1

Introduction to life and long-term health insurance

1.1 Summary

In this chapter we lay out the context for the mathematics of later chapters, by describing some of the background to modern actuarial practice, as it pertains to long-term, life contingent payments. We describe the major types of life insurance products that are sold in developed insurance markets, and discuss how these products have evolved over the recent past. We also consider long-term insurance that is dependent on the health status of the insured life, rather than simply survival or death. Finally, we describe some common pension designs.

We give examples of the actuarial questions arising from the risk management of these contracts. How to answer such questions, and solve the resulting problems, is the subject of the following chapters.

1.2 Background

The first actuaries were employed by life insurance companies in the early eighteenth century to provide a scientific basis for managing the companies' assets and liabilities. The liabilities depended on the number of deaths occurring amongst the insured lives each year. The modelling of mortality became a topic of both commercial and general scientific interest, and it attracted many significant scientists and mathematicians to actuarial problems, with the result that much of the early work in the field of probability was closely connected with the development of solutions to actuarial problems.

The earliest life insurance policies were annual contracts; the purchaser would pay an amount, called the **premium**, to the insurer, nominating an individual whose life was insured under the contract. The insured life might be the purchaser, but could also be a third party. If the insured life died during the year that the contract was in force, the insurer would pay a predetermined lump sum, the **sum insured**, to the policyholder or his or her estate. Each

year the premium would increase as the probability of death increased. If the insured life became very ill at the renewal date, the insurance might not be renewed, in which case no benefit would be paid on the insured life's subsequent death. Over a large number of contracts, the premium income each year should approximately match the claims outgo. This method of matching income and outgo annually, with no attempt to smooth or balance the premiums over the years, is called **assessmentism**. This method is still used for group life insurance, where an employer purchases life insurance cover for its employees on a year-to-year basis.

The radical development in the later eighteenth century was the level premium contract. The problem with assessmentism was that the annual increases in premiums discouraged policyholders from renewing their contracts. The level premium policy offered the policyholder the option to lock in a regular premium, payable perhaps weekly, monthly, or annually, which was fixed for the term of the contract. This was much more popular with policyholders, as they would not be priced out of the insurance contract just when it might be most needed. For the insurer, the attraction of the longer contract was a greater likelihood of the policyholder paying premiums for a longer period. However, a problem for the insurer was that the longer contracts were more complex to model, and offered more financial risk. For these contracts actuarial techniques had to develop beyond the year-to-year modelling of mortality probabilities. In particular, it became necessary to incorporate financial considerations into the modelling of income and outgo. Over a one-year contract, the time value of money is not a critical aspect. Over, say, a 30-year contract, it becomes a very important part of the modelling and management of risk.

Another development in life insurance in the nineteenth century was the concept of **insurable interest**. This was a requirement in law that the person contracting to pay the life insurance premiums should face a financial loss on the death of the insured life – an insurance payout should not leave the beneficiary financially better off than if the insured life had not died. The insurable interest requirement ended the practice where individuals would insure persons (often public figures) with no connection to the purchaser, as a form of gambling. It also, importantly, removed the incentive for a policyholder to hasten the death of the insured life. Subsequently, insurance policies tended to be purchased by the insured life, and in the rest of this book we use the convention that the policyholder who pays the premiums is also the life insured, whose survival or death triggers the payment of the sum insured under the conditions of the contract.

The earliest studies of mortality include life tables constructed by John Graunt and Edmund Halley. A life table summarizes a survival model by

specifying the proportion of lives that are expected to survive to each age. Using London mortality data from the early seventeenth century, Graunt proposed, for example, that each new life had a probability of 40% of surviving to age 16, and a probability of 1% of surviving to age 76. Edmund Halley, famous for his astronomical calculations, used mortality data from the city of Breslau in the late seventeenth century as the basis for his life table, which, like Graunt's, was constructed by proposing the average ('medium' in Halley's phrase) proportion of survivors to each age from an arbitrary number of births. Halley took the work two steps further. First, he used the table to draw inference about the conditional survival probabilities at intermediate ages. That is, given the probability that a newborn life survives to each subsequent age, it is possible to infer the probability that a life aged, say, 20, will survive to each subsequent age, using the condition that a life aged zero survives to age 20. The second major innovation was that Halley combined the mortality data with an assumption about interest rates to find the value of a whole life annuity at different ages. A whole life annuity is a contract paying a level sum at regular intervals while the named life (the annuitant) is still alive. The calculations in Halley's paper bear a remarkable similarity to some of the work still used by actuaries in pensions and life insurance.

This book continues in the tradition of combining models of mortality with models in finance to develop a framework for pricing and risk management of long-term policies in life and health insurance. Many of the same techniques are relevant also in pensions mathematics. However, there have been many changes since the first long-term policies of the late eighteenth century.

We usually use the term **insurance** to refer to a contract under which the benefit is paid as a single lump sum, either on the death of the policyholder or on survival to a predetermined **maturity date**. (In the UK it is common to use the term **assurance** for insurance contracts involving lives, and 'insurance' for contracts involving property.) An **annuity** is a benefit in the form of a regular series of payments, usually conditional on the survival of the policyholder.

1.3 Traditional life insurance contracts

1.3.1 Introduction

The three traditional forms of life insurance are term, whole life and endowment insurance. These policies dominated insurance markets until the 1980s, and in some countries are still popular today. However, the design of life insurance has broadened significantly in the past few decades. In this section we describe features of traditional life insurance policies. In the next section we review more modern developments.

1.3.2 Term insurance

Term (or temporary) insurance pays a lump sum benefit on the death of the policyholder, provided death occurs before the end of a specified term. Typical contract terms range from 10 to 30 years.

The premiums for term insurance are usually very small relative to the sum insured, because the insurer has to pay a death benefit on only a small proportion of the policies issued. If a policyholder who is aged 40 purchases 10-year term insurance, then the probability that the insurer will pay any death benefit on the policy (which is just the probability that the life dies before age 50) might be around 2%. So, around 98% of such policies will expire with no death benefit payable, and the premiums from these policies subsidize the benefits for the 2% for which the death benefit must be paid.

A term insurance policyholder may choose to lapse their policy, which means the policyholder would cease paying premiums. In this case, their insurance cover would also cease, and there would be no further payments by the policyholder or the insurer.

The main purpose of term insurance is for family protection. For a relatively low monthly cost, term insurance can protect the policyholder's spouse and children against financial hardship in the event of the policyholder's death.

Another use of term insurance is to protect businesses against losses arising from the deaths of key employees. In this case the business pays the premiums and receives the sum insured if the insured life dies during the term. The business must demonstrate insurable interest. This type of insurance used to be called **Key Person Insurance**, but is now more commonly called **COLI** for Company Owned Life Insurance.

Most term insurance policies offer a level sum insured, funded by level monthly or annual premiums. A variant is **decreasing term insurance** where the death benefit decreases over the term of the policy. Decreasing term insurance may be used in conjunction with a home loan; if the policyholder dies, the outstanding loan is paid from the term insurance proceeds, sparing the policyholder's family from the potential difficulty, expense and distress of struggling to make the loan payments after the policyholder's death. The death benefit under the policy can be set to match the outstanding loan on the home in each year of the contract.

Renewable term insurance offers the policyholder the option to renew the policy at the end of the original term, without further evidence of the policyholder's health status, up to a maximum age. For example, if the maximum age is 85, a policyholder purchasing a 20-year renewable term policy at age 45 would have the option to renew the policy at age 65 for another 20 years. The premium for the renewed contract would be greater than the original, because the probability of paying the death benefit increases, but

insurability is guaranteed, which means that renewal does not depend at all on the health status of the policyholder at that time. In **yearly renewable term insurance** each individual contract has just a one-year term.

Convertible term insurance offers the policyholder the option to convert to a whole life insurance at the end of the original term, or, for a renewable term policy, at the end of the first or second renewal. The premium would be recalculated for the new whole life policy, depending on the age at conversion. Conversion does not depend on or require evidence of the policyholder's health status at conversion, but there may be a maximum age at which conversion is permitted (typically, around age 75).

1.3.3 Whole life insurance

Whole life, or **permanent**, **insurance** pays a lump sum benefit on the death of the policyholder whenever it occurs. For regular premium contracts, the premium is often payable only up to some maximum age, typically 80–90 years old. The point of whole life insurance is that it should pay the death benefit whenever the policyholder dies, and it would not meet this objective if policies lapsed at older ages through non-payment of premiums, whether because of financial strain or through decreasing ability of the policyholder to manage their affairs in older age.

In general, whole life insurance will be significantly more expensive than term insurance, relative to the death benefit, as the probability of paying the death benefit (ignoring lapses) is 100%.

Traditionally, if a whole life policyholder decides to discontinue the policy after an initial period, they would be eligible for a **cash value** or **surrender value**, representing the investment part of the accumulated premiums. In the early years of a whole life policy, the cash values tend to be very low. In later years they may be substantial, though typically very much less than the sum insured. Recently it has become common in some countries to offer whole life policies with no cash value payable on surrender.

Life insurance works by combining premiums with the investment income (earned by investing the premiums), such that, by the time the policyholder dies, the premiums plus the investment income are sufficient, on average, to pay the sum insured. For short-term policies, premiums cover most of the sum insured. For long-term contracts the investment income becomes a much more significant component. The term of a whole life contract may be very long – a policy sold to a life aged 40 may still be in force over 50 years later – so the contribution of investment income is much more significant than for most term insurance. However, predicting investment returns over very long terms is very difficult, so insurers tend to calculate premiums using very conservative (that is, low side) assumptions about investment returns, resulting in relatively

high premiums. This means that most policyholders (the ones who do not die very early) earn quite low rates of return on their premiums, compared with, for example, simply paying the same amount into a mutual fund investment over the same period.

Suppose an insurer is pricing a whole life policy at a time when long-term interest rates of 7% per year are available, for a 20-year investment. The insurer may calculate the premium assuming 6% per year interest, allowing the difference of 1%, which is called the **interest spread**, to cover profit and allow a margin for adverse experience. The risk to the insurer is that interest rates could fall below 6% per year at some point during the contract, in which case the interest earned on investments made at that time would fall short of the amount required to pay the sum insured. On the other hand, if the insurer is more cautious, perhaps assuming only 3% per year interest, the investment part of the policy will look quite unattractive for the customer, compared with the 7% per year available from direct investment.

One solution to this problem is to charge the higher premium, but promise to pay back to the policyholder some share of the profits if the investments do well. If the investment experience is very poor, there will be no profits, and no profit-sharing. This is the principle behind **participating** insurance, where 'participating' refers to the policyholders' participation in the distribution of profits. Participating insurance is also called **par** insurance for short, and is more commonly called **with-profit** outside North America.

The policyholders' share of profits is called **dividends** in North America, and **bonuses** elsewhere. We use the term 'dividend' when the profit share is distributed in the form of cash (or cash equivalent, such as a reduction in premium), and 'bonus' when the profit share is distributed in the form of additional insurance. In fact, the form of distribution is an important design feature for participating insurance, with different jurisdictions favouring different distribution methods, with the following being the most common.

Cash refunds may be distributed at regular (e.g. annual) intervals, based on the profit emerging in the preceding year. This is common for North American participating insurance.

Premium reductions work very similarly to cash refunds. The profit allocated to the policyholder for the year may be used to reduce the premiums due in the period to the subsequent allocation date.

Increased death benefits are determined by using the emerging profit to increase the death benefit.

There are many variants of these methods. It is common in North America for policyholders to be given some choice about the distribution – for example,

offering a cash dividend as a standard benefit, but with options to convert to additional death benefit.

In the UK, profits were invariably distributed in the form of benefit increases. Bonuses would be awarded in two stages. **Reversionary bonuses** are applied to contracts in force, increasing the benefits by a specified percentage. There are three variations:

- *simple reversionary bonus* means the bonus rate is applied to the original sum insured only;
- *compound reversionary bonus* means the bonus rate is applied to the total of the sum insured and previous reversionary bonuses;
- *super-compound reversionary bonus* is a method with two bonus rates each year, the first applying to the original sum insured, and the second applying to the total of previous bonus declarations.

Terminal bonuses are used to top up the sum insured when the benefit is finally paid. Separating the profit share into reversionary and terminal bonuses allows insurers to take a more cautious attitude to distributing unrealized capital gains.

It is important to note that for all traditional participating whole life insurance, dividends and bonuses are never negative. Only profits are shared, not losses.

Profit distribution methods for participating insurance have an important impact on actuarial management, and on the techniques for pricing and marketing of policies. We note some of the more important considerations here.

- Cash dividends are attractive to policyholders; they are easy to understand, and offer flexibility. If a policyholder is in financial difficulty, the cash may enable the policyholder to maintain the policy longer, as it can be used to offset premiums. If the policyholder wants to increase their death cover, the cash bonus can be used to buy more insurance – but at greater cost, generally, than the reversionary bonus, as it constitutes a new policy and therefore incurs new policy expense charges.
- Cash dividends may be taxable. If the policyholder has no need for the cash, it is not likely to be a tax-efficient asset.
- Reversionary bonuses are more complex for policyholders to understand, but do offer a tax efficient distribution, that is also consistent with the purpose of the policy – to provide a death benefit over long terms.
- Insurers may offer a limited share of profits for policies that are surrendered. This can be particularly unfair when profit is distributed as reversionary bonus, as policyholders who contributed to the profits each year may only receive a small proportion of them on surrender. If profits are shared through

cash dividends, then, at most, the policyholder would lose one-year's profit share on surrender.

- Cash dividends require the insurer to liquidate assets, which may not be in the best interests of maximizing return. Reversionary bonus means that profits remain under the insurer's management, and so provides more potential for future profit for the insurer.
- Generally, insurers prefer to offer smooth bonuses and dividends, that is, with little variation from year to year. This is generally easier with reversionary and terminal bonus, as the actual payment is delayed until the policy matures.
- Cash dividends are expensive to operate, if every policyholder is paid a dividend each year.

Traditional participating whole life insurance is still popular in North America, but is no longer widely available in the UK or Australia, though some non-participating whole life policies are still marketed. One reason for the relatively greater success of the product in North America is that insurers there offered larger and more predictable cash values, so that policyholders could achieve a reasonable return on their premiums, even if they surrendered the contract.

Whole life insurance may be used by policyholders in a number of ways, for example as follows.

- For older lives, simplified whole life insurance may be used to cover funeral expenses. These policies have a relatively low sum insured, tend to be non-participating, and do not offer any cash value on surrender.
- For older, higher-wealth lives, whole life insurance may be used to reduce inheritance taxes, if the proceeds from insurance are taxed at a lower marginal rate than directly inherited wealth.
- For younger lives, participating whole life insurance can provide a simple, passive long-term investment opportunity, with the advantage of substantial death benefit (compared with premiums) in the event of early death.

As mentioned above, for some policies, particularly those designed for older policyholders (for tax planning or funeral expenses), cash values may not be offered on surrender of the contract. This can reduce the premiums, as the excess funds from lapsed contracts subsidize the remaining policies. This is called **lapse-supported** insurance. However, if policyholders have the ability to sell their policies to a third party, then the lapsation profits may be very low. This has led to the rise of Stranger Owned Life Insurance, or **STOLI**, where an investment firm (usually specializing in this business) makes a cash payment to a policyholder who wants to surrender their policy. The investment firm

then takes over the payment of premiums as long as the original policyholder survives, and receives the sum insured on their death. If the value of the cash settlement to the original policyholder plus the cost of premiums after the sale of the policy is less than the value of the sum insured, then the investment firm makes a profit. If the original policyholder survives longer than expected, then the firm may make a loss. Often the policy is exchanged for a very deep discount on the sum insured, even for quite elderly policyholders, allowing the investment firm to make very significant profits on a large proportion of their policies.

1.3.4 Endowment insurance

Endowment insurance offers a lump sum benefit paid either on the death of the policyholder or at the end of a specified term, whichever occurs first. This is a hybrid of a term insurance and a fixed term investment. If the policyholder dies, the sum insured is paid, just as it would be under a term insurance; if the policyholder survives, the sum insured is paid at the end of the contract. Similarly to whole life insurance, the probability of a payout on an endowment insurance, ignoring lapses, is 100%. For similar reasons to the whole life case, endowment insurance typically offers cash values on early surrender, and may be issued in participating or non-participating forms.

Endowment insurance is no longer offered through mainstream insurers in North America or the UK. The main purpose of endowment insurance is as an investment, but the low returns offered and lack of flexibility meant that the contract could not compete with an increasing variety of pure investment options that became widely available in the latter part of the twentieth century. Traditional endowment insurance then evolved into modern insurance/investment hybrids such as the Universal Life or Unit Linked policies described later in this chapter.

It is interesting to note, however, that traditional endowment insurance policies are increasing in popularity in developing nations, notably for microinsurance, where the amounts involved are small. In this context endowment insurance policies may be used in conjunction with microfinance, to support small sum lending to individuals and small businesses who may not have access to traditional banking services.

1.3.5 Options and variations on traditional insurance

Insurance riders are optional benefits that a policyholder can select at the issue of a contract. In this section we describe some common riders and other variations associated with traditional insurance.

Joint life insurance: For term and whole life insurance, policies issued on 'joint lives' have premiums and benefits that depend on the survival of two people, typically spouses. The most common format is a first-to-die policy, where the death benefit is paid on the first death of the couple, provided (for term insurance) that the death occurs within the policy term.

Joint life policies are increasingly popular as households increasingly rely on the earnings of two partners, not just one.

Multiple life insurance: Similarly to joint life policies, multiple life term insurance policies offer a benefit payable on the first death, or on each death, within a specified group of individuals, provided death occurs within the term of the policy. This feature is commonly used to insure business partners.

Guaranteed cash values: As discussed above, whole life and endowment insurance policies usually offer cash values on surrender. The policyholder may be able to lock in guaranteed cash values by paying an additional premium. In some jurisdictions, guaranteed cash values are required by law.

Policy loans: For policies that offer cash values, policyholders may be able to borrow money from the insurer, using the cash value of the policy as collateral. A common use for this is to pay premiums when the policyholder cannot otherwise raise the necessary funds. When the policy with an outstanding policy loan attached becomes a claim (or is surrendered), the sum insured (or cash value) is reduced by the amount of the outstanding loan and interest.

Allowing policy loans increases the chance that the policyholder will continue with the policy, rather than surrender. Since, in most cases, continuing with the policy offers a better opportunity for the insurer to make profits, it is to the insurer's advantage to facilitate policy loans. Consequently, the interest rates charged on policy loans may be quite low, relative to market rates.

Accelerated benefits due to terminal illness: Under this rider, the death benefit will be paid early if the policyholder can provide medical documentation that they are suffering from a terminal illness, and are not expected to live more than one year.

The early payment of benefit on terminal illness is called an **accelerated death benefit**. For the insurer, paying the claim slightly early reduces the incentive for the policyholder to sell the policy on to a third party, in a special kind of end-of-life STOLI called a **viatical settlement**.