



April 2018

## CMI Research data – our intentions in relation to the GDPR

### Background

Much of the CMI's research activity is based on policyholder data that we seek from insurance companies for the Annuities, Assurances and Income Protection investigations and pension scheme member data that we seek from actuarial consultancies for the SAPS investigation. We refer to this as "Research data".

We have reviewed our approach to Research data, ahead of the implementation of the General Data Protection Regulation (GDPR) in May 2018; this document sets out our intentions.

### Is CMI Research data covered by the GDPR?

Under the GDPR, "*Personal data*" means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person."

As we do not ask for the name, address or National Insurance number in Research data, the vast majority of records within a CMI dataset could not be related to a particular individual. However, a few may and we are therefore:

- 1 Amending the data we seek, to minimise this likelihood, and
- 2 Continuing to treat all Research data as if it were personal data from a legal and data security perspective.

### Lawful basis for processing personal data

Under the GDPR, data controllers need to:

- identify the lawful basis for processing personal data,
- document this, and
- advise individuals of this in their privacy notice.

The GDPR also states that: "*The processing of personal data for purposes other than those for which the personal data were initially collected should be allowed only where the processing is compatible with the purposes for which the personal data were initially collected. ... further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall not be considered to be incompatible with the initial purposes.*"

We have been advised that the CMI's Research data should fall within this definition; consequently it may not be **necessary** for data controllers to explicitly mention in their privacy notice that information may be used for research purposes. However – for transparency – it may be desirable to do so, if there is potential that the data being submitted to CMI could be considered personal data.

The following wording may be suitable for life insurers submitting policyholder data:

*We may also use your personal data in research into the mortality and morbidity experience of policyholders in general. This may include the provision of data, anonymised as far as possible, to a recognised external authority, such as the [Continuous Mortality Investigation](#) (CMI) (a subsidiary of the Institute and Faculty of Actuaries), which analyses experience on our behalf.*

The following wording may be suitable in the case of pension scheme members:

*The Scheme Actuary may also use your personal data in research which assists actuaries in providing actuarial advice to pension schemes – for example research into the mortality experience (life expectancy) of pension scheme members in general. This may include the provision of data, anonymised as far as possible, to a recognised external authority, such as the [Continuous Mortality Investigation](#) (CMI) (a subsidiary of the Institute and Faculty of Actuaries), which analyses mortality experience on our behalf.*

## CMI's status under the GDPR

Under the CMI's current Terms & Conditions for Data Contributors, to the extent that Research Data submitted to the CMI includes any personal data, the CMI will act as a data processor on behalf of the data contributor. We have been advised that – with some minor modifications to current practices – this position remains valid under the GDPR; we therefore intend to reflect this positioning in our updated Terms & Conditions.

Going forward, as a processor, the CMI will use such personal data only for the purposes of the insurer- or scheme-specific analysis carried out for that data contributor. Where the CMI undertakes further research and analysis to prepare industry-wide analysis, it will do that only with anonymised and aggregated data created in the course of the insurer- or scheme-specific research. The GDPR does not therefore apply to such further research activities.

## Data minimisation

We have reviewed a number of aspects of our data requirements. As an example, our discussion of the need for the exact date of birth (DoB) concluded that:

- a) In some cases, exact DoB does increase the chances of an individual becoming identifiable.
- b) With regard to accuracy:
  - Only collecting the month and year of birth is adequate; it is also easy for data contributors to implement.
  - Data with quarterly-or half-yearly rounding would still be acceptable.
  - Less precise dates – such as annual dates – unacceptably reduce the accuracy of the analyses.

Consequently, we are amending our data requirements for all investigations to encourage data contributors to submit only month and year of birth. We will continue to accept and process data submissions containing exact values, but will only retain and use month and year. If data contributors wish to round dates of birth to the nearest quarter- or half-year, we will also accept and process such data however we will not process any data that is rounded to annual dates.

We also thought it could be helpful to share our thinking on certain other aspects of the data requirements;

1. **Date of retirement.** As retirement often occurs on the 60<sup>th</sup> or 65<sup>th</sup> birthday, use of the exact date could inadvertently reveal the exact DoB. We will therefore treat this consistently with DoB for both Annuities and SAPS data.
2. **Other dates.** We also reviewed the need for exact date of death (DoD) and agreed that we should continue to request this; as the rounding of both DoB and DoD would make age at death less accurate. In addition, it was not obvious that using a rounded DoD would significantly impact on the potential identifiability of individuals. We do not think that other exact dates, such as date of exit, would be well-known, so they should not make a person more identifiable. We therefore see no need to change our current practice of seeking exact dates here.
3. **Very high benefit amounts.** As with exact DoB, we recognise that very high amounts potentially increase the chance of certain people becoming identifiable. In addition, we currently make limited use of the exact amount, as (to date) our analyses have grouped data into relatively wide amounts bands.

However there is no simple pragmatic approach, akin to collecting only the month of birth, as capping extreme amounts would reduce the weighting of this group in amounts-weighted analyses and the use of an average would be complicated as this could be insurer- or scheme-specific, vary with age and change over time.

For SAPS, we intend to encourage data contributors to provide a figure of £999,999 for any record with an annual amount over £100,000 and we will then re-weight these values in our analyses. If (some) data contributors are comfortable supplying the exact amount then we will treat that consistently with other large amount data in “all schemes” analyses.

We intend using the same value for Annuities, for consistency, and values of £1,000,000 for Assurances and £75,000pa for Income Protection.

4. **Very high ages.** Although there is an increased chance of identifiability at very high ages – say 105+ – due to the low numbers of lives at such ages in the population we intend to continue to request such data as this is an area of increasing relevance to longevity risk.

## Socio-economic indicators

CMI has been keen to expand the quality of its analyses by collecting some form of socio-economic indicator. Currently:

- Full postcode is an optional field in our data requirements for life insurers for this purpose, but so too are several alternatives and the lack of consistency between these has made it difficult to produce combined results.
- We do not seek any such indicator for SAPS. Previously we sought full postcode but we withdrew this when it became apparent that too few data contributors were submitting this.

We consider full postcode to be inappropriate under the GDPR and have sought views from a sample of data contributors on a proposal that the CMI should instead collect a measure based on the Index of Multiple Deprivation (IMD) for each record. We are pleased that this was widely-supported.

We therefore intend to make available a tool that will enable the value to be determined from an individual's postcode, via mapping to a census data area<sup>1</sup> for which IMD measures are calculated<sup>2</sup>. Our intention is that we will map to semi-deciles, which give greater flexibility than deciles but are sufficiently broad to avoid any risk that the IMD measure could be reverse-engineered to identify the individual.

## Next steps

This document sets out our current intentions.

We intend to issue revised Terms & Conditions, Data Handling Protocols and data requirements documents later this month. We will also issue an IMD mapping, and supporting documentation, as soon as we can.

---

<sup>1</sup> For example, lower-layer super output areas (LSOAs) in England and Wales.

<sup>2</sup> For example, the latest data for England is available here: <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015>. As the four constituent countries of the UK each produce their own IMDs using different methodologies, they are not directly comparable. We are currently considering using the method for standardising scores, based on the underlying elements for income and employment, published by [Abel, Barclay and Payne \(BMJ, 2016\)](#) in a CMI mapping tool to collect consistent data for the UK.