cancellations (3%)	2,823	2,910	3,000	5,476	-14,209
Renewals Offered	-	-	-	177,058	(177,058)
Lapses (15%)	-	-	-	26,558	-26,558
New Business (28%)	-	-	-	49,500	+49,500
In Force at 1.1.75	91,267	94,090	97,000	200,000	482,357

APPENDIX 3

CLAIM FREQUENCY ANALYSIS Year of account ending 31st December 1974

		Company ref	Country	Class	Risk Group	Year
For DOT use	100					

Country:	Class:	Risk group:
United Kingdom	Motor vehicle insurance business	PRIVATE CARS

		Gross premiums 1	Units of exposure 2 vehicle/years
		£000	
A Exposure during year of account: (*a*) under contracts commencing during the previous financial year	1	800	26000
(*b*) under contracts commencing during the year of account	2	950	27,000
(*c*) total (*a*) + (*b*)	3	1750	52,000
B Exposure carried forward to financial year following the year of account in respect of contracts commencing during the year of account	4	850	26000

		Number of claims
C Number of claims: (*a*) originating in earlier financial years and first notified during the year of account	5	500
(*b*) closed in earlier financial years and reopened during the year of account	6	200
(*c*) originating during the year of account and notified before closure of the records for that year ..	7	6000
(*d*) estimated number of claims originating during the year of account but not notified before the closure of the records for that year	8	300
(*e*) estimated total number of claims attributable to the year of account (*c*) + (*d*)	9	6300

		Per cent
D Claim frequency in year of account (where units of exposure are given)	10	12.1

APPENDIX

CLAIM SETTLEMENT ANALYSIS Year of account ending 31st December 1974

	Company ref	Country	Class	Risk group	Year of origin	Year of account	No of claims
For DOT use 300							

Country:	Class:	Risk group:	Year of origin:	Total number of claims attributable to the year of origin
United Kingdom	Motor vehicle insurance business	PRIVATE CARS	1970	8000

Year		Number of claims closed in year: at no cost	Number of claims closed in year: at some cost	Number of claims outstanding at the end of the year	Amount of payments made in the year in settlement or on account	Aggregate payments made up to the end of the year	Claims outstanding at the end of the year: Payments on account included in column 6	Claims outstanding at the end of the year: Estimated payments remaining to be made	Total amount paid and outstanding at the end of the year (Total of columns 6 and 8)
1		2	3	4	5	6	7	8	9
					£000	£000	£000	£000	£000
Year of origin 1970	1	1,500	3,500	3000	350	350	70	350	700
Year of account 1971	2	500	1,700	300	165	515	50	125	640
Year of account 1972	3	15	180	300	55	570	40	90	660
Year of account 1973	4	5	60	200	80	650	30	60	710
Year of account 1974	5	2	60	100	70	720	30	30	750