IAQS | INSTITUTE OF ACTUARIAL & QUANTITATIVE STUDIES

Class: M.Sc Sem 3

Subject : Actuarial Practice 1

Chapter: Unit 4 Chapter 1

Chapter Name: Handling Data

Table of contents

Personal data and data legislation

Big data

Data governance

Risks associated with the use of data

Operational data requirements

Data quality

Data issues for employee benefit schemes

Checks on data

Lack of ideal data

Industry wide data collection schemes

Risk classification and reduction of heterogeneity

Table of contents

Personal data and data legislation

Big data

Data governance

Risks associated with the use of data

Operational data requirements

Data quality

Data issues for employee benefit schemes

Checks on data

Lack of ideal data

Industry wide data collection schemes

Risk classification and reduction of heterogeneity

1 Personal Data and Data Legislation

Personal data

Organizations often accumulate large amounts of information relating to individuals as part of their ongoing operations. The increasing use of technology has now made it possible to **collect**, **store** and **use** very large amounts of information about individuals in ever more diverse ways.

Organizations have particular responsibilities when **acquiring** and **maintaining personal data**.

Personal data relates to information in respect of an individual where the individual can be **identified**, or where the data combined with other information could allow the individual to be identified.

Personal data may usually consist of: Name, address, email address, date of birth, marital status, occupation, etc.

Organizations have an **ethical responsibility** to deal responsibly with personal data.

In particular, they need to balance the **privacy of individuals** with the need of the organization to make fair and reasonable use of the personal data in their operations.

1 Personal Data and Data Legislation

Data protection legislation

Many countries have **data protection laws** to safeguard the rights of individuals with regard to how organizations can process and maintain personal data.

While the relevant regulations vary by jurisdiction, the objectives and expected behavior are often similar.

Examples of legislation that are broadly similar include the Data Protection Act in the UK, Personal Information Protection and Electronic Documents Act in Canada, and Personal Data (Privacy) Ordinance in Hong Kong.

However, not all countries have **equivalent data protection legislation**. For example, the USA has much less stringent personal data / privacy laws or regulations than the UK. Organizations need to take extra care where **data is being transferred between countries**, even if the purpose is valid.

An example of common data protection principles is contained in the Data Protection Act in the UK, which has eight principles that must be followed when processing personal data.

1 Personal Data and Data Legislation

Personal data must:

- be **processed** fairly and lawfully
- be obtained and processed for **specified purposes**
- be **adequate**, **relevant** and **not excessive** for the purposes concerned
- be **accurate** and, where necessary, kept up to date
- **not be kept longer** than necessary for the purposes concerned
- be processed in accordance with the **individual's rights** under the Act
- be processed **securely**
- not be **transferred to a country or territory** outside the European Economic Area unless that country or territory ensures an adequate level of protection.

1 Personal Data and Data Legislation

The consequences of **non-compliance** with the relevant data protection laws when processing personal data can be significant.

Laws with respect to **confidentiality** and **data protection** have become more stringent in the recent years, as the number of **breaches** and **compromises** on data protection have increased. This is largely due to an increase in cyber attacks.

In addition to prosecution and/or financial penalties, breaching data protection rules could lead to **adverse publicity** which can lead to significant reputational damage for an organization.

The ability to identify the individual to whom the information relates is crucial to the definition of personal data.

For **anonymous data** (i.e., where that individual cannot be identified) the obligations on an organization are often considerably less. For example, in the UK anonymous data does not constitute personal data and the duties and obligations of the Data Protection Act do not apply.

1 Personal Data and Data Legislation

In early August 2023, the Indian Parliament passed the **Digital Personal Data Protection Act (the "Act").** The Act applies only to personal data that is maintained in digital form and applies to the processing of personal data outside of India only if such processing is "in connection with an activity related to offering goods or services to data principals [(i.e., data subjects)] within the territory of India."

Highlights of the Bill

- The Bill will apply to the processing of digital personal data within India where such data is collected online, or collected offline and is digitized. It will also apply to such processing outside India, if it is for offering goods or services in India.

- Personal data may be processed only for a lawful purpose upon consent of an individual. Consent may not be required for specified legitimate uses such as voluntary sharing of data by the individual or processing by the State for permits, licenses, benefits, and services.

- Data fiduciaries will be obligated to maintain the accuracy of data, keep data secure, and delete data once its purpose has been met.

1 Personal Data and Data Legislation

- The Bill grants certain rights to individuals including the right to obtain information, seek correction and erasure, and grievance redressal.
- The central government may exempt government agencies from the application of provisions of the Bill in the interest of specified grounds such as security of the state, public order, and prevention of offences.

1 Personal Data and Data Legislation

Exclusivity means that one party is restricted from buying, selling or otherwise partnering with other parties than the one on the other side of the contract. Exclusivity is used in contracts to limit what one party can do, usually to the commercial advantage of the other party

Sensitive personal data

By **sensitive personal data**, we are referring to personal data, that is disclosed to a **third party** without the consent of the concerned party, that can cause him shame and public humiliation

Sensitive personal data is generally subject to much **stricter regulation** than ordinary personal data. For example, it may be the case that sensitive personal data can only be processed when one of the following conditions has been satisfied:

- The data subject has given explicit consent.
- It is required by law for employment purposes.
- It is needed in order to protect the vital interests of the individual or another person
- It is needed in connection with the administration of justice or legal proceedings.

1 Personal Data and Data Legislation

Sensitive personal data can include information related to:

- racial or ethnic origin
- political opinions
- religious or other similar beliefs
- membership of trade unions
- physical or mental health condition
- sexual life
- convictions, proceedings and criminal acts.

Table of contents

Personal data and data legislation

Big data

Data governance

Risks associated with the use of data

Operational data requirements

Data quality

Data issues for employee benefit schemes

Checks on data

Lack of ideal data

Industry wide data collection schemes

Risk classification and reduction of heterogeneity

2 Big Data

Big data refers to the large, diverse sets of information that grows at ever-increasing rates. It encompasses the volume of information, the velocity or speed at which it is created and collected, and the variety or scope of the data points being covered. Big data often comes from data mining and arrives in multiple formats.

The increased use of technology has now made it possible for the public and private sector to **collect and analyse** very large data sets of information.

Big data can be characterized by:

- very **large data sets**
- data brought together from **different sources**
- data which can be **analyzed** very quickly – such as in real time.

2 Big Data

Data protection considerations for big data

If personal data is held by a company, then the company needs to comply with the relevant data protection rules. Given the large amount of information that could be held on an individual, **privacy considerations** are likely to be a concern for individuals whose data is held.

Anonymization can potentially aid big data analytics, as it means that the information being analyzed is no longer considered personal data.

Data anonymization is the process of protecting private or sensitive information by erasing or encrypting identifiers that connect an individual to stored data.
For example, you can run Personally Identifiable Information (PII) such as names, social security numbers and addresses through a data anonymization process that retains the data but keeps the source anonymous

Anonymization can assist organizations to carry on research or develop products and services. It also enables these organizations to give an assurance to the people whose data was collected that the organization is not using data that identifies them for big data analytics.

2 Big Data

A key **feature** of big data is using '**all**' the data, which contrasts with the **concept of data minimization** in the data protection principles. This raises questions about whether big data is **excessive**, while the variety of data sources often used in big data analytics may also prompt questions over whether the personal information being used is relevant.

Organizations need to be clear from the outset what they **expect to learn** or be able to achieve by processing the data, as well as satisfying themselves that the data is relevant and not excessive.

Organizations that hold **big data** also need to be **transparent** when they collect data, and explaining how the data will be used is an important element in complying with data protection principles. The complexity of **big data analytics** will not be an acceptable excuse for failing to obtain consent where it is required.

Regulators expect organizations that hold big data to be proactive in considering any **information security risks** posed by big data.

Data governance is becoming increasingly important for holders of big data. This must take account of data protection and privacy issues.

3 Data Governance

- **Definition**

Data governance is the term used to describe the overall management of the availability, usability, integrity and security of data employed in an organization.

- **Data Governance Policy**

A data governance policy is a documented set of guidelines for ensuring the proper management of an organization's data.

A data governance policy will set out guidelines with regards to:

- the specific roles and responsibilities of individuals in the organization with regards to data
- how an organization will capture, analyze and process data
- issues with respect to data security and privacy
- the controls that will be put in place to ensure that the required data standards are applied
- how the adequacy of the controls will be monitored on an ongoing basis with respect to data usability, accessibility, integrity and security.

The data governance policy will also provide a mechanism for ensuring that the relevant legal and regulatory requirements in relation to data management are met by the organization.

3 Data Governance

- **Data Governance Risks**

Organizations that do not have adequate data governance procedures can be exposed to risks relating to:

- legal and regulatory non-compliance
- inability to rely on data for decision making
- reputational issues
- incurring additional costs (for example fines and legal costs).

A sound data governance policy should therefore provide the organization's stakeholders (staff, management, regulator, shareholders and policyholders, amongst others) with confidence that the organization is dealing appropriately with the data it holds.

3 Data Governance

- **Mergers and Acquisitions**

Where businesses are combined by merger or takeover, one of the key issues is whether the data for the two businesses should be combined onto one system and, if so, which.

The saving in **overhead costs** such as system maintenance and management is frequently cited as a justification for the transaction. In practice, the costs of converting the data from one working system to another are high.

New developments are carried out on one system and the other is left to decline as a **legacy system**, often requiring proportionately higher maintenance costs. Thus, the aim of cost saving is often not achieved.

There is a **risk in aggregating data** sourced from difference systems and a data governance policy needs to address this risk.

Two companies will usually have different policies for data governance. One might have a stringent approach while the other may have a lenient one towards data governance. Decision needs to be taken to adopt either one of the policies or a combination of both to make use of best of both

Table of contents

Personal data and data legislation

Big data

Data governance

Risks associated with the use of data

Operational data requirements

Data quality

Data issues for employee benefit schemes

Checks on data

Lack of ideal data

Industry wide data collection schemes

Risk classification and reduction of heterogeneity

4 Risk Associated with the Use of Data

Data Risk

An actuary is faced with a range of possible risks when using data. Examples of possible risks are

- **The available data might contain errors or omissions**, which could lead to erroneous results or conclusions. For example, if we are expected to find the average marks achieved by 50 students in math, where we are given their respective marks. An error where a few marks are written in negative, may cause the class average to fall and will give us erroneous results.. Hence data needs to be checked for errors

- There may be **insufficient historical data** available to estimate credibly the extent of a risk, and the likelihood of the occurrence of that risk in future.

- Even where there is sufficient data to estimate credibly future experience in normal conditions, there may be **insufficient data available to provide a credible estimate** of a risk in very adverse circumstances, which may be necessary for some purposes (eg estimating the tails of a distribution). For example there may be insufficient data regarding a global recession, as the number of such events is less, hence predicting future experience based on such limited data, makes us question their credibility.

4 Risk Associated with the Use of Data

- Where there is insufficient data, it may be possible to **use data from other sources** (eg industry data, other countries, competitors), but there is a risk that data from other sources may not be a sufficiently good proxy for the risk being assessed.

- **Historical data may not be a good reflection of future experience**. This could be due to:

    - past abnormal events which may have an almost nil chance of recurrence in future
    - significant random fluctuations
    - future trends not being reflected sufficiently in past data
    - changes in the way in which past data was recorded
    - changes in the balance of any homogeneous groups underlying the data
    - heterogeneity with the group to which the assumptions are to relate for example smokers and non-smokers existing in the same group for a life insurance policy
    - the past data may not be sufficiently up to date
    - other changes – e.g., medical changes, social changes, economic changes etc.

4 Risk Associated with the Use of Data

- There are risks where an actuary attempts to group data into broadly **homogenous groups**. The risks associated with this are:
    - the individual data groups may be too small for a credible analysis
    - if data groups are merged so there is sufficient data in each group to be credible, the combined data set may not be sufficiently homogeneous.
- The available data may not be in a **form** that is appropriate for the purpose required.
- The available data may have been **collected for a purpose**, which means that it is not appropriate for a different purpose.
- A **lack of confidence** in the available data will reduce the confidence in an actuary's conclusions.

5 Operational Data Requirements

Main uses of data

Actuaries use data in all their work. The **interaction** between the **data requirements** for the various tasks that actuaries carry out can be complex and will vary from organization to organization. Essentially, however, for a given type of work the underlying data requirements will normally be similar.

The overriding principle is that the **data for all the tasks** should be controlled through one **single, integrated data system**. This helps with:

- Limits the possibility of data being **corrupted**
- Reduces **inconsistency** in data, between products or over a period of time
- The person in charge of entering and amending the information can be regulated
- When data is stored in a single system, its **accessibility** improves, as opposed to when it is stored in several systems, and one has to collate it from them
- It saves time as **reconciling data** from different systems will not be needed

5 Operational Data Requirements

However, this ideal is not always achieved in practice. In a smaller organization it is easier to ensure that the data used for different applications are consistent, because it is likely that the same small group of people will carry out the applications. Whereas in large institutions we have different teams having several members having access to the data , using it, and amending it.

Sources of data

- Publicly Available Data

For some purposes, data may only be required on a 'big picture' basis. Here, data will be publicly available from published company accounts and regulatory returns.

- Internal Data

Product providers need data relating to the individual risks that they provide cover for. The quantity and quality of these data are both important. Without sufficient quantity, data groupings will either be non-homogeneous or lack credibility. However, even where there are plenty of data available, poor-quality data will mean that any results produced are not reliable.

Table of contents

Personal data and data legislation

Big data

Data governance

Risks associated with the use of data

Operational data requirements

Data quality

Data issues for employee benefit schemes

Checks on data

Lack of ideal data

Industry wide data collection schemes

Risk classification and reduction of heterogeneity

6 Data Quality

Sources of data quality issues

Problems of data quality and quantity can be a result of:

- **poor management control** of data recording or its verification processes, such as wrongly spelled customer names or incomplete or obsolete information
- **poor design of the data systems** where the system may not be robust enough to handle large quantities of data, or it may be obsolete

This may not necessarily be a reflection on the current management, as good quality data cannot necessarily be obtained quickly. After implementing a process for maintaining extensive records, it may take many years for enough data to be collected for analysis purposes.

The availability of data of **good quality** and **quantity** will vary greatly between organizations and, within organizations, between the different classes of business.

6 Data Quality

Ensuring good quality data – the proposal form

When placing a value on liabilities, for healthcare, life and general insurers, the **prime information source** will be the details given on the **proposal form**. It is therefore important that it produces relevant and reliable information for the system, as this will eventually determine the level of cover and the policy cost to the policyholder

Questions need to be **well-designed** and **unambiguous**, so that the proposer will give the full, correct information and the underwriting department can process the application readily, adding any coding that is necessary.

6 Data Quality

Ensuring good quality data – the proposal form

Quantitative information is preferred over qualitative information and must be used where possible.

The information should be input in the computer system in a manner which is **consistent** with the proposal form, such as questions should be in the same order, so that its interpretation is easy and not tie consuming.

Input of the information and its validation must happen simultaneously by the system

In particular, the result of any medical or occupational underwriting will need to be added. For general insurance personal lines, the composition of the final premium from various rating factors will be important.

Rating factors are the characteristics of individual policyholders that will determine the cost of the risk.

6 Data Quality

Using Proposal Form Information in Claim Assessment

The information from the **proposal form** (together with any subsequent changes) will need to be held for several purposes, including **cross-checking against the claims information** at the time of any claim.

Any changes made to the policy are termed as **endorsements**. For example, if a policyholder extends his medical coverage to his spouse midway through the policy year, then this change needs to be updated in the data system

All policy details will need to be **retained** irrespective of the endorsements, as they are needed at the time of **assessing claims experience**. The policy's exposure to risk needs to be known throughout the year.

Holding the basic policy information should enable the **automatic checking** of the validity of the claim and the updating of the policy information (eg termination of cover in the event of a total loss under a general insurance policy or death under a life insurance policy).

The data requirements will depend on the type of benefits provided.

6 Data Quality

Ensuring good quality data – the claim form

Another important information source will be the details given on a **claim form**. Like the proposal form, it is important that this is designed with the aim of producing information that can be both **analyzed accurately** and also **transferred** easily to the computer system.

A claim form is a formal written request to an insurance company for compensation you believe you are entitled to under their rules or statutes

Data to be captured

As well as data relating to current risks covered, it is important to retain the **history of past policy** and **claim records**. Past records are key to **predict future scenarios**.

Data requirements would include details of the risks covered, details of the cover, details of claim (if recorded), status of present record, control dates (start, end dates of policy), etc

7 Data Issues for Employee Benefit Schemes

Introduction

Data for employee benefit schemes is usually provided by the sponsor of the scheme i.e., the employer. There may be occasions when the **actuary does not have full control** over the data available.

For example, when valuing benefits under an employee benefit scheme, the scheme sponsor will usually provide data on the operation of the scheme and the scheme membership.

Emphasis needs to be placed on the **quality of data**. The sponsor should be educated on its importance. The data must be complete, accurate, consistent with the previous data, etc.

Once the data has been provided by the sponsor, it needs to be validated and needs to be checked for errors.

7 Data Issues for Employee Benefit Schemes

Data requirements

Data will be required to place a **value on the benefit entitlements** of individuals. Data will be required in respect of:

- Individuals who have an entitlement to receive a benefit in the future. This would include information with respect to **pensioners**, **active members and deferred members**
- The data will need to be **sufficiently detailed** to provide all information that is likely to be financially significant to the level or timing of future benefits.
- Also, if the pension amount is with respect to the salary earned in the last few years, then detailed information of it needs to be provided.
- However, if a pension were also to be paid to a spouse after the death of the member, the existence and age of a spouse of a young member may not be financially significant as the marital status of the member may change in the future.

Table of contents

Personal data and data legislation

Big data

Data governance

Risks associated with the use of data

Operational data requirements

Data quality

Data issues for employee benefit schemes

Checks on data

Lack of ideal data

Industry wide data collection schemes

Risk classification and reduction of heterogeneity

8 Checks on Data

Before being put to further or used for further inference, the data needs to be **checked** thoroughly, or else it may affect the results it gives.

Verifying current data

Any **equivalent data** used when previously valuing benefits will be useful to the actuary as it will enable **reconciliations** to be performed that help to indicate the validity of the current data.

For example, for a benefit scheme the actuary will examine:

- The membership movements over the inter-valuation period – the number of members existing in the pension, plus any new entrants less those who have exited the scheme should match with those at the end of the reconciliation
- Changes in averages over the inter-valuation period – to have an idea of the changes in average salary, change in average retirement age etc.
- Individual records – some employees will result in larger pension liabilities such as senior managers or department heads etc.

8 Checks on Data

Use of accounting data

Where reserves are built up for benefits, a **balance sheet** and **income and expenditure statement** may exist.

This will provide information about the **total value of the assets** held and perhaps information relating to recent benefit outgo and premium / contribution income. This information will be useful in verifying other data or in considering the assumptions to be used.

If **audited accounts** exist, they will enable greater reliance to be placed on the figures when verifying the data. For instance, the accounts will show the company's financial position at the start and end of the accounting period. The asset value depicted by the balance sheet can be cross-checked against the asset data.

8 Checks on Data

Asset data

To place a value on assets that is reliable and consistent with a value placed on future benefits, it is necessary to obtain a **full listing of the individual assets** held.

These individual holdings should then be checked to determine whether they are permitted or are subject to valuation restrictions imposed by regulation or legislation.

8 Checks on Data

Assertions to be examined

Whether using data provided by their own organization or a third party, an actuary will have to **make and check certain assertions** about that data. Such assertions include:

- that a **liability** or **asset** exists on a given date
- that a liability is held, or an asset is **owned** on a given date
- that when an event is recorded, the time of the event and the associated income or expenditure are allocated to the **correct accounting period**
- that data is **complete**, ie there are no unrecorded liabilities, assets or events
- that the **appropriate value** of an asset or liability has been recorded.

8 Checks on Data

Checking the Assertions

A decision will then have to be made as to how these **assertions will be checked** and the level of detail that will be appropriate in checking them. Possible checks could include:

- Reconciliation of the total number of members/policies and **changes** in membership/policies, using **previous data and movement data.**
- Reconciliation of the **total benefit amounts** and **premiums** and changes in them, using previous data and movement data.
- The **movement data** should be checked against any appropriate accounting data, especially with regard to benefit payments.
- Checks should be made for any **unusual values**, such as impossible dates of birth, retirement ages or start dates.

8 Checks on Data

- Consistency between **salary-related contributions** and **in-payment benefit** levels indicated by membership data and the corresponding figures in the accounts.
- Consistency between the **average sum assured** or **premium** for each class of business should be sensible, and consistent with the figure for the previous investigation.
- Consistency between **investment income** implied by the asset data and the corresponding totals in the accounts.
- Where assets are held by a third party, reconciliation between the **beneficial owner's** and the **custodian's records.**
- Full **deed audit** for certain assets, such as checking the title deeds to large real property assets.
- Consistency between **shareholdings** at the start and end of the period, adjusted for sales and purchases, and also bonus issues, etc.
- **Random spot checks** on data for individual members / policies or assets.

9 Lack of Ideal Data

When ideal data is unavailable

In the best-case scenario, an **ideal data** would be one that is 100% reliable in terms of quality and quantity. The main circumstances where ideal data are not available are because:

- The amount of data is **insufficient**, say there are not enough data points or numbers in the sample data
- Lack of **detailing in the data**, making it inappropriate for the intended purpose.

Insufficient volume of data

There may be **insufficient data** to provide a **credible result**. A provider may have recently launched a new product or branched out into a new target market.

Alternatively, the provider may simply be too small to attach any credibility to its own experience. This is particularly the case with benefit schemes, where very few employers will be of sufficient size to have credible experience to assess mortality rates before retirement.

9 Lack of Ideal Data

The use of summarized data

- When valuing benefits, it may be appropriate to use **summarized data** instead of **detailed membership data** in some circumstances.
- For instance. when one company is acquiring another company, it may not have access to all the data regarding the acquired company before acquisition. In this case a summarized data will be useful.
- However, it should be recognized that the **reliability of the values** will be reduced, as **full validation** of the data will be impossible.
- Additionally, the summarized data may miss **significant differences** between the nature of benefits that have been grouped together. In the case data may also not be sufficient such that it can be split into homogeneous groups, with each group having enough data.
- It is also unlikely that summarized data could be used to **value options or guarantees** that may or may not apply on an individual basis.

9 Lack of Ideal Data

- For example, consider a benefit scheme that promises the larger of the two calculations. To establish the benefit amount with accuracy, individual calculations will be needed to be carried out for each beneficiary. Otherwise, a beneficiary who is to receive a larger benefit amount will be grouped with members for whom the other calculation is larger, thus under-estimating the cost of a guarantee.

- Summarized data is therefore only suitable if such **inaccuracy** is recognized by the users of the results of the calculations.

- As for the actuary's work, he is obliged to mention any **discrepancies** in the data, has to mention the **quality of data** on which the advice is based upon and must also disclose the source of data

Table of contents

Personal data and data legislation

Big data

Data governance

Risks associated with the use of data

Operational data requirements

Data quality

Data issues for employee benefit schemes

Checks on data

Lack of ideal data

Industry wide data collection schemes

Risk classification and reduction of heterogeneity

10 Industry-wide Data Collection Schemes

What are industry-wide data collection schemes

- In some countries there are organizations that **collect data** from their member offices and then make available **summaries of all the data** to their members. For example, in the UK, the Association of British Insurers collects and collates a wide variety of insurance data.

- This cannot be used in place of policy data to establish **provisions** for a particular policy or scheme but could be used to **determine bases or be used in product pricing**.

- One of the best examples is the Continuous Mortality Investigation Bureau of the Institute and Faculty of Actuaries in the UK, which does a large amount of work on **mortality and morbidity statistics**. The forum also collates and summarizes the data, pointing out the impact of any events such as pandemic or medical advancements on it.

- The **volume of data** that can be collected from across a whole industry greatly improves the statistical significance of the resulting analysis.

10 Industry-wide Data Collection Schemes

Potential Benefits from Using Industry-wide Data Collection Schemes

- An insurer participating in an industry-wide scheme has the prospect of being able to **compare** its **own experience** with that of the **industry** as a whole (or that part of it represented by the participating insurers) with regard to both the **overall level** and the **pattern** of the experience by the categories into which the data are classified. Any significant differences point to a need for explanation.

- Since an insurer is likely to be seeking to expand by **attracting business** from its **competitors**, it may be important to have an indication of the ways in which the characteristics of the business it is seeking may differ from those of the business it already has.

10 Industry-wide Data Collection Schemes

Possible reasons for heterogeneity

When using industry-wide data, there is potential for **distortions arising from heterogeneity**. This is because the data supplied by different organizations may not be precisely comparable because:

- companies operate in **different geographical or socio-economic sections** of the market. For example, an insurance company based in a developed city will have policyholders with better mortality rates and larger coverage than one based in a small town.

- the policies sold by different companies are **not identical**. For example, the products offered by companies may differ in terms of benefit where the benefit may be a fixed lumpsum or a with profit

- **sales methods are not identical**. Some companies may sell their policies via brokers, and some may directly sell to their customers via an online portal

10 Industry-wide Data Collection Schemes

Possible reasons for heterogeneity

- the companies will have **different practices**, eg underwriting or claim settlement standards. Some companies may undertake rigorous underwriting where detailed medical information is required while others may have a lenient approach requiring only the basics
- the **nature of the data** stored by different companies will not always be the same
- the **coding** used for the risk factors may vary from organization to organization.

10 Industry-wide Data Collection Schemes

Other problems with industry-wide data

Other problems with using industry-wide data may be:

- the data will usually be **less detailed**, or **less flexible**, than those available internally. The data may be in terms of a broad perspective instead of focusing on accurate details
- external data are often much more **out-of-date** than internal data
- the data quality will depend on the **quality of the data systems** of all of its contributors. Any errors in data collection made by one company such as assigning the wrong genders, may invalidate the entire data set.
- not all **organizations contribute**, and the organizations that do contribute are **not representative** of the market as a whole.

10 Industry-wide Data Collection Schemes

Other sources of data

- It may also be possible to obtain data from a **reinsurer**.
- An insurer may cater to a particular class or group of people offering specific products, whereas a **reinsurer** has a **wider scope** as it offers cover to several insurers in the market who sell different products.
- As a result, they have **access to wider data**, which can be useful to the insurer especially when launching a new product in the same market or entering a new market.
- Another source of data would be **national statistics**, where government bodies give information about demographics and any changes in them

11 Risk Classification and Reduction of Heterogeneity

What is the aim of risk classification

- When carrying out **data analysis**, having **sufficient data** is a prerequisite, so as to ensure the credibility of results. However, care must be taken that balance between the importance of having **sufficient data** and the **issue of heterogeneity** is maintained.

- The main aim of **risk classification** is to obtain **homogeneous data**.

- The reduction of heterogeneity within the data for a group of risks makes the experience in each group **more stable** and **characteristic** of that group. Furthermore, it enables the data to be used more appropriately for projection purposes.

- This is important when **monitoring claims** and **mortality experience**. Any **heterogeneity** in data groups will serve to **distort the results** and can lead to setting provisions that are too big or too small and calculating premiums or contributions that are incorrect.

11 Risk Classification and Reduction of Heterogeneity

- Ideally data to be analyzed should be split into **homogeneous groups**, for example, by age and gender in a mortality investigation.

- However, where data is **scarce**, such as for numbers of deaths at young ages, splitting data into homogenous groups may result in **data groups that are too small** to enable any credible analysis to be carried out.

- In such cases data may need to be combined into **groups** which are **less homogeneous**, but which are large enough to be credible. Whenever data is to be analyzed there needs to be a balance between splitting the data into homogeneous groups and having sufficient data in each group to enable a credible analysis to be carried out.

- There is also a need to carry out **sensitivity testing** to check that if the data are grouped in a different way, the same results are obtained.

Topics covered

- Personal data and data legislation
- Big data
- Data governance
- Risks associated with the use of data
- Operational data requirements
- Data quality
- Data issues for employee benefit schemes
- Checks on data
- Lack of ideal data
- Industry wide data collection schemes
- Risk classification and reduction of heterogeneity