B.A.J. **9**, IV, 875-901 (2003)

HIGHLY PENETRANT GENETIC DISORDERS: AN ATLAS FOR CRITICAL ILLNESS INSURANCE

By A. S. Macdonald and S. Y. Yang

ABSTRACT

For estimates of critical illness (CI) insurance premiums in the presence of a known mutation leading to a genetic disorder, the key quantity is the penetrance, that is the probability $q(x)$ that the disease has developed by age x. This function is often estimated in the genetics literature, though typically with large confidence intervals. In this paper we suggest that the main features of real penetrance functions can be represented reasonably well by simple one-parameter families of functions, which can be scaled to fit the age range and lifetime penetrance. This gives a simple, direct, pragmatic way to obtain quick estimates of CI premium rates from published penetrance estimates, and also some indicative bounds for such premium rates, which are useful since confidence intervals usually cannot be estimated. To aid this process, as a short-cut to the solution of Thiele's equations in a multiple-state model, we give extensive tables in another report (Macdonald & Yang, 2003).

KEYWORDS

Critical Illness Insurance; Genetics; Mutation Penetrance; Premium Rating

CONTACT ADDRESS

A. S. Macdonald, Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh EH14 4AS, U.K. Tel: +44(0)131-451-3209; Fax: +44(0)131-451-3249; E-mail: A.S.Macdonald@ma.hw.ac.uk

1. Introduction

1.1 *Genetics and Actuarial Modelling*

Advances in human genetics, and especially the advent of DNA-based genetic tests, have raised many questions about insurance, relating to possible discrimination if insurers are allowed access to genetic information, or to possible adverse selection if they are not. In the United Kingdom in particular, government's approach to these questions has led to close scrutiny of the scientific basis of underwriting — so-called 'evidence-based underwriting' — which has in turn led to a need for actuarial models. Daykin *et al.* (2003) or Macdonald (2003) give recent reviews, which we will not reprise here.

Most actuarial research has concentrated on the single-gene disorders. These are the conditions in which a mutation in a single gene leads to onset

of a specific disorder with high probability. If onset (and death) tends to be delayed until adult ages, mutation carriers raise families like anyone else, the mutation may be inherited, and the disease is seen to 'run in families'.

Genetic tests for single-gene disorders divide the population into two or more subgroups, namely mutation carriers and non-carriers (there may be more than two because of genetic heterogeneity). The multiple-state framework is ideally suited to modelling the discrete sub-populations defined by a single-gene disorder, see Macdonald (2003) or Gutiérrez & Macdonald (2003) for examples. Each genetic disorder is characterised by:

(a) the transition intensities in the models representing onset and progression of the disease; and
(b) the frequency with which mutations are present in the population.

This is the domain of genetic epidemiology. Note that mutation frequencies matter only when studying the potential cost of adverse selection, because they define the size of the relevant group. We will not consider them further here.

1.2 *Genetic Epidemiology: the Penetrance Function*

Geneticists call the cumulative incidence of a genetic disorder the *penetrance*. 'Penetrance by age x', which we denote $q(x)$, is the probability that the disease has appeared by age x in a known mutation carrier, assuming that all other decrements (such as death) are absent. It is similar to the so-called single-decrement life table associated with one of the decrements in a multiple-decrement table. Let μ_x denote the associated rate of onset defined by:

$$q(x) = 1 - \exp\left(-\int_0^x \mu_t dt\right). \tag{1}$$

Because of equation (1), it does not matter whether we work with studies that estimate penetrance or that estimate rates of onset, as long as they are reported at enough ages.

Unfortunately, studies providing estimates of age-related rates of onset of disease or of death are not yet plentiful in the genetic epidemiology literature. Figure 1 gives an example of penetrance estimates obtained from the genetics literature (there are more examples in Section 3). Ford *et al.* (1998) estimated the penetrance of mutations in the BRCA1 gene, the event being onset of breast cancer. Point estimates of $q(x)$ were given at ages 30, 40, 50, 60 and 70, with 95% confidence intervals. Figure 1 also shows a function fitted to these values by Macdonald, Waters & Wekwete (2003a). It is important to emphasise that the underlying data from which these point estimates were obtained were *not* published, only the estimates themselves.

(a) Macdonald, Waters & Wekwete (2003b) applied this fitted penetrance function (among others) to multiple-state models of critical illness

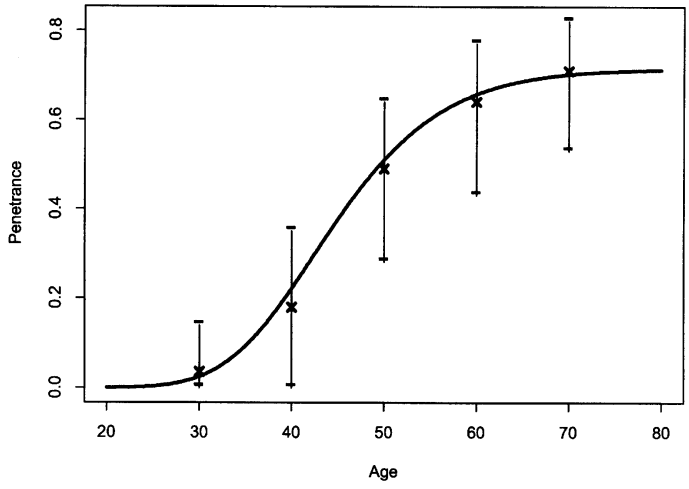

Figure 1. Observed values (×) and 95% confidence intervals of breast cancer penetrance associated with BRCA1 mutations, based on Ford *et al.* (1998); also shown is the fitted function from Macdonald, Waters & Wekwete (2003a)

insurance. Once these models are programmed, re-running them to study a different genetic disorder is often simply a case of plugging in the appropriate penetrance function.

(b) The confidence intervals are very wide, even although the study was based on a worldwide collection of affected families (this often happens with rare genetic disorders). Clearly, such stochastic uncertainty carries through to any financial quantities that appear in actuarial problems. This is an important question that has not yet been studied.

(c) To quantify this uncertainty, we would usually specify a model in which a continuous random variable **T** represented age at onset of breast cancer. The published data alone (which are shown *in toto* in Figure 1) are not sufficient to parameterise this model using any standard methods from survival analysis (although the full underlying data might be).

The preceding observations motivate this paper. From (a), we see that it would be useful quickly to map out the effect of any single-gene disorder on insurance premiums and adverse selection. Then, as genetic epidemiology yields more and better penetrance estimates in future, their importance for insurance can be seen immediately, almost like looking up a table. One aim of this paper is to provide just such an 'atlas', in respect of critical illness insurance.

From (b) and (c), we see that it is also desirable to have some measure of the uncertainty in actuarial quantities arising from the uncertainty of the penetrance estimates, even when only these estimates themselves may be available, and the full underlying data are not.

1.3 *An 'Atlas' of Single-Gene Genetic Disorders and Critical Illness Insurance*

The main features of penetrance, of importance for actuarial models, are:

(a) the maximum reached by $q(x)$ (generally less than 1), denoted Q;
(b) the minimum age at which symptoms (or death) can occur (that is, at which $q(x) > 0$), denoted x_m;
(c) the age at which the maximum of $q(x)$ is attained, denoted x_M; and
(d) the shape of the function $q(x)$, which determines the severity of the disorder.

This is a reasonably small number of factors. Table 1 shows x_m, x_M and Q in respect of several single-gene disorders. Actuaries are possibly more interested than are clinicians in the penetrance over the whole age range, since the need to consider insurance contracts for any reasonable combination of age and term means that no one segment of the penetrance curve is more important than any other segment.

We proceed by defining a simple parametric family of model penetrance curves, to give a representative range of 'shapes', which can then be scaled to cover the ranges of x_m, x_M and Q that are found in practice. The resulting 'grid' of model penetrance curves can then be used in an actuarial model to obtain and to tabulate premium rates. Given estimates and confidence intervals of any new penetrance curve, from the epidemiological literature, it is then a simple matter to read off approximate values or ranges of critical illness premium rates. Given estimates of mutation frequencies also, we could do the same for the costs of adverse selection, but here we look only at premium rates.

Table 1. Boundaries of fitted penetrance functions for several single-gene disorders

Study	Disorder	Min age x_m	Max age x_M	Final penetrance Q (%)	Fitted functional form
Guttierez *et al*. (2003)	APKD (Churchill)	20	73	47.2%	μ_x = Beta
Guttierez *et al*. (2003)	APKD (USRDS)	20	80	51.1%	μ_x = Beta
Macdonald *et al*. (2003a)	BC (BRCA1)	20	80	71.2%	μ_x = Gamma
Macdonald *et al*. (2003a)	BC (BRCA2)	30	80	95.5%	μ_x = Gamma
Macdonald *et al*. (2003a)	OC (BRCA1)	30	80	43.8%	μ_x = Gamma
Macdonald *et al*. (2003a)	OC (BRCA2)	40	80	41.8%	μ_x = Gamma
Smith (1998)	HD	0	80	100%	$q(x)$ = Normal
Gui *et al*. (2002)	EOAD (Presenilin-1)	20	80	100%	μ_x = polynomial

1.4 *Plan of this Paper*

In Section 2 we discuss suitable one-parameter families of model penetrance functions. In Section 3 we show that a Beta distribution (or occasionally a sine function) with a proportional hazards (PH) distortion provides a reasonable representation of the penetrance of several important disorders. In fact, by superimposing the modelled penetrance functions, including confidence intervals where these are available, on our one-parameter families of functions, we get a clear impression of the great degree of uncertainty that attaches to published penetrance estimates.

In Section 4 we describe briefly a model for critical illness (CI) insurance in which onset of a genetic disorder may be a reason for a claim. This is used to compute tables of representative extra premiums for each family of model penetrance functions. These tables are too extensive to include, and except for a few examples we omit them from this paper, but we have deposited them in the libraries of the Faculty of Actuaries and Institute of Actuaries, as Occasional Actuarial Research Document No. 42 (Macdonald & Yang, 2003). In Section 5 we give some examples of premium rates and ranges of premium rates. Our conclusions are in Section 6.

2. Model Families of Penetrance Curves

Experience suggests that at least some penetrance curves may be close to symmetrical. Huntington's disease is perhaps the most studied of all genetic disorders, and the Normal distribution has been suggested as a good model for its penetrance by several authors (Bell, 1934; Wendt & Drohm, 1972; Roos *et al.*, 1991; Wilkie, 2000). Thus we propose to take some simple symmetric functions as starting points, and obtain simple parametric families from them by applying suitable 'distortions'. Examples of such symmetric functions are:

(a) a sine function on $[-\pi, \pi]$, suitably scaled;
(b) a Normal distribution function (see above); this does not have a finite range, but as long as it is scaled so that $[x_m, x_M]$ is about -3 to $+3$ standard deviations this is immaterial; and
(c) a Beta distribution function $\beta(p_1, p_2)$; this does have a finite range, and can produce different shapes of penetrance curves by suitable choice of the parameters p_1 and p_2.

Figure 2 shows the following sine function, Normal distribution function and Beta(5,5) distribution function, scaled to represent a disorder with onset after age 20, fully penetrant by age 60 ($x_m = 20$, $x_M = 60$, $Q = 1$):

$$q^{\sin}(x) = \frac{1}{2}\left(\sin\left(\frac{2x - x_m - x_M}{2(x_M - x_m)}.\pi\right) + 1\right) \tag{2}$$

$$q^{\text{normal}}(x) = \int_{x_m}^{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{3(2t - x_m - x_M)}{x_M - x_m}\right)^2\right) dt \tag{3}$$

$$q^{\text{beta}}(x) = \int_{0}^{x} \frac{\Gamma(10)}{\Gamma(5)\Gamma(5)} \left(\frac{t - x_m}{x_M - x_m}\right)^4 \left(1 - \frac{t - x_m}{x_M - x_m}\right)^4 dt. \tag{4}$$

We can see that all of these have the right general properties. Other functions would have too, these are just plausible candidates, and the usual question of fitting an adequate model to the data arises.

(a) Genetic epidemiologists often fit simple parametric or semi-parametric models of incidence rates, such as piecewise linear, piecewise exponential or Weibull, usually by likelihood methods. Given the data, so could actuaries.

(b) Geneticists often publish non-parametric estimates of penetrance (equivalently, survival free of disease), usually Kaplan-Meier survival functions. Note the distinction between this empirical reporting and the modelling described in (a).

(c) The problem, which has been a feature of almost all actuarial models of genetics and insurance, is that such models need fitted estimates as in (a), but all that is available are empirical reports as in (b), or (as in

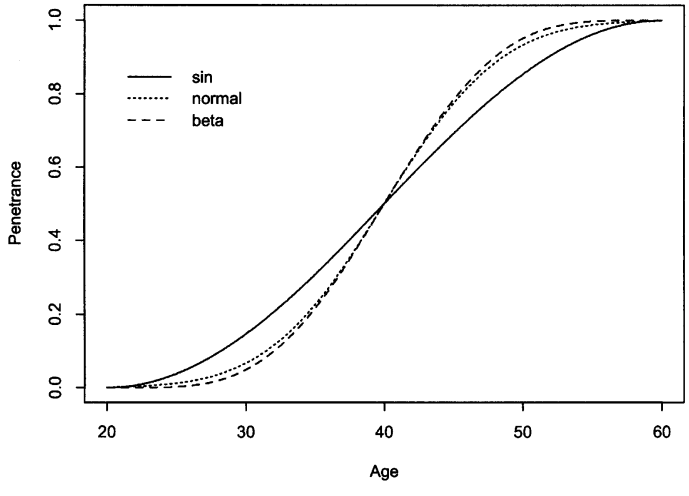

Figure 2. An example of a sine function, Normal distribution function and Beta distribution function, scaled to represent a disorder with onset after age 20, that is 100% penetrant by age 60 ($x_m = 20$, $x_M = 60$, $Q = 1$)

Figure 1) very sparse point estimates. Data are practically never available. This means that the only option that the actuary has is to find a model penetrance function that is a reasonable approximation to whatever evidence is to hand.

In the absence of the data, there is no question of formal procedures such as model selection or finding a best-fitting function. It is not unlike the old problem of graduating very sparse mortality data, but without the data. What we must do is as follows:

(a) We need to choose one or more plausible families of model penetrance functions. The motivation for this is entirely empirical, based on comparisons with actual penetrance functions that are, necessarily, fairly informal. This is the subject of Section 3.

(b) Empirical evidence suggests that some disorders have ages at onset distributed rather more evenly over the relevant age range, and this leads to relatively 'flat' penetrance functions. Others seem to have onset ages more concentrated in the middle of the age range, but with less frequent cases at lower and higher ages; this leads to a relatively 'steep' penetrance function. In Section 3 we find that different families are needed to represent both of these patterns.

(c) Although many penetrance functions estimated from genetic data seem to be nearly symmetrical, not all are; some disorders have onset skewed towards higher or lower ages within quite a long range of ages (possibly an indication of heterogeneity of the underlying mutations, but this is usually hard to verify). To meet this, we extend our basic symmetrical functions, each to a one-parameter model family of penetrance functions, by applying a distortion that shifts the onset of the disorder towards earlier or later ages.

In this paper we consider a proportional hazards transformation: if $q(x)$ is the basic symmetrical penetrance curve, and μ_x the corresponding hazard (rate of onset), then any positive constant k defines a new penetrance function $q_k(x)$ by:

$$q_k(x) = 1 - (1 - q(x))^k = 1 - \exp\left(-\int_0^x \mu_t dt\right)^k = 1 - \exp\left(-\int_0^x k\mu_t dt\right). \tag{5}$$

It is convenient to parameterise the distortion in terms of the penetrance itself. For example, a one-parameter family may be uniquely specified in terms of the penetrance at any interior age, the middle age $(x_m + x_M)/2$ being an obvious choice (the baseline function, by symmetry, will always have penetrance $Q/2$ at this age). A two-parameter family could be specified in terms of the penetrances at ages

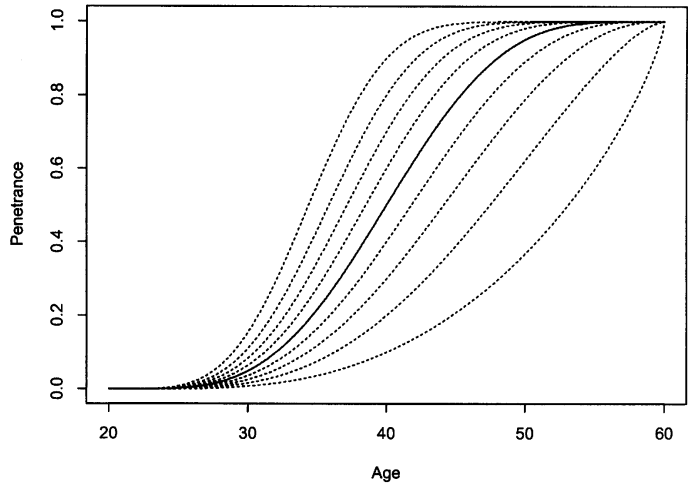

Figure 3. A Beta(5,5) distribution as a hypothetical penetrance curve, and a proportional hazards family based upon it (penetrances from 10% to 90% at age 40 (the middle age) shown)

$(x_m + x_M)/3$ and $2(x_m + x_M)/3$, and so on. For the proportional hazards family above, penetrance of $(1 + r)Q/2$ at age $(x_m + x_M)/2$ (for $-1 < r < 1$) is obtained by choosing:

$$k = 1 + \frac{\log(1 - r)}{\log(0.5)}. \tag{6}$$

Figure 3 gives an example, in which $x_m = 20$, $x_M = 60$, $Q = 1$, the basic symmetric function is Beta(5,5) (suitable scaled), distorted by proportional hazards to give penetrance of 10%, 20%, ..., 80%, 90% at the middle age, 40 years.

After a great deal of experimentation, which we do not describe here, we found that the Beta and sine functions, as basic penetrance curves, and the proportional hazards distortions gave results as good as any other one-parameter family, and from now on our discussion centres on them. Section 3 will make it clear that attempts to find any best-fitting model are quite inappropriate here, given the absence of data and (when given) the very wide confidence intervals of the penetrance estimates (see Figure 1).

For critical illness (CI) insurance, the financially important part of the age range is about 20–65. With this in mind, we choose the age intervals shown in

Table 2. Model penetrance functions tabulated in Macdonald & Yang (2003)

Function	Age range	Function	Age range
Beta	0–80		
Beta	0–60		
Beta	0–40	sine	10–80
Beta	10–80	sine	10–60
Beta	10–60	sine	10–40
Beta	20–80	sine	20–80
Beta	20–60	sine	20–60
Beta	30–80	sine	30–80
Beta	30–60	sine	30–60
Beta	40–80	sine	40–80
Beta	40–60	sine	40–60

Table 2, and five lifetime penetrance levels: 20%, 40%, 60%, 80% and 100%. We will show, in Section 3, that these include penetrance curves acceptably close to at least some examples drawn from real genetic disorders.

3. Comparison of Real and Hypothetical Penetrance Curves

3.1 *Studies of Penetrance*

In this section we compare some penetrance curves taken or estimated from the epidemiological literature (see Table 1) with our one-parameter families of penetrance curves. This serves two purposes:

(a) It should give a clear idea of how an 'atlas' such as this can be used in practice; given a penetrance function in respect of some disorder, it can quickly be compared with one or more of the families of hypothetical penetrance curves, matching it for age range and lifetime penetrance, and the nearest fitting members found, perhaps giving a range of reasonable approximations. Then, by referring to the tables constructed as in Section 5, and available in full in Macdonald & Yang (2003), the extra premiums for CI insurance can immediately be read off, or at least upper and lower limits may be found.

(b) It illustrates what large errors may be present in estimates of penetrances, because of the typically modest sample sizes (by actuarial standards). This is especially important in actuarial work, where questions of insurance premiums for individuals need quite detailed estimates of risk. This 'atlas' may also be used, therefore, to indicate how much uncertainty there might be in insurance premiums based upon quite uncertain epidemiological data. If original data are available, we might be able to improve upon this by modelling it directly, but this is quite unusual.

Note that we are considering only estimation error here; that is, we suppose that the estimates were based on data that are a fair, unbiased sample of the relevant population, by which we mean those people who might be applying for insurance. This ignores ascertainment bias, which is a major feature of genetic epidemiology: the epidemiologist studies families in which the disorder is present, and in the absence of a prospective, population-based study design it is very likely that families have been included precisely *because* they have an unusually strong family history.

However, it is not clear that ascertainment bias ought always to be removed, or adjusted for, in addressing insurance questions. If it is the case that genetic tests are only available in a clinical setting, and that the only reason for being referred to a genetics clinic by one's primary carer is the presence of a family history, then those persons applying for insurance with knowledge of a genetic test result will be drawn from a population similar to that which might be sampled in epidemiological studies. In other words, the penetrance affected by ascertainment 'bias' is, in fact, correct for the insured population. However, if access to genetic information was less restricted, this argument might fall away.

In the following sections we consider some of the studies in the actuarial literature. In considering how well the families of Beta (or sine) functions represent those based on epidemiological studies, we should remember that ages above 60 are of little importance for CI insurance, and that ages over about 50 may often be of little relevance in the context of severe single-gene disorders. Where possible, we comment on the selection of cases into the original studies, and possible ascertainment biases (see Hodge (2002) for a discussion of ascertainment bias).

3.2 *APKD, based on Churchill* et al. *(1984)*

Churchill *et al.* (1984) gave a Kaplan-Meier estimate of the distribution of the time to the first of end-stage renal disease (ESRD, meaning kidney failure) or death, among persons affected by APKD. Death quickly follows ESRD unless dialysis and transplant services are available, and ESRD would give rise to a critical illness insurance claim. Gutiérrez & Macdonald (2003) used this estimate to model the rate of onset of ESRD. The estimate and their fitted penetrance curve are shown in Figure 4, along with an appropriate family of model (Beta) penetrance functions with $Q = 0.48$. We see that the symmetric Beta function itself, with no proportional hazards distortion, is an excellent fit up to age 60.

This study was based on 140 persons from 17 affected kindreds (100 documented with APKD, 32 suspected and eight unknown). Kindreds were ascertained by a survey of nephrologists in Newfoundland and Labrador, and as many members as possible were scanned by ultrasound to determine the presence of APKD. Steps were taken to reduce ascertainment bias, for example by excluding the index cases through which the kindreds had been identified.

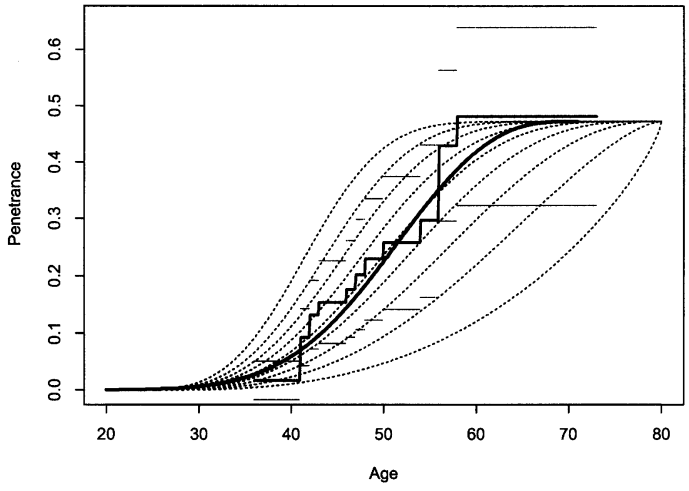

Figure 4. Observed values and graduated penetrance curve of APKD with 95% confidence interval, based on Churchill *et al.* (1984) and Gutiérrez & Macdonald (2003), compared with a proportional hazards (Beta) family

3.3 *APKD, based on USRDS Data*

The United States Renal Disease Survey (USRDS, 1999) provided data on the incidence of ESRD caused by APKD in the U.S.A., from which Gutiérrez & Macdonald (2003) obtained rates of onset. Figure 5 shows the corresponding penetrance curve, compared with an appropriate family of model (Beta) penetrance functions with $Q = 0.51$. Again, the fit is very good; the model penetrance function with penetrance 20% of the maximum at the middle of the age range is a good representation.

Because they are based on population data, these estimates are free of the ascertainment bias that arises from the selection of small samples, but they may be affected by the usual biases of incomplete disease registration, wrong diagnoses and incomplete censuses.

3.4 *Breast Cancer, BRCA1 and BRCA2 Mutations*

Breast and ovarian cancers (BC and OC) have relatively rare familial forms that are known to be caused by mutations in the BRCA1 and BRCA2 genes, and these have been intensively studied. Figures 6 and 7 show the point estimates and their 95% confidence intervals (Ford *et al.*, 1998), compared with suitable (Beta) families of model penetrance functions. Also shown are the penetrance functions fitted by Macdonald, Waters & Wekwete (2003a). In Figure 6 $Q = 0.71$, and in Figure 7 $Q = 0.955$.

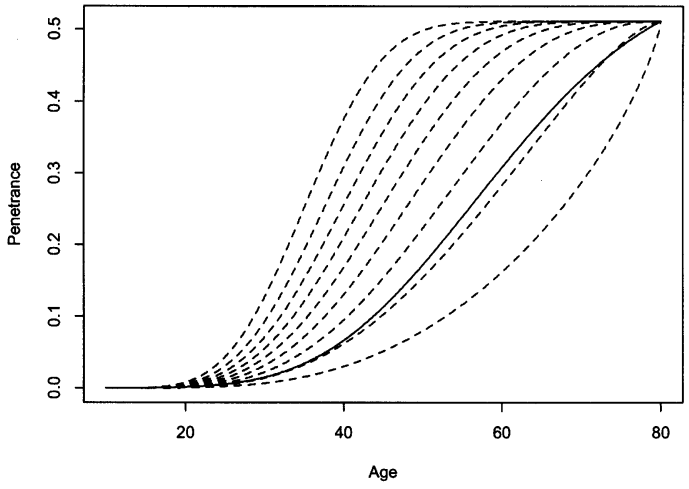

Figure 5. Penetrance curve of APKD based on USRDS data and Gutiérrez & Macdonald (2003), compared with a proportional hazards (Beta) family

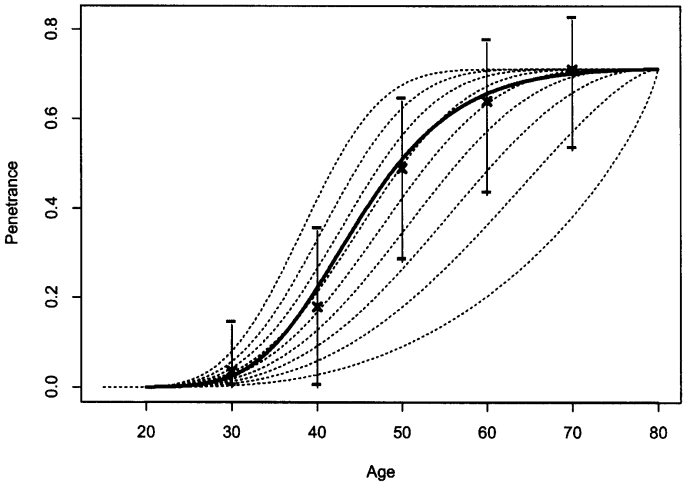

Figure 6. Observed values, 95% confidence intervals and fitted penetrance curve for breast cancer, BRCA1 mutations, based on Macdonald, Waters & Wekwete (2003a), compared with a proportional hazards (Beta) family

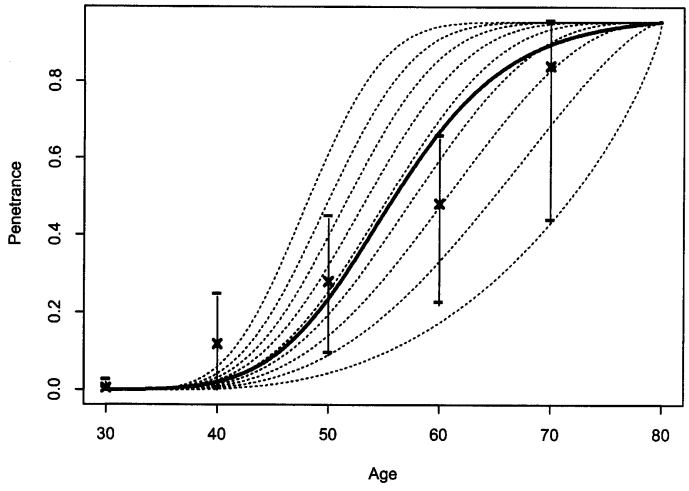

Figure 7. Observed values, 95% confidence intervals and fitted penetrance curve for breast cancer, BRCA2 mutations, based on Macdonald, Waters & Wekwete (2003a), compared with a proportional hazards (Beta) family

A most important point is that the model penetrance functions should be compared with the point estimates and not the previously fitted penetrance functions (the bold lines). The actuary faced with the original data would have only the former to go on. In this case, it so happens that the penetrances have been modelled in a previous actuarial research paper, so it is of some interest to show them as well, but strictly they are beside the point. This applies to subsequent figures as well.

Again, these model penetrance functions provide quite good fits up to age 60. For BRCA1 mutations, penetrance seems to be slightly shifted towards earlier ages, with penetrance reaching about 60% of Q by the middle of the age range, while for BRCA2 mutations the penetrance function is close to symmetrical. However, the confidence intervals show how much room there is for doubt. For example, in Figure 6, five (nearly six) of the model penetrance functions lie within all the confidence intervals. Moreover, these are pointwise confidence intervals, and confidence bands (calculated simultaneously for all five ages) would be even wider.

Ascertainment bias is a large issue. Families from all over the world were selected for the study, run by the international Breast Cancer Linkage Consortium (BCLC), precisely on the basis of multiple affected members in several generations. Much of the literature since Ford *et al.* (1998) has focused on the difference among mutation carriers in unselected families and those in the BCLC families, with inconclusive results so far.

3.5 *Ovarian Cancer, BRCA1 and BRCA2 Mutations*

Mutations in BRCA1 and BRCA2 are also associated with ovarian cancer. Figures 8 and 9 show point estimates of penetrance (to onset of ovarian cancer) based on Ford *et al.* (1998), for BRCA1 and BRCA2 mutations, respectively. Ascertainment bias is a strong possibility, since these are the same BCLC families referred to in Section 3.4. Also shown are fitted curves from Macdonald, Waters & Wekwete (2003a). This time the model penetrance curves are based on sine functions, as these gave a slightly better representation. No confidence intervals are shown in Figure 8, since Ford *et al.* (1998) gave estimates for the penetrance of BRCA1 mutations in respect of breast and ovarian cancers combined; these are shown in Figure 10, with a Beta family of model penetrance curves for comparison.

In all of Figures 8 to 10, the fitted curves seem reasonable, and the families of model penetrance curves do provide close approximations to the fitted curves, but the confidence intervals do not give very strong support for these estimates.

3.6 *Huntington's Disease*

As mentioned before, several authors have suggested that a Normal distribution represents well the age at onset distribution of Huntington's disease. Since this is one of the clearest single-gene disorders, with practically 100% penetrance, it has been studied for a long time, since well before the responsible gene was identified, and large studies have been carried out. Subsequently, however, the mechanism of Huntington's disease has been found to be a variable number of repeats of the trinucleotide CAG in the Huntington gene, and lower ages at onset are associated with larger numbers of CAG repeats (see Gutiérrez & Macdonald, 2002a, 2002b). What appeared to be the simplest and clearest example of a single-gene disorder has, therefore, turned out to be extremely heterogeneous, and earlier estimates of penetrance ignoring the CAG repeat length have become less relevant. Moreover, although it may be possible to estimate moments of the fitted parameters (of a Normal age-at-onset distribution), these do not by themselves provide confidence intervals for the penetrance curve, which might be found by some procedure such as bootstrapping.

As an example, Figure 11 shows the Normal penetrance curve suggested by Wilkie (2000), compared with a Beta family of model penetrance curves. Of course, we would get a better (in fact, perfect) fit if we used a Normal family of model penetrance curves, but this does serve to show that the Beta family can provide a good approximation.

Ascertainment bias is often less of an issue with Huntington's disease, because several countries or regions have compiled near-complete registers of affected families. This is the case for Wilkie (2000), who used onset rates based on the Leiden register in the Netherlands.

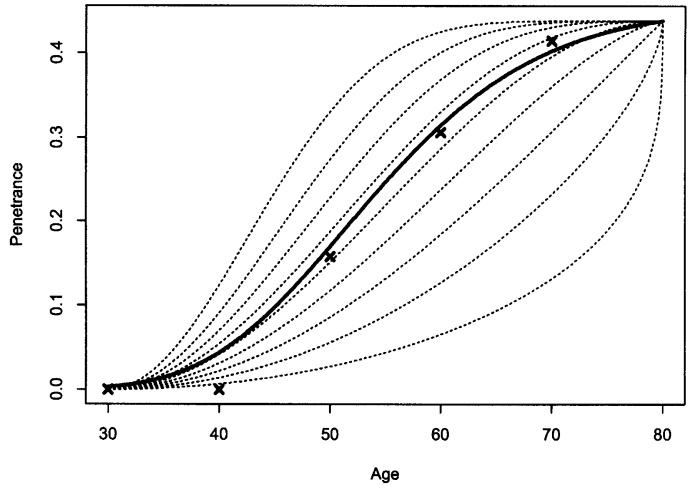

Figure 8. Observed values and fitted penetrance curve for ovarian cancer, BRCA1 mutations, based on Macdonald, Waters & Wekwete (2003a), compared with a proportional hazards (sine) family

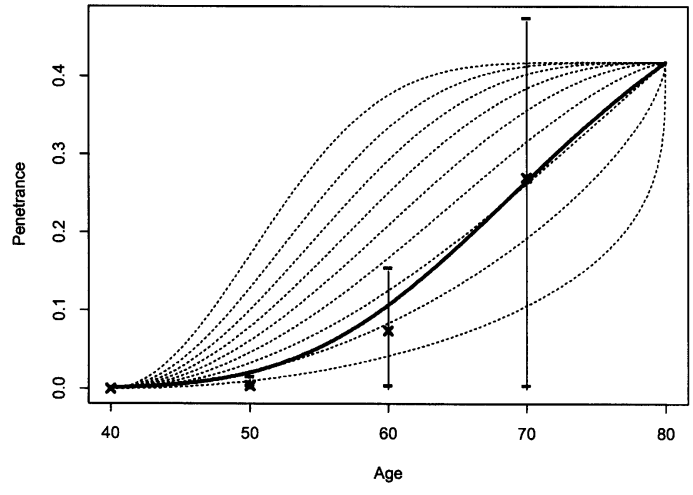

Figure 9. Observed values, 95% confidence intervals and fitted penetrance curve for ovarian cancer, BRCA2 mutations, based on Macdonald, Waters & Wekwete (2003a), compared with a proportional hazards (sine) family

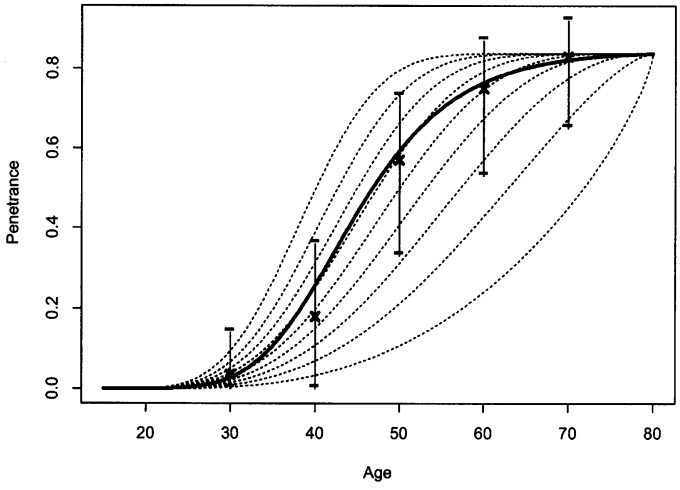

Figure 10. Observed values, 95% confidence intervals and fitted penetrance curve for breast and ovarian cancer combined, BRCA1 mutations, based on Macdonald, Waters & Wekwete (2003a), compared with a proportional hazards (Beta) family

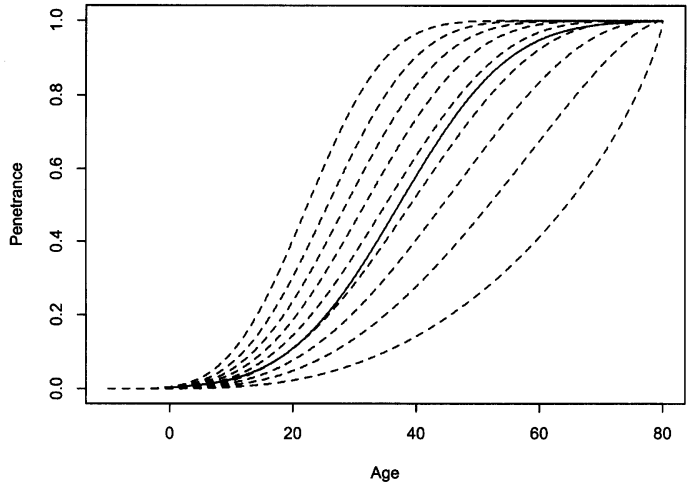

Figure 11. Penetrance curve of Huntington's disease, based on Wilkie (2000), compared with a proportional hazards (Beta) family

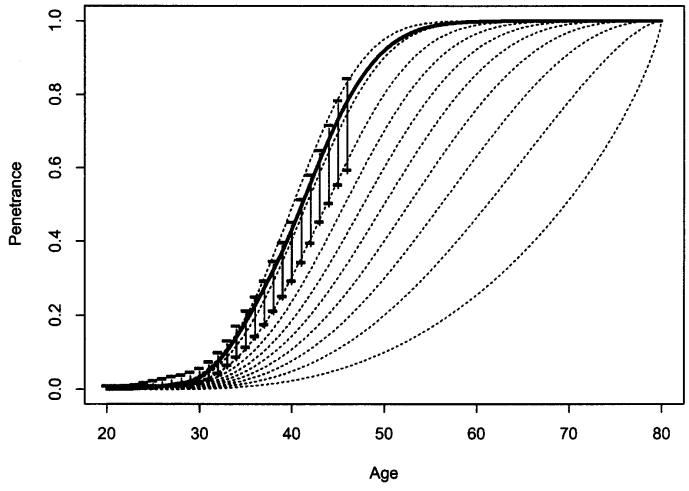

Figure 12. Graduated penetrance curve with 95% confidence interval of early-onset Alzheimer's disease, Presenilin-1 mutations, based on Gui & Macdonald (2002), compared with a proportional hazards (sine) family

3.7 *Early Onset Alzheimer's Disease, Presenilin-1 Mutations*

Gui & Macdonald (2002) estimated the penetrance of Presenilin-1 mutations, one of the known causes of early-onset Alzheimer's disease. This is a version of Alzheimer's disease, normally a disorder of old age, occurring before age 60. Their estimate was derived from a smoothed Nelson-Aalen estimate, and therefore took the form of a numerical estimate of the entire function $q(x)$, with confidence intervals at integer ages obtained by simulation. For reasons explained in Gui & Macdonald (2002), the estimate was unreliable after about age 45, and even at earlier ages was probably subject to ascertainment bias and certainly affected by missing data.

Figure 12 shows the penetrance curve up to age 50, with confidence intervals, compared with a Beta family of model penetrance curves. (In this case, the fitted penetrance curve has a penetrance of just over 90% of Q at the mid-point of the age range, so we show an additional member with penetrance of 95% of Q at that age.) We see that, even in this extreme case, the Beta family provides a good approximation.

3.8 *A Comment on the Choice of Model Penetrance Function(s)*

Here we have superimposed model penetrance functions on some empirical penetrance functions or point estimates, and have commented

on goodness of fit, but only informally. We could try to find a 'best-fitting' model curve more formally, but we have not, for the following reasons.

(a) The quality of the *published* data is, as we have emphasised throughout, often very poor, and confidence intervals, when given, very large. Elaborate fitting procedures would usually by largely spurious. In these circumstances, it makes more sense to look for a range of reasonable answers.

(b) Model penetrance functions are only a means to an end, which is approximate extra premiums, here for CI insurance. Our aim is to allow the user to obtain these with an economy of effort commensurate with the quality of the data, *via* the tables in Macdonald & Yang (2003).

3.9 *Summary*

Table 3 summarises the disorders discussed here and the model penetrance curves that we have used. Our model penetrance curves, based on a symmetrical onset distribution and a simple one-parameter family of transformations, may not reproduce any the fitted penetrance curves perfectly, but do often come close, especially over ages 20–60. Given the great uncertainly that usually attaches to penetrance estimates, it provides a satisfactory basis for representing any penetrance estimates that might appear in the genetics literature.

Table 3. Summary of single-gene disorders and the model penetrance curves

Actual disorder	Age range $[x_m, x_M]$	Baseline function $q(x)$	Maximum penetrance Q	Penetrance at midpoint of age range of closest-fitting curve
APKD (Churchill)	[20,80]	Beta	47.2%	50% of Q
APKD (USRDS)	[10,80]	Beta	51.1%	20% of Q
BC (BRCA1)	[20,80]	Beta	71.2%	60% of Q
BC (BRCA2)	[30,80]	Beta	95.5%	50% of Q
OC (BRCA1)	[30,80]	sine	43.8%	50% of Q
OC (BRCA2)	[40,80]	sine	41.8%	20% of Q
BC & OC (BRCA1)	[15,80]	Beta	83.8%	60% of Q
BC & OC (BRCA2)	[30,80]	Beta	97.4%	50% of Q
BC & OC (BRCA1/2)	[15,80]	Beta	99.6%	60% of Q
HD	[−10,80]	Beta	100%	40% of Q
EOAD	[20,80]	Beta	100%	90% of Q

4. A Model of Critical Illness Insurance

Figure 13 shows a simple model suitable for pricing and reserving for critical illness (CI) insurance contracts. State 2 represents the onset of the genetic disorder in question, state 3 the occurrence of any other event that triggers a claim under the contract, and state 4 death before any claim occurs. This is a continuous-time model parameterised by the transition intensities $\mu_{jk}(x)$, which are functions of the age x. These models are now standard in the actuarial literature, and we refer to Hoem (1988) for details.

Note that onset of the genetic disorder does not always trigger a CI insurance claim. Where 'onset' means kidney failure (as in APKD) or a diagnosis of cancer, then a valid claim would result, but in respect of neurological disorders such as Alzheimer's disease or Huntington's disease, 'onset' often means the earliest detectable signs of cognitive impairment, and a claim might only arise some time later when there is serious disability. This affects Huntington's disease in particular, because the disorder can progress quite slowly.

For any given genotype, level net premiums under a CI insurance contract of any age and term can be calculated in the usual way as:

$$\frac{\text{EPV of unit benefit}}{\text{EPV of unit annuity}}$$

solving Thiele's equation for the expected present values (EPVs) numerically by any convenient method. We used a Runge-Kutta algorithm with a step-size of 0.0005 years, but many standard mathematics packages could also be used.

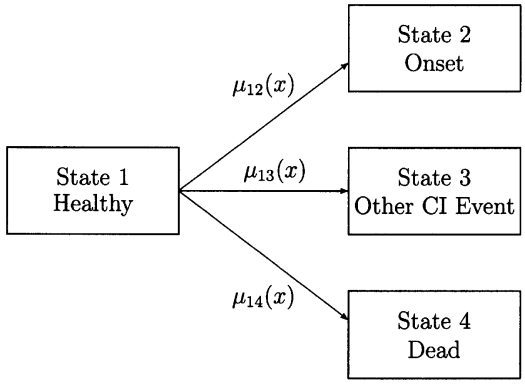

Figure 13. A multiple state model for onset of a genetic disorder in critical illness insurance

The transition intensities $\mu_{13}(x)$ and $\mu_{14}(x)$, representing 'non-genetic' events, are taken from Gutiérrez & Macdonald (2003), and are described briefly in the Appendix.

5. Examples: Breast/Ovarian Cancer and Early Onset Alzheimer's Disease

5.1 *Tables of Extra Premiums*

Once parameters have been obtained as in Table 3, by simple graphical methods, premium rates can of course be obtained by direct solution of Thiele's equations. However, this is not (yet) a standard procedure for most actuaries, so, as a short cut, we have produced representative tables.

Our tables show the percentage extra premiums for mutation carriers, relative to the standard net premiums for level CI insurance cover. The mutation penetrance is represented by the families of model penetrance curves. Tables 4 to 7 show these extra premiums for females, with onset in the range 10–80 and 20–80 years of age, penetrance based on a Beta family, and maximum penetrance Q of 80% and 100%. Macdonald & Yang (2003) contains the full range of tables, for males and females, described in Section 2, of which these are examples.

5.2 *Breast/Ovarian Cancer and BRCA1 Mutations*

We consider the onset of either breast or ovarian cancer. Figure 10 and Table 3 showed that BRCA1 mutations were represented well by one of the Beta families of model penetrance curves on the age range 15–80, specifically the curve which reached 60% of the maximum penetrance (83.8%) at age 47.5. Taking $Q \approx 0.8$, reference to Tables 4 and 6, shows that the extra premium for CI cover from (say) age 30 to age 50 is approximately:

$$(1{,}484 + 975)/2 \approx 1{,}230\% \tag{7}$$

or, taking $Q \approx 0.838$ (though such 'accuracy' is certainly unwarranted), we get an extra premium of about:

$$(0.81\,(1{,}484 + 975) + 0.19\,(2{,}063 + 1{,}272))/2 \approx 1{,}312\%. \tag{8}$$

Macdonald, Waters & Wekwete (2003b, Table 6) gave an extra premium of 1,423%, though with a slightly different model of the non-genetic CI claims.

5.3 *Early Onset Alzheimer's Disease and Presenilin-1 Mutations*

Figure 12 showed that Presenilin-1 mutations were represented well by a Beta family of model penetrance curves on the age range 20–80, specifically the fitted penetrance lay a little above the curve which reached 90% of the

Table 4. Percentage extra premium for level CI cover for females with Beta family of model penetrance curves, maximum penetrance 80%, age at onset in range 10–80 years

Percentage of maximum penetrance at mid-age	Age 20 at entry				Age 30 at entry			Age 40 at entry		Age 50 at entry
	Term 10 yrs	Term 20 yrs	Term 30 yrs	Term 40 yrs	Term 10 yrs	Term 20 yrs	Term 30 yrs	Term 10 yrs	Term 20 yrs	Term 10 yrs
10%	168.1%	212.9%	202.3%	177.5%	243.5%	215.8%	186.4%	204.7%	180.0%	170.2%
20%	354.8%	443.5%	409.1%	342.0%	509.8%	441.3%	364.7%	419.5%	354.2%	330.5%
30%	565.2%	695.8%	621.6%	495.8%	804.1%	678.0%	536.3%	644.5%	520.1%	470.6%
40%	806.1%	975.4%	841.5%	641.9%	1133.9%	928.2%	703.2%	878.7%	674.7%	575.9%
50%	1088.5%	1290.9%	1071.6%	784.3%	1510.3%	1195.0%	868.1%	1119.8%	813.4%	627.4%
60%	1430.3%	1655.7%	1316.6%	928.6%	1950.8%	1483.6%	1035.4%	1361.0%	928.8%	605.5%
70%	1865.0%	2094.0%	1585.6%	1083.5%	2486.1%	1802.4%	1211.9%	1586.0%	1007.8%	499.3%
80%	2466.2%	2656.2%	1898.6%	1265.0%	3179.2%	2168.5%	1409.4%	1751.4%	1022.8%	321.8%
90%	3463.7%	3488.7%	2317.9%	1517.4%	4202.5%	2624.8%	1653.9%	1713.7%	899.6%	123.1%

Table 5. Percentage extra premium for level CI cover for females with Beta family of model penetrance curves, maximum penetrance 100%, age at onset in range 10–80 years

Percentage of maximum penetrance at mid-age	Age 20 at entry				Age 30 at entry			Age 40 at entry		Age 50 at entry
	Term 10 yrs	Term 20 yrs	Term 30 yrs	Term 40 yrs	Term 10 yrs	Term 20 yrs	Term 30 yrs	Term 10 yrs	Term 20 yrs	Term 10 yrs
10%	210.3%	266.9%	254.4%	224.1%	306.2%	273.0%	237.2%	261.4%	232.5%	224.8%
20%	444.4%	557.7%	517.8%	436.2%	645.7%	565.6%	473.7%	549.9%	475.9%	470.4%
30%	708.6%	878.1%	792.5%	639.0%	1027.0%	882.3%	712.5%	872.1%	733.7%	741.9%
40%	1012.0%	1236.1%	1081.4%	836.6%	1462.5%	1229.6%	958.1%	1238.0%	1011.2%	1046.9%
50%	1368.7%	1643.8%	1389.6%	1034.6%	1971.1%	1617.7%	1218.2%	1662.6%	1317.7%	1397.1%
60%	1801.9%	2120.7%	1725.4%	1241.4%	2584.4%	2063.0%	1505.1%	2171.1%	1668.9%	1812.4%
70%	2355.3%	2702.1%	2104.5%	1470.7%	3360.5%	2596.5%	1841.6%	2809.6%	2094.9%	2329.9%
80%	3125.5%	3462.8%	2562.0%	1750.0%	4427.5%	3286.5%	2276.9%	3680.5%	2664.0%	3034.3%
90%	4416.1%	4625.4%	3207.5%	2158.3%	6183.6%	4348.7%	2961.7%	5104.7%	3592.9%	4196.6%

Table 6. Percentage extra premium for level CI cover for females with Beta family of model penetrance curves, maximum penetrance 80%, age at onset in range 20–80 years

Percentage of maximum penetrance at mid-age	Age 20 at entry				Age 30 at entry			Age 40 at entry		Age 50 at entry
	Term 10 yrs	Term 20 yrs	Term 30 yrs	Term 40 yrs	Term 10 yrs	Term 20 yrs	Term 30 yrs	Term 10 yrs	Term 20 yrs	Term 10 yrs
10%	17.0%	75.2%	110.6%	117.9%	111.4%	134.9%	135.3%	153.7%	149.0%	153.9%
20%	36.1%	157.8%	225.6%	227.4%	234.6%	277.8%	265.6%	319.0%	297.2%	306.6%
30%	57.6%	249.7%	345.9%	329.3%	372.6%	430.3%	391.6%	498.2%	444.4%	454.0%
40%	82.5%	353.4%	472.5%	424.6%	529.7%	594.9%	514.5%	694.1%	590.5%	589.2%
50%	111.8%	472.9%	607.3%	514.8%	712.5%	774.8%	636.0%	910.8%	735.8%	701.5%
60%	147.6%	614.7%	752.9%	602.2%	931.9%	975.4%	759.4%	1154.1%	880.6%	773.8%
70%	193.7%	790.3%	914.3%	690.5%	1207.9%	1206.2%	890.4%	1433.0%	1025.5%	780.4%
80%	258.5%	1025.1%	1101.3%	786.7%	1583.9%	1486.9%	1041.0%	1761.7%	1170.4%	684.2%
90%	368.7%	1394.0%	1343.0%	910.2%	2193.0%	1874.8%	1245.7%	2163.2%	1307.2%	438.4%

Table 7. Percentage extra premium for level CI cover for females with Beta family of model penetrance curves, maximum penetrance 100%, age at onset in range 20–80 years

Percentage of maximum penetrance at mid-age	Age 20 at entry				Age 30 at entry			Age 40 at entry		Age 50 at entry
	Term 10 yrs	Term 20 yrs	Term 30 yrs	Term 40 yrs	Term 10 yrs	Term 20 yrs	Term 30 yrs	Term 10 yrs	Term 20 yrs	Term 10 yrs
10%	21.3%	94.1%	138.6%	148.1%	139.5%	169.4%	170.8%	194.2%	190.1%	199.8%
20%	45.1%	197.5%	283.5%	287.3%	294.3%	350.9%	339.0%	408.4%	387.9%	417.8%
30%	72.0%	312.8%	435.8%	418.4%	468.4%	547.1%	505.7%	647.4%	595.8%	658.4%
40%	103.1%	443.1%	597.3%	542.7%	667.5%	761.8%	672.8%	918.6%	817.3%	928.3%
50%	139.8%	593.6%	770.5%	662.0%	900.6%	1000.5%	843.4%	1232.8%	1058.5%	1237.4%
60%	184.6%	772.6%	959.3%	779.5%	1182.1%	1272.2%	1023.1%	1608.3%	1329.6%	1602.7%
70%	242.3%	995.0%	1170.8%	900.2%	1539.3%	1593.2%	1222.2%	2078.3%	1650.7%	2055.8%
80%	323.3%	1293.5%	1419.5%	1034.6%	2031.8%	1997.9%	1463.1%	2716.1%	2067.5%	2668.0%
90%	461.2%	1765.8%	1747.6%	1211.2%	2845.3%	2589.9%	1814.0%	3748.1%	2725.4%	3667.7%
95%	598.3%	2194.1%	1994.6%	1350.9%	3624.4%	3091.9%	2121.0%	4719.8%	3346.8%	4634.4%

maximum penetrance (100%) at age 50. From Table 7, we see that the extra premium in that case is 2,590%. Gui & Macdonald (2002b) gave an extra premium of 2,796%.

5.4 *Other Disorders*

Tables 8 and 9 compare extra CI premiums for a range of mutations, ages at entry and policy terms:

(a) using the penetrance estimates from published actuarial models (Table 8); and
(b) using the tables (Macdonald & Yang, 2003) and the parameters from Table 3, interpolating where necessary (Table 9).

The agreement is very good. The largest discrepancies are for very young lives and short terms, where both the epidemiological evidence and the fitting of the penetrance curve are likely to be based on very few observed cases.

Table 8. Extra percentage of level net premium for level CI cover for females with mutations conferring risk of several genetic disorders, from previous actuarial studies; see Section 3 for references

	Age 20 at entry				Age 30 at entry			Age 40 at entry		Age 50 at entry
Disorder	Term 10 yrs	Term 20 yrs	Term 30 yrs	Term 40 yrs	Term 10 yrs	Term 20 yrs	Term 30 yrs	Term 10 yrs	Term 20 yrs	Term 10 yrs
APKD (Ch)	105%	249%	330%	289%	342%	397%	335%	459%	368%	322%
APKD (US)	233%	298%	282%	232%	344%	304%	245%	290%	232%	197%
BRCA1	457%	1126%	1075%	775%	1637%	1392%	982%	1466%	996%	638%
BRCA2	4%	83%	3010%	392%	132%	391%	475%	592%	662%	935%
BRCA1 & 2	461%	1193%	928%	1239%	1756%	1649%	1231%	1996%	1525%	1558%
EOAD	535%	1940%	1858%	1275%	3151%	2796%	1940%	4078%	2922%	3612%
HD	4301%	3416%	2374%	1650%	3232%	2296%	1625%	2019%	1492%	1326%

Table 9. Extra percentage of level net premium for level CI cover for females from our atlas tables

	Age 20 at entry				Age 30 at entry			Age 40 at entry		Age 50 at entry
Disorder	Term 10 yrs	Term 20 yrs	Term 30 yrs	Term 40 yrs	Term 10 yrs	Term 20 yrs	Term 30 yrs	Term 10 yrs	Term 20 yrs	Term 10 yrs
APKD (Ch)	66%	277%	350%	291%	413%	436%	345%	481%	362%	289%
APKD (US)	226%	281%	257%	213%	320%	273%	222%	252%	207%	182%
BRCA1	828%	1194%	1093%	811%	1525%	1313%	967%	1378%	1018%	883%
BRCA2	0%	67%	317%	397%	109%	401%	482%	627%	684%	952%
BRCA1 & 2	993%	1447%	1342%	1010%	1883%	1668%	1264%	1890%	1499%	1708%
EOAD	461%	1766%	1748%	1211%	2845%	2590%	1814%	3748%	2725%	3668%
HD	3892%	2970%	2070%	1462%	2632%	1887%	1368%	1626%	1250%	1189%

5.5 *Uncertainty*

It is also possible, in some cases, to obtain some information about the stochastic uncertainty of insurance premiums based on estimates of penetrance functions for which we have some measure of uncertainty. Taking BRCA1 and breast/ovarian cancer combined as an example, the confidence intervals in Figure 10 show that a penetrance at the mid-age point of about 40–70% of the maximum penetrance Q would not be ruled out, and (continuing the example) Tables 4 and 6 with $Q \approx 0.8$ then suggest an extra premium of 762–1,504% for the 20-year policy for a woman aged 30.

A range of extra premiums suggested by the tables is not a confidence interval, and where possible it would be preferable to find a confidence interval. This is usually difficult, however, unless the original data from the genetical study are available. In the case of Figure 10, a parametric function was fitted to the five point estimates by least squares, using the confidence intervals to derive approximate weights. This is purely curve fitting, not modelling, and we do not obtain a variance matrix for the parameters that would allow us to find moments of actuarial functions by (for example) bootstrapping. The question of the uncertainty of premium rates based on medical studies, genetical or otherwise, has not received much attention and is worthy of further research.

6. Conclusions

The key quantity for translating epidemiological knowledge about single-gene disorders into actuarial models is the penetrance function. This can be represented by suitably scaled distribution functions, and we have found Beta and sine functions to be very suitable.

In almost all cases, penetrance estimates are based on quite small samples, and have wide confidence intervals. In the absence of the underlying data, these are not sufficient to obtain confidence intervals of derived

quantities such as insurance premiums, an important gap in the actuarial toolkit. However they do show:

(a) that simple representative functions (like our Beta functions) are often adequate, and sometimes very good, as models of the penetrance function; and
(b) that the confidence intervals of the penetrance estimates will often admit a wide range of these representative penetrance functions, even if we cannot assign probabilities to them.

A wide spectrum of representative penetrance functions can be defined in terms of just a few quantities, namely age range $[x_m, x_M]$, maximum penetrance Q, and 'shape'. 'Shape' is defined by the choice of baseline function and a distortion that shifts the ages of onset up or down; we found a one-parameter proportional hazards transform to be suitable, conveniently parameterised in terms of the penetrance in the middle of the age range.

Compared with estimated penetrance functions from the actuarial literature, these simple families, *via* the associated tables (that are available in Macdonald & Yang (2003)) gave very acceptable estimates of extra premiums for CI insurance, in respect of known mutation carriers. Moreover, the same tables allow us to suggest reasonable bounds for such extra premiums. These bounds are not confidence intervals, which are usually unobtainable from published penetrance estimates: the bigger problem of obtaining the statistical properties of premium rates from the statistical properties of penetrance estimates remains to be solved, and it deserves attention in future.

Using these tables, any actuary can superimpose penetrance estimates from the growing genetics literature, whether in the form of Kaplan-Meier, actuarial or point estimates, on a plot of Beta or sine functions to obtain rough values for x_m, x_M, Q and the mid-age penetrance, and hence read off approximate CI extra premiums, and get some idea about bounds for CI extra premiums.

Acknowledgements

This work was carried out at the Genetics and Insurance Research Centre at Heriot-Watt University. We would like to thank the sponsors for funding, and members of the Steering Committee for helpful comments at various stages. One author (SYY) is supported by a Heriot-Watt University Scholarship.

References

Bell, J. (1934). Huntington's chorea. *Treasury of Human Inheritance*, **IV**, 1-67.

Churchill, D., Bear, J., Morgan, J., Payne, R., McManamon, P. & Gault, H. (1984). Prognosis of adult onset polycystic kidney disease re-evaluated. *Kidney International*, **26**, 190-193.

Daykin. C.D., Akers, D.A., Macdonald, A.S., McGleenan, T., Paul, D. & Turvey, P.J. (2003). Genetics and insurance — some social policy issues. *British Actuarial Journal*, **9**, 787-874.

Dinani, A., Grimshaw, D., Robjohns, N., Somerville, A.S., Spry, A. & Stafforth, J. (2000). A critical review: report of the critical illness healthcare study group. Presented to the Staple Inn Actuarial Society.

Ford, D., Easton, D.F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D.T., Weber, B., Lenoir, G., Chang-Claude, J., Sobol, H., Teare, M.D., Struewing, J., Arason, A., Scherneck, S., Peto, J., Rebbeck, T.R., Tonin, P., Neuhausen, S., Barkardottir, R., Eyfjord, J., Lynch, H., Ponder, B.A.J., Gayther, S.A., Birch, J.M., Lindblom, A., Stoppa-Lyonnet, D., Bignon, Y., Borg, A., Hamann, U., Haites, N., Scott, R.J., Maugard, C.M., Vasen, H., Seitz, S., Cannon-Albright, L.A., Schofield, A., Zelada-Hedman, M. & the Breast Cancer Linkage Consortium (1998). Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *American Journal of Human Genetics*, **62**, 676-689.

Gui, E.H. & Macdonald, A.S. (2002). A Nelson-Aalen estimate of the incidence rates of early-onset Alzheimer's disease associated with the Presenilin-1 gene. *ASTIN Bulletin*, **32**, 1-42.

Gutiérrez, M.C. & Macdonald, A.S. (2003). Adult polycystic kidney disease and critical illness insurance. *North American Actuarial Journal*, **7**:2, 93-115.

Hodge, S.E. (2002), Ascertainment. In *Biostatistical genetics and genetic epidemiology*. Ed. Elston, R., Olson, J. & Palmer, L., John Wiley.

Hoem, J.M. (1988). The versatility of the Markov chain as a tool in the mathematics of life insurance. *Transactions of the 23rd International Congress of Actuaries,*, Helsinki **S**, 171-202.

Macdonald, A.S. (2003). Genetics and insurance: what have we learned so far? *Scandinavian Actuarial Journal*, 2003, 324-348.

Macdonald, A.S., Waters, H.R. & Wekwete, C.T. (2003a). The genetics of breast and ovarian cancer. I: a model of family history. *Scandinavian Actuarial Journal*, 2003, 1-27.

Macdonald, A.S., Waters, H.R. & Wekwete, C.T. (2003b). The genetics of breast and ovarian cancer. II: a model of critical illness insurance. *Scandinavian Actuarial Journal*, 2003, 28-50.

Macdonald, A.S. & Yang, S.Y. (2003). Tables of representative premiums for critical illness insurance policies in the presence of model genetic disorders. Occasional actuarial research document No. 42, Faculty and Institute of Actuaries.

Roos, R.A.C., der Vlis, M.V., Hermans, J., Elshove, H.M., Moll, A.C., van de Kamp, J.J.P. & Bruyn, G.W. (1991). Age at onset in Huntington's disease: effect of line of inheritance and patient's sex. *Journal of Medical Genetics*, **28**, 515-519.

Smith,C. (1998). *Huntington's chorea — a mathematical model for life insurance*. Swiss Reinsurance Company, Zurich.

United States Renal Diseases System (1999). *Annual report of the United States Renal Diseases System*. USRDS.

Wendt, G.G. & Drohm, C. (1972). Die Huntingtonsche chorea. Eine populations gentische studie. *Advances in Human Genetics*, **IV**, 1-121.

Wilkie, A.D. (2000). Report by the Independent Actuary on the application by the Association of British Insurers to the Genetics and Insurance Committee for approval to use genetic test results for insurance risk assessment — use of Huntington's disease test in life insurance. Report available by request to the Genetics and Insurance Committee.

APPENDIX

THE CRITICAL ILLNESS INSURANCE MODEL

Gutiérrez & Macdonald (2003) obtained the following model for CI insurance based on medical studies and population data. Full references can be found in that paper.

(a) Rates of onset were found for:

(1) *Cancer (excluding non-malignant skin cancers)*: For males:

$$\mu_x^c = \exp(-11.25 + 0.105x) \quad (x < 51)$$
$$\mu_x^c = 0.2591585 - 0.01247354x + 0.0001916916x^2 - 8.952933 \times 10^{-7}x^3 \quad (x \geq 60)$$

with linear interpolation between ages 51 and 60, and for females:

$$\mu_x^c = \exp(-10.78 + 0.123x - 0.00033x^2) \quad (x < 53)$$
$$\mu_x^c = -0.01545632 + 0.0003805097x \quad (x \geq 53).$$

(2) *Heart attack*: for males:

$$\mu_x^h = \exp(-13.2238 + 0.152568x) \quad (x < 44)$$
$$\mu_x^h = (-0.01245109 + 0.000315605x) \quad (x > 49)$$

with linear interpolation between ages 44 and 49, and for females:

$$\mu_x^h = 0.598694\left(\frac{0.15317^{15.6412}\exp(-0.15317x)x^{14.6412}}{\Gamma(15.6412)}\right).$$

(3) *Stroke*: for males:

$$\mu_x^s = \exp(-16.9524 + 0.294973x - 0.001904x^2 + 0.00000159449x^3)$$

and for females:

$$\mu_x^s = \exp(-11.1477 + 0.081076x).$$

(b) 28-day survival factors for heart attack and stroke victims were taken from Dinani *et al.* (2000) (this relates to the common contractual condition, that payment depends on surviving for 28 days). Let p_x^h and p_x^s be the 28-day survival probabilities after the first-ever heart attack or

stroke, respectively, and $q_x^h = 1 - p_x^h$, $q_x^s = 1 - p_x^s$ the corresponding mortality rates. From Dinani *et al.* (2000), $q_x^h = 0.21$ at ages 20–80 for females, and q_x^h for males is given in Table 10. From the same source, $p_x^s = (0.9 - 0.002x)/0.9$ for both males and females.

(c) Other minor causes of CI insurance claims amount to about 15% of those arising from cancer, heart attack and stroke. Therefore the aggregate rate of CI claims is:

$$\mu_x^{CI} = 1.15(\mu_x^c + p_x^h \times \mu_x^h + p_x^s \times \mu_x^s).$$

(d) Population mortality rates (English Life Tables No. 15) were adjusted to exclude deaths which would have followed a CI insurance claim.

Table 10. 28-day mortality rates ($q_x^h = 1 - p_x^h$) following heart attack; based on Dinani *et al.* (2000)

Age	q_x^h	Age	q_x^h	Age	q_x^h	Age	q_x^h
20–39	0.15	47–52	0.18	58–59	0.21	65–74	0.24
40–42	0.16	53–56	0.19	60–61	0.22	75–79	0.25
43–46	0.17	57	0.20	62–64	0.23	80+	0.26