

Continuous Mortality Investigation

Working Paper 57

CMI Life Office Mortality ‘Per Policy’ Results: Consultation on the Future Format of Results

Contents

Executive Summary	2
1. Introduction	3
2. The future format of results from Per Policy data: Introduction	5
3. Results from Per Policy data: Detailed Results	6
4. Future Developments.....	15
5. Results from Per Policy data: Areas for Consultation.....	17
References.....	19
Appendix 1: Example file size for Detailed Results produced under Approaches 1 and 2	20
Appendix 2: Per Policy data fields and possible Attributes.....	22

September 2011

© 2011 Institute and Faculty of Actuaries

The text in this document may be reproduced free of charge in any format or medium providing that it is reproduced accurately and not used in a misleading context. The material must be acknowledged as Institute and Faculty of Actuaries copyright and the title of the document specified.

The Institute and Faculty of Actuaries and the CMI do not accept or assume any responsibility for the use of this document by any party in any context. This document does not provide any form of guidance and should not be relied on as such.

Executive Summary

The CMI Life Office Mortality investigation has, until recently, collected what is termed ‘scheduled’ data. That is, the data provided by contributing offices has been in a census format containing the total numbers of policies (in force and deaths) split by age and duration. No policy specific details were requested from contributors which means that the level of analysis that could be carried out on the data has been limited.

In recent years the CMI has been switching to ‘Per Policy’ data collection for life office mortality data. The aim is to capture more detailed information which will allow analyses that would previously have been impossible as well as improving the accuracy of the analysis.

The CMI Life Office Mortality Committee has now decided on the methodology and format of results to be used initially using Per Policy mortality data. These are set out in Working Paper 56, published simultaneously with this paper. With the level of detail included in the Per Policy submissions, the Committee has the opportunity to extend these analyses greatly. However the initial results will not use much of the additional data that is available in Per Policy data and this paper seeks views on how the results should be developed in future.

The Committee proposes issuing two distinct types of results in future: “Summary Results” and “Detailed Results”. Summary Results are very similar to the initial results described in Working Paper 56 and are only briefly discussed in this paper.

In contrast, Detailed Results could be very different from any results previously issued by the CMI; consequently these results are described and discussed at length and the Committee seeks views on many aspects of these results. Detailed Results could contain many of the fields from the original data submitted to the CMI, except for those that could infringe the confidentiality of any individual or any data contributor, together with values of exposed to risk, expected and actual deaths. We anticipate these will allow considerable flexibility with regard to areas that are determined by the Committee in Summary Results but, as a result, they will be very large files that require considerable manipulation by the user to produce useable results.

In particular, the Committee is consulting on two distinct approaches to the structure of records for Detailed Results; one mirrors the format of the data that is submitted to the CMI, whilst the second aggregates across policy records but at a very granular level. As a result of the confidentiality issues and the very large file sizes arising under the first approach, the Committee is proposing to use the second approach for the record structure for both Individual Office and All Office Detailed Results.

Proposals are also set out in the paper for the content of the records in Detailed Results, the Product Categories to be used, and the medium to be used for sending out results. Views are sought on all these proposals.

A number of areas that are currently outside the scope for results using Per Policy data are also discussed in this paper, including the use of postcode.

The paper concludes with a number of specific questions on which feedback is sought in relation to the future development of Per Policy results. Responses are requested by 30 November 2011.

Continuous Mortality Investigation

Working Paper 57

CMI Life Office Mortality ‘Per Policy’ Results: Consultation on the Future Format of Results

1. Introduction

1.1 Background

The CMI Life Office Mortality investigation has, until recently, collected what is termed ‘scheduled’ data. That is, the data provided by contributing offices has been in a census format containing the total numbers of policies (in force and deaths) split by age and duration. No policy specific details were requested from contributors which means that the level of analysis that could be carried out on the data has been limited.

In recent years the CMI has been switching to ‘Per Policy’ data collection for life office mortality data. The aim of this is to capture more detailed information which will allow analyses that would previously have been impossible as well as improving the accuracy of the analysis.

Under the Per Policy submission requirements, offices are asked to submit a separate record for each benefit of each life insured on each policy for each period that this benefit is in force with unchanged details within a calendar year. This means that more than one record per year is required for many policies. For example, an additional record is required if a policy is taken out of force during the year and brought back into force as is the case with an alteration. Each record should occupy one “row” in the medium of submission (e.g. one row in a spreadsheet, one database record or one text line). The information requested in each record is set out in the [Coding Guide](#).

A significant number of offices have now supplied Per Policy data and considerable progress has been made in verifying and processing this data. However, the additional time and effort required by offices and the CMI Secretariat to validate Per Policy data means that it has not yet been possible to issue results based only on Per Policy data. Some of the areas within the Per Policy data requirements that are causing difficulty are discussed in section 2 of Working Paper 56, published simultaneously with this paper.

The CMI Life Office Mortality Committee has now decided on the methodology and format of results to be used initially using Per Policy mortality data. These are set out in Working Paper 56. With the level of detail included in the Per Policy submissions, the Committee has the opportunity to extend these analyses greatly. However the initial results will not use much of the additional data that is available in Per Policy data and this paper seeks views on how the results should be developed in future.

1.2 The scope of this Working Paper

The results issued by the CMI in recent years comprise the results for each Individual Office, which are sent to that data contributor only, and the All Office results, which are sent to all member offices. These are produced for both a single year's results and for four years' results combined (quadrennial results).

The CMI has also published a summary of the experience of each successive quadrennium in CMI Reports or Working Papers. These publications are not considered within this paper.

As with Working Paper 56, this paper focuses on mortality data. Many of the considerations are also relevant to Per Policy data for the Critical Illness investigation; however the change in data format is less radical for the Critical Illness investigation which was already collecting individual data records containing life- and policy-specific details. Consequently, the Critical Illness Committee has yet to consider in detail how it will adapt its results for Per Policy data. As far as possible, it will seek consistency with the Life Office Mortality investigation. This has benefits both to the CMI, in terms of systems and processes, but more importantly to practitioners seeking to understand CMI results. However, particular features of the two datasets may lead the Critical Illness Committee to take a different approach in some areas.

1.3 The structure of this Working Paper

The Committee proposes issuing two types of results in future: "Summary Results" and "Detailed Results". An overview of the two types of results is set out in section 2. Summary Results are very similar to the initial results described in Working Paper 56 and hence are not described in detail in this paper whilst Detailed Results are discussed in section 3.

A number of areas that are currently outside the scope for results using Per Policy data are discussed in section 4. The specific questions on which feedback is sought in relation to future results are then set out in section 5.

1.4 Next Steps

Feedback is sought from all interested parties on the development of Per Policy results, with the consultation period ending 30 November 2011.

2. The future format of results from Per Policy data: Introduction

In this section we consider the results that the CMI will regularly issue in future for Life Office Mortality using Per Policy data.

We are consulting on issuing two distinct types of results using Per Policy data which we refer to as “Summary Results” and “Detailed Results” in the remainder of this paper. In this section we provide a brief overview of the two types of results.

2.1 Summary Results

“Summary Results” are expected to be a natural extension of the results that the CMI has previously issued and are very similar to the initial results described in Working Paper 56. Hence, they do not need any further description in this paper.

One of the principal aims of moving to Per Policy data was to produce more valuable analyses by including new risk factors. With the level of detail included in the Per Policy submissions, the CMI has the opportunity to analyse experience using a greatly increased number of categories. However for Summary Results to be manageable, they must necessarily lose much of the detail that is available in Per Policy data. It is therefore proposed that the experience for these additional factors will initially be investigated by the Committee with the results reported in ad hoc reports, such as Working Papers; such factors will then only be incorporated into regular reports if they are adjudged to be valid risk factors.

In the consultation we seek views on these Summary Results, including the categories by which the results are presented.

2.2 Detailed Results

The second type of results we are consulting on are referred to as “Detailed Results”; these could contain many of the fields from the original data submitted to the CMI (except for those that could infringe the confidentiality of any individual or any data contributor). Values of exposed to risk, expected and actual deaths will also be included. We anticipate these will allow considerable flexibility with regard to areas that are determined by the Committee in Summary Results but, as a result, they will be very large files that require considerable manipulation by the user to produce useable results.

Further discussion of Detailed Results is contained in section 3. These results could be very different from the results previously issued by the CMI. Consequently the consultation seeks views on many aspects of these results.

3. Results from Per Policy data: Detailed Results

As noted earlier, the second type of results we are consulting on are referred to as “Detailed Results”. These are described and discussed in the section.

Many of the areas discussed in this section are inter-related – in particular two possible structures for the records contained in Detailed Results are described first, under 3.1 below, and the content of the records (under each structure) is then described under 3.2. The structure and content are then discussed together under 3.3. Other areas are discussed in the remainder of this section, concluding with an initial proposed format for Detailed Results in section 3.10.

3.1 Record structure

Two approaches to the structure of records for Detailed Results are considered below.

Approach 1: Data-level records

Under this approach the results records would correspond to the records in the original data, with each results record relating to an individual benefit under a policy. Each record would contain (some of) the data fields submitted to the CMI for that policy/life/benefit (for example, gender). In addition, the CMI Secretariat would append fields to each record, such as the age and duration (at the start of the year) and exposed to risk, actual and expected claims for the four age and duration combinations that could be attained during the investigation year. The corresponding fields for exposed to risk, actual and expected claims on an Amounts basis would also be appended.

This approach is perhaps best explained by an example. Suppose a data submission consists of just the following four records, three of which are in force throughout 2010 and one is an exit due to death:

Record Type	Sex	Date of Birth	Date of Benefit Commencement	Type of Exit	Date of Death
I	M	02/01/1950	01/03/2000	-	-
I	M	01/02/1950	02/01/2000	-	-
I	M	01/11/1949	01/12/1999	-	-
O	M	01/03/1950	01/02/2000	D	01/10/2010

NB only a small number of data fields are shown and, in particular, the example considers lives analysis only.

Under Approach 1, there would necessarily be an identical number of results records. Consequently the records might look as follows:

Record Type	Sex	Date of Birth	Date of Benefit Commencement	Type of Exit	Date of Death	Age last at start of year (x)	Curtate Duration at start of year (t)
I	M	01/01/1950	01/03/2000	-	-	59	9
I	M	01/02/1950	01/01/2000	-	-	59	9
I	M	01/11/1949	01/12/1999	-	-	60	10
O	M	01/03/1950	01/02/2000	D	01/10/2010	59	9

x,t			x+1,t			x,t+1			x+1,t+1		
Exposure	A	E									
0.0027	0	0.0000	0.1589	0	0.0009	-	-	-	0.8384	0	0.0048
0.0027	0	0.0000	-	-	-	0.0822	0	0.0004	0.9151	0	0.0052
0.8329	0	0.0048	0.0822	0	0.0005	-	-	-	0.0849	0	0.0005
0.0849	0	0.0004	-	-	-	0.0767	0	0.0004	0.5890	1	0.0034

Note that:

- A = Actual claims and E = Expected claims;
- Each data record contributes exposure (and hence expected claims) in up to three age/duration cells during the investigation year;
- Exposure is calculated on a day-count basis, consistent with the initial methodology set out in Working Paper 56;
- Expected claims are calculated in this example using the AMC00 table; and
- No values of Actual/Expected are included since these are meaningless at an individual record level.

Approach 2: Result-level records

Under Approach 2, the results records would aggregate across all the records in the original data which share common features at any time during the year; we refer to these common features as “Attributes” in the remainder of this paper. Each results record would contain the set of Attributes plus values for the aggregated exposure, actual and expected claims on both a Lives and an Amounts basis. The records would therefore be much closer to the information provided in Summary Results.

Using the same four records as above, and limiting the Attributes to gender, age and duration, the results records under Approach 2 would be:

Sex	Age Last (x)	Duration (t)	Exposure (Years)	Actual	Expected	100 A/E
M	59	9	0.0904	-	0.0005	-
M	59	10	0.1589	-	0.0008	-
M	60	9	0.1589	-	0.0009	-
M	60	10	3.1753	1	0.0182	5493.2263
M	61	10	0.0822	-	0.0005	-
M	61	11	0.0849	-	0.0005	-

Note that the number of results records in this example exceeds the number of data records, although one would expect the converse to apply for larger datasets. An example of the numbers of records under each structure is contained in Appendix 1.

3.2 The content of records

One of the principal aims of moving to Per Policy data was to produce more valuable analyses by including new risk factors. In this section we consider which fields from the overall dataset can be made available to users for analysis purposes in Detailed Results. These may be quite different between the two approaches described above:

- In theory, all of the data fields requested in the Per Policy coding guide could be included in the records under Approach 1, for Individual Office results, together with additional fields (such as exposure) generated by the CMI.
- Some of the data fields would need to be removed from records if Approach 1 is adopted for All Office results in order to preserve the confidentiality of individuals and data contributors.
- Under Approach 2, including a very large number of Attributes could produce as large a data file as Approach 1, with little aggregated data in each record. It may therefore be easier to consider which factors are included as Attributes, rather than which are excluded.

In addition to using a data field directly as an Attribute in Detailed Results (examples include gender and smoker status) they can also be used indirectly, where one or more data fields are used by the CMI Secretariat to determine a new field for inclusion in the Detailed Results. An example of such an “Indirect Attribute” is age, which will be calculated by the CMI from the date of birth supplied in the data.

Appendix 2 sets out the data fields that are requested in Per Policy data and the possible use for each data field, in particular whether it can be used Directly or Indirectly, which may vary according to whether the record structure is “data-level records” (Approach 1) or “results-level records” (Approach 2) and whether the results are at an Individual Office or All Office level. The possible scenarios for each data field are set out in Appendix 2 along with comments to clarify their possible use.

The contents of records are now considered further separately for their different uses.

Approach 1 – Individual Office

For Individual Office results produced under Approach 1 there is, in theory, no limit on the number of Attributes that could be included in the Detailed Results. All of the fields provided to the CMI in the original data can be provided back to the data contributor, together with additional fields derived from the data, such as Product Category, age and duration. This will provide offices with all the information needed to perform their own analyses and the additional fields will assist offices to produce these consistent with Summary Results (for example, by showing both the office’s product type and the Product Category to which it has been assigned by the CMI Secretariat).

The number of records in the data will equal the number originally supplied; each additional Attribute increases the length of each record.

Approach 1 – All Office

The principal issue that surfaces with All Office Detailed Results produced under Approach 1 is confidentiality. Data fields such as date of birth and postcode cannot both be included without breaching policyholder confidentiality, whilst the product type may be unique to an office so its inclusion could breach office confidentiality. As a result the Attributes may need to be limited, and the record format different, compared with Individual Office Detailed Results. Appendix 2 indicates which data fields cannot be included as Attributes in All Office Detailed Results under Approach 1.

In addition to the removal of certain fields, additional measures might be required to preserve confidentiality:

- The provision of life-specific records may allow the identity of individuals to be surmised, even if the exact date of birth has been removed, for example where very large benefit amounts occur at ages where data is scarce. This could perhaps be overcome by grouping together lives over a certain age (say 90), rather than including individual records.
- Considerable care will be required to ensure that no combination of fields (for example Product Category, Distributional Channel and Rate of Increment) results in the records of a single office becoming identifiable.
- Some data fields are optional, which may mean that the CMI does not receive sufficient data to allow their inclusion in the Detailed Results, without a further risk that an individual office's experience can be identified. (Note that this also applies to mandatory fields with an "unknown" option.)

Such measures potentially lead to further inconsistencies between Individual Office and All Office Detailed Results under Approach 1.

Issues may also arise from the size of All Office Detailed Results files – see the example in Appendix 1.

Approach 2

There are clear benefits to using as many fields as possible as Attributes, so that they are available to users for their own analyses (subject to data contributors' consent). However as the number of Attributes is increased, the number of results records increases at a faster rate so file size may become an issue for All Office Detailed Results. Furthermore the volume of data within each record reduces, again giving rise to possible confidentiality issues for combinations of Attributes, as discussed above.

Consequently – if Approach 2 is adopted for the record structure – it may be appropriate to limit the number of Attributes, at least initially to those with known value.

3.3 Discussion

There are advantages and disadvantages to each of the approaches to the record structure and the inevitable consequences for the contents of records. These are discussed below. In some cases, the arguments depend on whether one is considering Detailed Results at an Individual Office level, to be sent back only to the relevant data contributor, or at an All Offices level. These are therefore considered separately.

Individual Office Results

The biggest potential advantage to Approach 1 is that (if required) every field that was provided in a data record could be returned in the corresponding results record. This provides a considerable amount of flexibility to users to undertake their own analyses. For example if the results records include “date of commencement”, users could undertake analyses of experience with duration measured in months or quarters. (Such analyses would obviously not be possible under Approach 2, if policy duration is included in the results record using curtate duration in years.)

Note that the analyses that could be undertaken under Approach 1 are not restricted to the fields that are actually contained in the results records. Only a small proportion of the universe of fields that can be derived indirectly is referred to in Appendix 2. To take one example, in the Per Policy data requirements we request “Date of policy commencement” and “Date of benefit commencement”; where an office allows multiple benefits under a single policy, the latter will be used indirectly to determine duration and underwriting year (if not, the date of policy commencement will be used). Both could be included as fields themselves, but the two could also be used to derive an additional field, “Initial or subsequent benefit”, and thereby investigate the experience of “subsequent benefits” compared to “first benefits”, if data volumes permitted.

An additional advantage is that users would be able to see the contribution made by each specific policy (indeed the policy identifier can be included within each results record). This might make it easier for users to understand and reproduce their Individual Office Summary Results and to reconcile the results with internal analyses, if required.

One disadvantage is that users might need to undertake greater work in order to generate results consistent with those in Summary Results or published in a CMI Working Paper, for example using a different table to calculate expected claims.

Another disadvantage is that Approach 1 would lead to large results files, as there would be one results record for each life under each policy. This might impact on the medium used to dispatch results, discussed in 3.7, below. An example of the potential file sizes under the two approaches is contained in Appendix 1.

Note that many of the data fields should not vary from one calendar year to the next, so – in theory at least – additional fields could be appended to the records in the following year, for example in relation to the exposure and expected and actual claims (for those in force in the previous year; new records would be added for new policies). Consequently, the number of records would not increase rapidly for results covering multiple calendar years.

The feasibility of “extending the records” in subsequent calendar years has not yet been investigated. It does, of course, presume that data records can be consistently identified from one year to the next and additional practical issues would arise from a late-reported death or a change in smoking status, for example. Such changes would need to be identified within results records in such a way that the date of change is apparent for users seeking to produce results consistent with Summary Results already issued for prior years. This may mean that the number of “stationary” fields, across years, is insufficient to justify this approach. If it proved impractical for the CMI to extend records across calendar years, then Approach 1 could be applied to results records for each individual year; users would then need to access multiple files to produce results across calendar years.

Approach 2 is likely to be more immediately familiar to users, since the results records would resemble the current results format more closely, as well as the proposed format of Summary Results. Consequently it should be easier for users to produce analyses consistent with CMI results. The size of the file might also be smaller, depending on the number of Attributes included in each record.

The principal disadvantage would be that the fields in each record would need to be restricted in order to keep the results files to a manageable size. Consequently fields that might be of interest to users would not be made available. Note that this argument applies to fields derived indirectly, as well as those contained in the original data submissions.

All Office Results

Many of the considerations discussed above apply to both the Individual and All Office results. Extra considerations relating to the All Office results are explored below.

The main disadvantage of using Approach 1 for the All Office Detailed Results is confidentiality. As discussed earlier, if All Office Detailed Results were produced under Approach 1, the CMI would have to ensure that confidentiality is protected for both individual lives and individual offices. This means that some fields would need to be removed from the All Office results, compared with the Individual Office results, and other measures might also be required, resulting in an inconsistency between Individual Office and All Office results that had not previously existed and reducing the value of the Detailed Results to users. There might also be inconsistencies in the fields that can be made available from one year to the next.

In contrast, under Approach 2 the results records would not contain any policy- or life-specific information and hence there will be fewer confidentiality issues.

3.4 Product Categories

As noted earlier, the “product type” (supplied in Per Policy data) will not be included in All Office Detailed Results, as this may be unique to an office so its inclusion could breach office confidentiality; hence the Product Categories are a key field for users of All Office Detailed Results. Although product type can be included in Individual Office Detailed Results, we envisage the Product Category will also be included to allow data contributors to see how their data has been mapped to the Product Categories.

Product Categories are considered in more detail in Working Paper 56, which sets out a proposed hierarchy. Unlike the Summary Results, there is no requirement to restrict the number of Product Categories to a manageable level in Detailed Results – the greatest flexibility is achieved by using the lowest level possible as users will be able to aggregate these records up to higher level categories as they wish. Subject to protecting office confidentiality within All Office results, there does not seem to be any reason to use anything other than the lowest level of Product Category in Detailed Results.

3.5 Expected Claims

For Detailed Results, all the information will be at a sufficiently granular level (e.g. individual ages) to allow users to apply their own choice of table. However it might still be useful for

expected claims to be included as a check; these would be calculated using a single table chosen by the Committee, consistent with that used in Summary Results.

3.6 Statistical Tests

As noted in Working Paper 56, statistical tests are currently provided in the scheduled mortality results. Users will have considerable flexibility to perform their own analyses on the data contained in the Detailed Results, and the provision of any statistical tests would therefore be more complex than in the Summary Results.

3.7 Medium for results

The key requirements for the medium for Detailed Results are the ability to hold a large volume of data and the capability to allow users to undertake their own analyses.

In terms of file size, commonly-used media (such as an Excel Spreadsheet or Access Database) may be too restrictive. Excel 2007 can hold up to 1,048,676 rows of data but earlier versions are limited to only 65,536 rows. Access databases are subject to a 2Gb limit. The volume of data in Detailed Results is dependent on the Results Structure (see 3.1); in particular Approach 1 will result in very large files whilst Approach 2 could also generate very large files, depending on the number of Attributes (see Appendix 1).

Consequently, although either Excel or Access would provide considerable flexibility for users, both media are likely to be inadequate for Detailed Results files unless these are produced using Approach 2 for the Results Structure with the number of Attributes curtailed (which obviously limits the analyses users can undertake).

The most suitable media would therefore appear to be a Text File. This would contain the information to allow a user to perform their own analyses but would necessitate a separate tool to produce any useable results.

To ensure that users have the necessary information to use the results accurately and consistently with published analyses, the Committee could provide a brief “user guide” with examples. Alternatively, the CMI might develop a tool to generate Summary Results internally and it may be feasible for this to be made available to users. The current intention is to use an OLAP cube, a form of data warehouse akin to a pivot table but with the ability to operate on much larger amounts of data. (Specifically, the CMI anticipates using Microsoft SQL Server Analysis Services as its OLAP cube.)

In contrast, if the fields included under Approach 2 were curtailed to a file size that enabled Detailed Results to be provided in an Excel Spreadsheet or an Access database, then standard pivot tables or queries could be provided, to facilitate the manipulation of Detailed Results. Users would then have the option of adjusting these pivot tables or queries to perform their own analyses.

Some form of secure web-based access might provide the optimal medium for Detailed Results. However the CMI is unsure whether all potential users would be able to access a single site and would be comfortable with the implications for data security (considered in 3.8 below). Consequently a more established form of transmission, such as CD, might be used initially.

3.8 Data Security

It is envisaged that the CMI Secretariat will zip, encrypt and password-protect Detailed Results before they are dispatched, whichever of the media discussed above is used. The encryption key and password will be advised via a separate means – probably telephone (if results are dispatched via e-mail), or e-mail (if results are dispatched via CD).

For Summary Results, we assume that the CMI Secretariat continues to send results to a named individual in each member office and that, once received, security becomes the responsibility of that organisation. However data contributors may have additional concerns for Detailed Results, for example arising from the existence of a CD containing a very large amount of office-specific information; if so, we would obviously be keen to understand these concerns.

Additional measures may need to be considered if Approach 1 is adopted for the Record Structure for Individual Office Detailed Results, if dates of birth and/or postcode are included, as these could then be considered personal data in the context of the Data Protection Act 1998.

3.9 Confidentiality

The CMI has an overriding responsibility to protect the confidentiality of individuals, where they are capable of being identified within data or results, and also to protect office-specific data and results.

There is clearly an increased risk of individuals being identified in Detailed Results, compared with Summary Results, particularly if Approach 1 is adopted for the Record Structure as the results records would relate to individual policies. As discussed previously, this will necessitate restricting the fields included in the results records for All Office Detailed Results.

The measures required to protect office confidentiality were also discussed earlier and are considerably increased if Approach 1 is adopted for the Record Structure. Given the number of fields that might be included in results records, it would not be straightforward for the CMI Secretariat to check the number of offices contributing data under every possible combination of fields, to ensure that there is no category representing a single office.

3.10 The Proposed Format of Detailed Results

The structure of Detailed Results (and indeed whether they are produced at all) will depend on the feedback to this consultation exercise. However the Committee considered that it might be helpful to readers to set out its thinking, as a summary to the preceding discussion.

Record structure

As a result of the confidentiality issues arising from Approach 1, we propose to use Approach 2 (Results-level records) for the record structure of All Office Detailed Results.

Although the confidentiality issues are less significant for Individual Office results, the Committee believes that consistency between the All Office and Individual Office results is desirable. We are therefore proposing to also use Approach 2 for the record structure of Individual Office Detailed Results.

The content of records

We propose to use the following Attributes (subject to data volumes) directly from the data fields received: territory, medical type, sex, smoker status (where relevant), original type of entry, single or joint life, rated or non-rated, ABI new business code and distribution channel.

In addition, the following Attributes will be derived by the CMI Secretariat and appended: Product Category, retirement type (where relevant), age last birthday and curtate duration.

The exposed to risk, actual and expected claims will also be included (on both a Lives basis and an Amounts basis), aggregated across all the records in the original data which share common values of that set of Attributes at any time during the year.

Product Categories

As discussed above, subject to protecting office confidentiality in All Office results, it is proposed to use the lowest level of Product Category shown in Appendix 2 of Working Paper 56 in Detailed Results.

Expected Claims

The expected claims will be calculated using the same basis as used in the Summary Results.

Statistical Tests

It is not proposed to include any statistical tests within Detailed Results.

Medium for results

The Committee proposes to issue Detailed Results as a text file.

The CMI would look into the possibility of providing a data tool to analyse Detailed Results in due course if there were sufficient demand.

Data Security

Initially, Detailed Results will be encrypted and dispatched on a password-protected CD sent via courier. The encryption key and password will be sent separately via e-mail.

The CMI would look into the possibility of providing Detailed Results via a secure website if there were sufficient demand.

4. Future Developments

In this section we consider a number of developments that are currently out of scope for results using Per Policy data, but where we are keen to receive feedback to influence the direction and prioritisation of our future work.

4.1 Postcode

If Detailed Results are produced using Approach 1 for the Record structure (i.e. “Data-level records”, see section 3.1), the full postcode could theoretically be included as each record relates to an individual life under a policy. However it is highly likely that a combination of full postcode with date of birth (and gender) is regarded as “personal data” in terms of the Data Protection Act 1998. Consequently the Committee would not include the full postcode in All Office Detailed Results. Although the issue differs for Individual Office results, we are not sure whether there is sufficient benefit in providing a data field back to the office that supplied it to warrant the CMI including full postcode, given the implications for additional data security measures required to safeguard personal data. The CMI would therefore NOT include the full postcode in Individual Office results unless specifically requested to do so by data contributors, if these were to be produced using Approach 1.

This question does not arise in the context of Summary Results, or for Detailed Results using Approach 2 for the Record structure (i.e. “Results-level records”), as the records then aggregate across multiple policies/lives and hence are not associated with a single postcode.

With the possible exception of Individual Office Detailed Results, noted above, in the context of this paper postcodes can therefore be considered as being used for two Indirect Attributes:

- I. One is to “map” the postcode to some form of socio-economic grouping, e.g. using Acorn or Mosaic. These could be included as an Attribute, in either Summary Results or Detailed Results, which would mean that results would indicate the relative mortality/morbidity experience for each socio-economic grouping (or combinations thereof) present in the data. The CMI has yet to undertake any work using such mappings but expects to commence initial analysis in the near future.
- II. A second is to use part of the postcode as an Attribute; for example the postcode PO16 7DZ could be shortened to PO (postcode area) or PO16 (postcode district). Analysis of the experience at this level may be of limited value, as the areas are likely to contain considerable heterogeneity however the geographical spread of the dataset could be considered. We would welcome feedback on the perceived value of including this as an Attribute, but note that legal opinion on the definition of personal data would be needed before we include this in All Office Detailed Results.

Both of these are therefore considered future developments as far as this paper is concerned, although we are keen to receive feedback to influence the direction of our work in this area.

4.2 Lapse analyses

This paper focuses on mortality and morbidity analyses but, within Per Policy data, the CMI is also now collecting information on other types of exit, including lapses and surrenders. In theory, it would therefore be relatively straightforward for the CMI to also generate analyses of these other types of exit, using similar or identical methodology and results formats to those adopted for mortality and morbidity. Some practical issues are bound to arise – in

particular, it might be necessary to introduce additional checks for reasonableness on the numbers of, and the dates relating to, these other types of exit.

The CMI is conscious that this represents an extension of the scope of its traditional role and so is potentially a sensitive area. As such we do not propose that the CMI produces such results, unless there is significant demand from users and consent from contributors. Views are therefore invited on the value of extending the CMI's traditional work in this manner.

4.3 Impaired Lives

From 1982, the CMI ran an investigation into the mortality of impaired assured lives, covering an extensive list of impairments, including Hypertension, Ischaemic Heart Disease, Cerebrovascular Disorders and Tumours.

The complexity of the data requirements, and the likelihood that such detail is not held on life insurers' principal systems, meant that this investigation only ever received data from a small number of insurers. When the first version of the Per Policy Coding Guide was issued, alongside [CMI Working Paper 19](#) in December 2005, the existing impairment codes were retained to try to continue capturing data from the few offices that were able to provide it.

A report on the experience in the years 1995-2006 is contained in [CMI Working Paper 36](#). The reduced volume of new business in the investigation cast doubt over the credibility of the results and the CMI decided to cease collecting data in its current form and the impairment codes were removed in version 1.6 of the [Coding Guide](#), in January 2010.

[Working Paper 36](#) also sought views on possible options for a revamped future investigation into impaired lives, perhaps covering annuities and critical illness insurance as well as assurances.

Responses to the consultation from data contributors did not indicate that the CMI would be able to collect detailed data on impaired lives in the short term, beyond an indicator of whether or not a life was subject to special terms. For the foreseeable future, CMI analyses will be limited to 'standard business' and 'non-standard business' with no further analysis of the non-standard benefits, for example by type of impairment.

The CMI believes that an impaired lives' investigation has strategic importance to the insurance industry, in demonstrating the need to underwrite, to charge additional premiums for impaired lives and also to help better understand trends in non-impaired mortality and will continue to explore options to reinstate an investigation. Views and suggestions are welcome on the value and practicalities of reinstating an impaired lives investigation.

5. Results from Per Policy data: Areas for Consultation

This section sets out the specific areas on which the Committee wishes to seek views.

- Q1. Do you have any comments regarding the provision of two distinct types of results, “Summary Results” and “Detailed Results”?
- Q2. Do you have any comments on the proposed format of Summary Results? Comments are specifically invited on the following aspects:
- a. The proposed categories for Summary Results; in particular:
 - Are there any other categories that might be valuable that have not been referred to in Working Paper 56.
 - b. The choice of Product Categories; in particular:
 - The proposed hierarchal structure of Product Categories as set out in Working Paper 56, and
 - The proposed level of Product Categories contained in the Summary Results.
 - c. The medium for results (i.e. Excel).
- Q3. Do you have any comments on the proposed format of the Detailed Results set out in section 3.10? Comments are specifically invited on the following aspects:
- a. The record structure; in particular:
 - Whether All Office Detailed Results should initially be issued using Approach 2, and
 - Whether Individual Office results using Approach 1 would be valuable if Approach 2 were adopted for All Office results.
 - b. The content of records; in particular:
 - Which Attributes should be included in results if Approach 2 were adopted for the record structure, and
 - Other Indirect Attributes that might be valuable that have not been referred to in this paper.
 - c. The medium for results (i.e. text file).
- Q4. Do you have any comments on the value of the statistical tests in the current Life Office Mortality tests? Do you have suggestions regarding tests that would be of value in future? Should such tests be generated for Individual Office Summary Results, All Office Summary Results or only used in published analyses (in Working Papers)?
- Q5. Do you have any comments on the proposals regarding transmission of CMI results? Would you be willing and able to download results from a secure website in the future? Do your responses vary between Individual Office and All Office results? What encryption software do you have available for accessing results?

- Q6. Views are requested on the value of continuity between the Summary Results and the results traditionally issued for Life Office Mortality, in particular:
- a. For what period might it be useful to make available Summary Results comparable with the results issued for earlier years (by mapping to the current “investigations”) in order to allow experience to be tracked across years?
 - b. Is there value in such mapping where the mix of offices in the underlying data has changed considerably or is this only valuable if it can be undertaken on a reasonably consistent dataset?
- Q7. Do you have any comments on the proposed use of postcode, in particular:
- a. “Mapping” to Acorn and Mosaic categories, or
 - b. Including part of the postcode as an Attribute (if legally permissible)?
- Q8. Should the CMI consider the feasibility of undertaking lapse analyses? Please provide reasons for your answer.
- Q9. Do you have any comments on the value and practicalities of reinstating an impaired lives’ investigation?
- Q10. If there were sufficient demand, the CMI would look into the possibility of providing a data tool to analyse the Detailed Results. Would the provision of a data tool be of value to you? Do you have any further comments or suggestions for this? Would you be able to use Microsoft’s SQL Server Analysis Services OLAP cube?
- Q11. Only the current data requirements have been considered in this paper – are there any additional fields that you think data contributors would be able to provide that might usefully be added for future analyses?

Responses on the points noted above – and indeed any other comments arising from this Working Paper – should be sent via e-mail to mortality@cmib.org.uk or in writing to: Rachel Cox, CMI, Cheapside House, 138 Cheapside, London, EC2V 6BW.

Responses are requested by 30 November 2011.

References

CMI Working Paper 19 : Per-policy data submission (December 2005)

CMI Working Paper 36 : The mortality of impaired assured lives: Report on 1995-2006 experience and consultation on the future of the investigation (January 2009)

CMI Working Paper 56: CMI Life Office Mortality 'Per Policy' Results: the Initial Methodology and Format (September 2011)

Per Policy Coding Guide; version 1.6 (January 2010)

Per Policy Coding Guide; version 1.7 (September 2011)

All of the above can be found at: <http://www.actuaries.org.uk/research-and-resources/pages/continuous-mortality-investigation>

Appendix 1: Example file size for Detailed Results produced under Approaches 1 and 2

The example below compares the number of results records and resulting file sizes for a set of Detailed Results produced under Approaches 1 and 2. The example dataset contains 2 million data records for mortality assurance contracts issued at standard rates. This is comprised of 1 million males and 1 million females. The example is intended to be realistic and to represent a data submission from a large office, based initially on a single investigation year.

Approach 1

Individual Office

Assume that Approach 1 is applied to the Detailed Results and all data fields supplied by data contributors are included within the results records, except postcode. In addition to the data fields supplied by data contributors, the following additional fields are appended to the results records for this example: Product Category, age and duration (at the start of the year) and the exposed to risk, actual and expected claims for the four age and durations that could be attained during the investigation year (on both a lives and amounts basis).

Under this approach, the Detailed Results will contain 2 million records and, assuming that the results are provided in the form of a text file, will be around 800mb. By zipping the file the size will be reduced by over 90%, to approximately 50mb. A file of this size is too large for e-mail but will easily be accommodated on a CD; if a secure website were used then, at typical internet access speeds, the file could take around 7 minutes to download.

Note that, in order to further reduce the file size, the Detailed Results could be split, say, by gender and a separate results file produced and/or dispatched for each.

All Office

If Approach 1 is used for All Office results, then these numbers will escalate proportionally to the data volumes, e.g. if the All Office dataset is 10 times the size of the office in this example, then the All Office Detailed Results will be approximately 10 times the size of the Individual Office Detailed Results. (NB This ignores any “saving” that arises from the removal of sensitive fields, as discussed in section 3.) This would lead to the production of an 8gb text file which could be reduced to 500mb by zipping.

Although it is possible to put a text file of this size onto a CD, it could be very slow to access the results, depending on the application used. As noted above, the size of the file could be reduced further by splitting the Detailed Results, e.g. by gender. Using this example, the result would be the production of two 250mb text files.

Approach 2

Individual Office

Now assume that Approach 2 is applied to the Detailed Results and the following attributes are used:

Attribute	Number of categories
Territory	1 (UK data only)
Age (last)	Around 80 (Individual ages)
Curtate Duration (0,1,2,3,4,5+)	6 (Grouped duration)
Gender	2
Smoker Status	2
Product Category	7 (to level 2)
Medical Status	2
Original Type of Entry	1
Single Joint Life	2
ABI New Business Code	9
Distribution Channel	4

Under this approach, the Detailed Results will contain around 25,000 records and assuming that the results are provided in the form of a text file, this example data set would lead to the production of a 2.5mb file. Again, by zipping the file the size will be reduced by over 90%, leaving the zipped text file at approximately 140kb. (As noted above, splitting the Detailed Results by gender would lead to a further reduction in the file size but this is probably unnecessary under Approach 2.)

Note that, although the numbers of categories indicated above may appear to give rise to nearly 2 million distinct combinations of Attributes, the number of records is much lower. This arises because many combinations of Attributes do not contain any data and can be suppressed; for example, there may not be older lives at short durations. In addition, the Attributes are not distinct – in particular there is considerable correlation between the Product Categories and the ABI New Business Codes. Hence the actual number of results records is considerably lower than the potential number.

Using this data submission and the Attributes detailed in the table above, Approach 2 leads to a considerable reduction in the file size of the Detailed Results compared to Approach 1.

All Office

If Approach 2 is used for All Office results, then the number of records and the file size will increase as a result of data occurring for some additional combinations of Attributes. The increase however will be modest and far less substantial than that between the Individual Office and All Office Detailed Results produced under Approach 1.

Quadrennial results

The example above is for an individual year. If quadrennial results are provided with exposure and claims split by record year, then the file sizes under Approach 1 could increase four-fold (unless it is possible to “extend the records” across the calendar years, as discussed in section 3.3).

Under Approach 2, quadrennial results could contain four times the number of records; alternatively the exposure, expected and actual claims for the four years could be included in a single set of records, which would be little larger than the individual year.

Appendix 2: Per Policy data fields and possible Attributes

The table below sets out the data fields that are requested in Per Policy data (extracted from version 1.7 of the Per Policy Coding Guide). The table shows the following:

- **Field.** This column simply shows the name given to the data field in the coding guide.
- **Format of Values.** This column shows, for each data field, the possible values that are permitted. Note that the coding guide itself contains additional clarification for many of the fields.
- **Mandatory?** indicates whether the data field is mandatory or optional. Fields indicated as Y* for “Mandatory?” are not mandatory if the value is the same for the entire data submission and is clearly specified in accompanying documentation (e.g. if the file only contains UK business for office 999 relating to 2006). Additional explanation of the fields indicated as N* is provided below the table.
- **Use?** indicates how we envisage the data field could be used in Detailed Results (see section 3), in particular whether:
 - It can be included directly in the results records under Approach 1, which in some cases differs between Individual Office (IO) and All Office (AO) results;
 - It can be used as an Attribute under Approach 2; and/or
 - It is likely to be used indirectly to determine other Attributes.
- **Comments**

Field	Format of Values	Mandatory?	Use?	Comments
Record type	I = In force at the end of the record year O = Policy taken out of force in the record year	Y*	Direct for Approach 1 only	
Office Number	NNN	Y*	Direct for Approach 1 IO only	
Record Year	YYYY	Y*	Direct	Required if results cover multiple years
Territory	1 = UK 2 = Republic of Ireland	Y*	Direct	Would only be included if receive sufficient Irish data

Field	Format of Values	Mandatory?	Use?	Comments
Product code	Any alphanumeric (up to 10 characters ^s)	Y*	Direct for Approach 1 IO or Indirect	This field is unique to offices; it can only be used directly for IO results or indirectly to map to the product category
Client identifier	Any alphanumeric (up to 10 characters ^s)	N	Direct for Approach 1 IO only	These fields are included in the data requirements principally for use in communications with offices, but could be included in Individual Office results under Approach 1
Policy identifier	Any alphanumeric (up to 10 characters ^s)	Y		
Benefit identifier	Any alphanumeric (up to 10 characters ^s)	Y (if >1 benefit)		
Sex	M, F	Y	Direct	
Medical type code	M = Life medically examined on entry N = Life not medically examined on entry but satisfactory evidence of health received P = Lives accepted after paramedical examination S = Lives accepted on minimum evidence of health via a shortened proposal form. U = Unknown/Undifferentiated W = Sold without underwriting	Y	Direct	Use depends on data volumes and the proportion coded as "U". Also required in order to map records to some of the scheduled Life Office Mortality investigations
Smoker status	N = Non-smoker S = Smoker U = Unknown/Undifferentiated	Y	Direct	
Date of Birth	DDMMYYYY	Y	Direct for Approach 1 or Indirect	This could be used as a direct Attribute for Individual Office results or indirectly to determine age

Field	Format of Values	Mandatory?	Use?	Comments
Original Type of Entry	C = Compensation case G = Effected by exercising a GIO N = New Business O = Other U = Unknown	Y	Direct	Use will depend on data volumes and the proportion coded as "U"
Date of policy commencement	DDMMYYYY	Y	Direct for Approach 1 or Indirect	These could both be used as Direct Attributes for Individual Office results. The date of benefit commencement will also be used indirectly to determine duration and underwriting year (unless an office does not issue "benefits", in which case date of policy commencement will be used).
Date of benefit commencement	DDMMYYYY	Y*		
Entry into Current Status	A = Alteration (on) C = Compensation case G = Effected by exercising a GIO H = Annuity benefits suspended as death suspected I = In force at previous submission N = New business O = Other Q = Claim being investigated R = Reinstatement from lapse or suspension T = Bulk transfer-in U = Unknown W = Commencement of a dependant's pension annuity	Y	Direct for Approach 1 only	These are intended for use in data validation but could be used for Approach 1 if calendar years' data are not combined
Movement on date	DDMMYYYY	Y		

Field	Format of Values	Mandatory?	Use?	Comments
Benefit maturity/expiry date	DDMMYYYY	Y	Direct for Approach 1 or Indirect	This could be used to determine term
Business Type	H = Hybrid N = Non profit U = Unit linked W = With-profits	Y*	Direct	This could be used directly or may help to determine product category. It is required to map some records to the scheduled Life Office Mortality investigations
Premium frequency	P = Recurrent Single premium R = Regular premium S = Single premium	N	Direct	
Premiums in payment or paid up	N = Paid up Y = Premium paying	N	Direct	
Single or joint life	D = Dual J = Joint life first event benefit or joint life annuity S = Single life benefit	N	Direct	Also required to map some records to the scheduled Life Office Mortality investigations
Rated or non-rated	N = Non-rated Y = Rated	Y	Direct	

Field	Format of Values	Mandatory?	Use?	Comments
Benefit type	DB = Stand Alone Death benefit SC = Stand Alone Critical Illness benefit AC = Accelerated Critical Illness benefit DC = Stand Alone Death component of a multiple benefit Death and Critical Illness policy CA = Accelerated Critical Illness component of a multiple benefit Death and Critical Illness policy CC = Stand Alone Critical Illness component of a multiple benefit Death and Critical Illness policy LA = Life annuity in payment benefit DA = Pension benefits in deferment NA = Pension annuity in payment to members retiring in normal health IA = Pension annuity in payment to pensioners retiring in ill-health PA = Pension annuity in payment where the health status of the pensioner at retirement is unknown XA = Pension annuity where it is not known whether the beneficiary is the member or a dependant WA = Pension annuity in payment to dependants including widow(er)s	Y*	Direct	This could be used directly or to determine Product Category
ABI new business code	NNN	Y	Direct	This could be used directly or to determine Product Category
Distribution channel code	A = Basic advice (i.e. Stakeholder products) B = Bancassurance I = IFA/Whole of market M = Multi-tie/Limited range N = Non-intermediated S = Single tie U = Unknown	Y*	Direct	

Field	Format of Values	Mandatory?	Use?	Comments
Location	Any alphanumeric area postcode	N	Direct for Approach 1 or Indirect	This could be used as Direct Attribute for Individual Office results or indirectly (see section 4.1)
Initial benefit amount	NNNNNNNNN.NN	N	Direct for Approach 1 or Indirect	These could be used as Direct Attributes under Approach 1 but can also be used indirectly to assign to a benefit amount bands.
Benefit amount at 'Movement on date'	NNNNNNNNN.NN	Y		
Benefit amount at end of year or 'Date of exit'	NNNNNNNNN.NN	Y		
Date of amount review	DDMM	Y (if)	Direct for Approach 1 only	Intended to only be used to determine amounts exposure
Type of increment / decrement	C = RPI subject to a cap D = Decreasing (non-Mortgage) F = Fixed rate increase I = Family Income Benefit L = LPI M = Decreasing (Mortgage) N = No increment (i.e. level) O = Other R = RPI W = With-profits U = Unknown	Y*	Direct for Approach 1 or Indirect	Primarily intended for data validation but could be used directly as an Attribute or indirectly to identify Product Category
Rate of increment / decrement	NN.NN Rate of increase or decrease in benefit	N (Previously "Y*")	Direct for Approach 1	
Previous Investigation Number	NN	N	Direct	Used to map records to the scheduled Life Office Mortality investigations
The following three fields should only be completed for pension annuities in payment.				

Field	Format of Values	Mandatory?	Use?	Comments
Pension Grouping	B = Other bulk purchase annuities (i.e. where the office is unable to identify whether this is “buy-in” or “buy-out” business) C = Buy-out bulk purchase annuities D = Buy-in bulk purchase annuities I = Individual annuities U = Unknown	Y*	Direct or Indirect	Intended to only be used indirectly to identify Product Category but could be used directly as an Attribute
Pension Source Type	O = Occupational pension P = Private pension (unknown source) Q = Private pension (personal pension) R = Private pension (income drawdown) S = Private pension (S226) U = Unknown	Y*	Direct or Indirect	Intended to only be used indirectly to identify Product Categories but could be used directly as an Attribute
Dependant’s proportion	NN.NN	N ^y	Direct for Approach 1	Intended to only be used in data validation
The fields below should be completed only for benefits exiting during the year.				
Date of exit	DDMMYYYY	Y	Direct for Approach 1	Intended to only be used to determine exposure

Field	Format of Values	Mandatory?	Use?	Comments
Type of exit	A = Alteration (off) – see section 4.8 B = Cover ceases due to a claim on another benefit C = Critical Illness claim paid D = Death claim paid E = Ex-gratia claim paid H = Annuity benefits suspended as death suspected L = Lapse M = Maturity / Expiry Q = Claim being investigated S = Surrender T = Terminal Illness claim paid U = Unknown X = Other exit	Y	Direct or Indirect	Intended to be used to segregate claims from other exits and, indirectly, to separate death and TI claims (for assurances). Could also be used as a direct Attribute if the CMI were to undertake lapse analyses (see section 4.2)
Date of claim	DDMMYYYY	N*	Direct for Approach 1 only	Primarily intended to be used to determine exposure
Date of notification of claim	DDMMYYYY	N*	Direct for Approach 1 only	Primarily intended to be used to determine exposure if Date of Claim not available
Date of claim admission	DDMMYYYY	N*		
Date of claim settlement	DDMMYYYY	N*		
Cause of CI Claim	Any Alphanumeric	N	Direct	Critical Illness analyses only

Two fields have N* in the Mandatory? field:

- The dependant's proportion need only be recorded for joint life annuities (see section 5.44 of the coding guide); and
- The various Dates of Claim are shown as N* but at least one of these four dates **must** be supplied for assurances and at least one of date of death or date of notification **must** be supplied for annuities. The date of the claim event (death or diagnosis) is the preferred field for both assurances and annuities.