



Data submissions to the CMI Income Protection investigation

What is the scope of the investigation?

The Income Protection investigation covers individual income protection cover, focusing on policies that are not medically rated. It does not currently cover group income protection cover; this part of the investigation was discontinued following the withdrawal of a number of offices leading to unacceptably high concentrations of data between offices.

What data do we need to supply?

The CMI adopts a flexible approach to data collection, and keeps the number of essential fields to a minimum, thereby making data submission as easy as possible for offices whilst still providing valuable aggregated results.

Until recently, the Income Protection investigation's data processing and analysis system had restricted the data that could be captured and analysed. However, a system upgrade (see [Working Paper 149](#)), means it is now possible to capture and utilise a wider range of data fields. We are actively encouraging submission of more data fields than have historically been requested and, provided sufficient data volumes are collected, plan to expand the range of factors analysed.

This document outlines the key data fields we are seeking from insurers for the Income Protection investigation and identifies these as "essential", for the standard analyses, or "desirable" for additional analyses.

Details of the data that the CMI needs are contained in the Appendix to this note.

What will the CMI do with our data?

Initial checks will be applied to ensure that we understand the data you have supplied; this will be done as soon as possible and we will then let you know whether we are able to proceed. In some cases we will need to check our understanding of your data. This is more likely while we seek to ensure good understanding of data being used in the upgraded systems, but we will seek to keep the number of queries to a minimum by focusing only on issues material to the results.

When the processing of your data is complete, we will send you a "Data Contributor Report", summarising the assumptions and mappings that we have applied to your data, and office-specific results. You will be asked to review the Data Contributor Report and the results to ensure that we have not misinterpreted your data.

Once data collection is deemed complete, the CMI will produce aggregated ("all offices") results. These results will be available to all CMI subscribers; as will commentary on key features contained in a Working Paper.

The CMI may also make high-level results available more widely, for example via [blogs](#), articles in *The Actuary* magazine or presentations open to all members of the IFoA.

Who will have access to our data?

Data submitted to the CMI, in any form, is subject to the [Terms and Conditions for Data Contributors](#), and is processed in accordance with CMI's Data Handling Protocols which address both data protection considerations and protecting the confidentiality of individual contributors' data. These are available on request.

Contact Details

If you have any questions, please contact us on incomeprotection@cmilimited.co.uk or 020 7776 3820.

June 2022

Appendix: Data submission guidelines for income protection

Flexibility

The CMI is flexible in terms of the format in which data is provided. Clearly though, each office's data will have to be converted into a standard format before being combined with other offices' data; so the CMI Secretariat needs enough explanatory information regarding your data to enable us to format the data appropriately. An accompanying explanation of the data fields and approach used, and consistency from year to year, would be greatly appreciated.

If possible, we encourage you to submit data files that you use for other purposes, such as internal experience analyses; these should necessitate less work for you to submit data and should also mean that the results we send back to you are more consistent with your own results.

We suggest that offices consider the data they can submit and then contact the Secretariat at an early stage in their data preparation to agree whether the data meets the CMI's needs. In particular, it may be helpful if offices submit a small volume of "test" data to allow the Secretariat to provide an opportunity for any issues to be resolved in advance of preparing a full data submission.

Personal data

Data supplied to the CMI must not include name, address or postcode of any policyholder. Consequently, the vast majority of records within a CMI dataset could not be related to a particular individual and would not, therefore, constitute "personal data" under data protection legislation. However, a few records may fall within the legal definition and we therefore:

- i. Seek to minimise this likelihood by using rounded dates of birth and capping benefit amounts (see below) in the data we use; and
- ii. Treat all such data as if it were personal data from a legal and data security perspective.

Principles

The principal analyses are intended to cover:

- Individual business only (not group).
- Income protection policies only.
- UK business, but small volumes of non-domestic business may be included.
- Policies issued at normal premium rates; however:
 - The Committee may consider producing high-level results for rated lives – it would therefore prefer ALL data to be submitted, with a rated/non-rated indicator.
 - Policies subject to exclusions can be included in data submissions, or excluded; however offices are requested to apply this consistently – in particular between in force records and claims. Please advise which approach you use.
- Direct business, gross of reinsurance.

The CMI is seeking to analyse experience in two areas:

- Claim inceptions (i.e. the frequency of claims); and
- Claim terminations (i.e. the length of claims).

This is done by comparing claim events (inceptions and terminations) with exposure. To do this we require:

- Details of each policyholder under each distinct benefit;
- Details of when each policy went on-risk and off; and
- Additional details regarding claims; the definition of these is considered further below.

Our standard analyses seek to analyse the experience:

- by deferred period;
- by CMI occupation class, so require details of insurers' occupation classes and how these are mapped from occupations;
- by smoker status; and
- by year, so the data should allow us to calculate exposure and allocate claims, recoveries and deaths by calendar year.

The tables below indicate which fields are “essential”, for our standard analyses, and which fields are “desirable”, to be able to analyse experience by additional factors.

Of the desirable fields the following are of particular interest:

- socio-economic status, using IMD measures;
- amounts weighting, which requires a measure of current benefit amount;
- sum insured band, requiring a measure of initial benefit amount;
- cause of sickness;
- distribution channel; and
- policy expiry date.

Essential data items

Data field	Comments
Record identifier	To enable ready identification of a record, should a query arise. This should be consistent between in-force records and claim records and, preferably, between submissions
Gender	
Date of birth	See comments below
Record type	In force, exit or claim, dependent on the structure of data that is supplied
Deferred period	
Occupation class	Accompanied by details of insurers' occupation classes and how these are mapped from occupations
Smoker status	
Policy commencement date	
Date of exit	
Type of exit	
Rating indicator	Not relevant if data is supplied just for “standard rate” cases
For claims:	
Date of falling sick	
Payment commencement date	
Payment cessation date	
Mode of cessation	Recovery, death or other mode of cessation, such as expiry, lump sum payment, ex-gratia commutation, benefit rate altered, end of limited benefit period

Please talk to us if any essential data items cannot be supplied.

Desirable data items

The following additional data fields are particularly desirable:

Data field	Comments
Territory of the life insured at outset	Unless the data can be regarded as UK-only
Initial benefit amount	See comments below
Current benefit amount	See comments below
Policy expiry date	See comments below (alongside dates of birth)
Rating type/level	e.g. medical/lifestyle, per mille rating
Maximum benefit period indicator	
Length of maximum benefit period	In years
Benefit guarantee indicator	
Level of benefit guarantee	
Type of benefit	Level/increasing/decreasing/waiver/lump sum/other
Mortgage related indicator	Whether sold in connection with a mortgage
Employment status	Employed/self-employed/unknown
Sickness definition	Own occupation, any occupation, ADLs
Socio-economic indicators	Measure based on the Index of Multiple Deprivation (IMD) – a postcode mapping tool, along with supporting documentation, is available on the CMI Data page of the IFoA website
Distribution channel	Distinguishing between Bancassurance, IFA/Multi-tie, Single-tie, Direct sales and Unknown (unless the data can be regarded as being from a single channel, in which case please advise us of this alongside the data extract)
Cause of sickness	For claims. This should be an indicator of the condition triggering the claim. If possible, it would be helpful to distinguish COVID-19 claims.
Benefit rate percentage	For claims

Dates of birth

To reduce the likelihood of individuals being identifiable, we encourage firms to always use “16” as the value for “dd”. However we will accept and process data submissions containing:

- Exact values, but will replace “dd” with “16” for the purposes of our analyses.
- Dates with quarterly- or half-yearly rounding; if so, please advise us of the rounding used.

It is often the case that expiry dates are on birthdays, if this is the case, we encourage expiry dates to be treated in the same way as dates of birth.

Note: We do not think that the other dates would be well-known, so this should not make a person more identifiable. We therefore request the exact date for these.

Benefit amounts

As with exact date of birth, we recognise that very high amounts potentially increase the chance of certain people becoming identifiable. Consequently, for both the initial and the current amount, we encourage firms to provide a value of “999999” for any record with an annual benefit over £75,000 and we will then re-weight these values in future analyses. However we will accept and process data submissions containing exact values, but will replace this with the substitute value for the purposes of our analyses.

Claims

We are seeking valid claims, in accordance with the policy terms, due for payment in each investigation year. This will include claims that commence during the investigation year or continue from a year prior to the investigation year.

We do not intend making any adjustment for Incurred But Not Settled (IBNS) claims or Incurred But Not Reported (IBNR) claims.

We assume that, when data is submitted, the claims data is ‘complete’, for the following reasons:

- CMI claim inceptions analyses are by year of payment commencement, using a deferred period-specific year of exposure. See Working Paper 149 for further details. This means that claims in a particular investigation year relate to sicknesses that commenced d weeks’ (where d is the length of the deferred period) prior to payment commencement. It is expected that policyholders will have notified offices of sickness during the deferred period, particularly for longer deferred periods, and therefore late reporting/settlement is likely to be less of an issue than it may be for other insurance policies.
- Data is submitted at least three months after the end of the investigation year, allowing time for late reporting/settlement.

The following should not be submitted as claims:

- Claims arising during a free cover period
- Ex-gratia claims

Investigation years

The CMI intends to produce aggregated (“all offices”) results for each calendar year and for each quadrennium. As each investigation year’s claim inceptions exposure period is deferred period-specific, and straddles two calendar years, we will require data covering both the investigation year itself and the previous year. Please contact us to find out which years of data we are currently focusing on.

Structure of data files

The CMI is happy to accept data files in a number of structures. It is preferable for the format to provide detailed policy data, such as:

- “Policy history” files, including the complete history of a policy in a single record.
- “Movements” data, indicating when policies go on-risk and off-risk within each calendar year.

“Census” data, comprising in force data at the end of each 31 December (including the prior year) and details of claims, which is consistent with the format historically requested, can also be submitted but it would be very helpful if policy commencement/exit dates could be provided to allow us to accurately determine when policies were on risk during the exposure period.

Other approaches may also be acceptable but please discuss them with us first.

Methods of submitting data

We are happy to accept data from your secure file transfer site or you can use the Barnett Waddingham site – please email us on incomeprotection@cmilimited.co.uk to agree the more convenient approach – in the following file formats:

- Text/CSV – commas should separate each field or the file should be arranged in fixed width format.
- Spreadsheet.
- Database – we request that database formats are Microsoft Access. For offices using different database software, we suggest that data be converted into Text/CSV.

Alternatively, data can be submitted by email but this should ONLY be done if you are confident that the data submission does not include any personal data. Where email is used, the attaching file should, where possible, be compressed. We encourage data contributors to encrypt and password-protect all data files. Passwords should be provided separately (by telephone) from the data.

The Secretariat may be able to accept other commonly used form of electronic data transmission but please contact us beforehand to check.